Certified Human Trajectory Prediction

Mohammadhossein Bahari*,[1] Saeed Saadatnejad*,[1] Amirhossein Askari Farsangi[1]
Seyed-Mohsen Moosavi-Dezfooli†,[2] Alexandre Alahi[1]

[1]EPFL [2]Apple
{mohammadhossein.bahari, saeed.saadatnejad}@epfl.ch

Abstract

Predicting human trajectories is essential for the safe operation of autonomous vehicles, yet current data-driven models often lack robustness in case of noisy inputs such as adversarial examples or imperfect observations. Although some trajectory prediction methods have been developed to provide empirical robustness, these methods are heuristic and do not offer guaranteed robustness. In this work, we propose a certification approach tailored for trajectory prediction that provides guaranteed robustness. To this end, we address the unique challenges associated with trajectory prediction, such as unbounded outputs and multi-modality. To mitigate the inherent performance drop through certification, we propose a diffusion-based trajectory denoiser and integrate it into our method. Moreover, we introduce new certified performance metrics to reliably measure the trajectory prediction performance. Through comprehensive experiments, we demonstrate the accuracy and robustness of the certified predictors and highlight their advantages over the non-certified ones. The code is available online: https://s-attack.github.io/

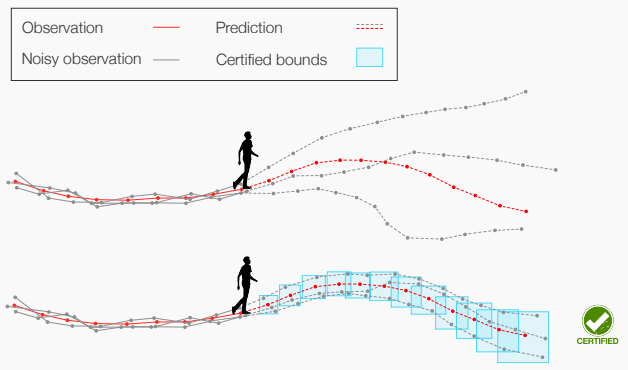

Figure 1. Illustration of the influence of noisy inputs on trajectory prediction models. The red trajectories depict clean observations and the corresponding predictions, while the gray trajectories show noisy observations along with their resulting predictions. The top part showcases the outputs of a standard trajectory prediction model, revealing unbounded predictions with noisy inputs. In contrast, the bottom part demonstrates the outputs of our trajectory predictor with guaranteed robustness. The model provides certified bounds (blue boxes) on the predicted outputs, ensuring that outputs remain within guaranteed regions despite input noise.

1. Introduction

Predicting the behavior of humans is a crucial task for the safe operation of autonomous vehicles and robots. The task, known as human trajectory prediction, aims to predict the future positions of humans given their past positions. It has received significant attention in recent years, with data-driven methods demonstrating remarkable performance [24, 50, 59]. Such progress prompts a critical question: *Are these methods reliable enough for real-world applications with noisy inputs?* Notably, it has been shown that these methods are susceptible to adversarial attacks, raising significant concerns regarding their robustness and security [11, 48, 54]. Moreover, recent findings indicate that in real-world autonomous driving pipelines, inputs of the prediction models are imperfect, resulting in performance drops [60]. This noise stems from upstream modules in the perception pipeline, such as detection and tracking. Therefore, it is crucial to study the robustness properties of trajectory predictors.

Previous works proposed heuristic approaches to improve the empirical robustness of the trajectory prediction models [12, 32, 66]. However, it has been shown that such heuristic approaches are ultimately ineffective against sufficiently powerful adversaries [2, 13, 55]. Therefore, it is essential to study certification techniques that provide *guaranteed* robustness *i.e.,* to guarantee that given any confined input noise, the models' outputs fall within certified bounds. Given the black-box nature of the prediction models, the

* Equal contribution.
† Work done while at Imperial College London.

bounds on the outputs also deliver reliability to the system which is crucial for autonomous vehicles. We illustrate in Figure 1 that even slight perturbations in observation inputs considerably changes model's predictions. Nevertheless, predictions from a model with guaranteed robustness invariably stay within its certified bounds.

In this work, we propose a certification method for trajectory prediction based on "Randomized Smoothing" [18]. In principle, randomized smoothing transforms a base model into a smoothed model by adding random perturbations to the input, and then aggregating the outputs. Up to our knowledge, we are the first to study this in trajectory prediction, encountering various challenges: (1) How can we transform the randomized smoothing technique widely applied in image classification to the multi-output regression task of trajectory prediction? Moreover, it is a time-series regression task that does not inherently have a maximum output range, which is essential for randomized smoothing. How can we define a range for the outputs? (2) While randomized smoothing is known to hurt performance in classification [18], to what extent does it hurt the accuracy of trajectory predictors? Is there a way to maintain the accuracy? (3) Finally, trajectory predictors are often multi-modal, generating multiple output modes. How can these multi-modal outputs be accommodated in the certification process? In order to address the aforementioned challenges, (1) we adapt two randomized smoothing approaches based on mean [18] and median [17] aggregation functions to trajectory prediction and compare their performances. We also propose an adaptive clamping strategy to pose maximum output ranges. (2) To mitigate the performance degradation resulting from randomized smoothing, we propose a denoiser as a pre-processor suppressing the noise before feeding to the smoothed model. Given the success of diffusion models, we design an unconditional diffusion-based denoiser tailored for trajectory data. (3) Finally, we address the multi-modality challenge by proposing a new certification definition that can accommodate multi-modal outputs.

We conduct experiments employing state-of-the-art trajectory prediction models trained on Trajnet++ benchmark [34]. We develop smoothed trajectory prediction models with guaranteed robustness and demonstrate both their accuracy and the certified bounds of their predictions. The results highlight the advantages of the smoothed models over non-certified models in noisy input settings. They also indicate that the most accurate models are not necessarily the most robust. In addition, we show that common performance metrics for the trajectory prediction task are unreliable as they cannot account for the potential input noises. To address this, we introduce new certified metrics, equipped with the certified bounds.

In summary, our contributions are as follows:

- We are the first to introduce certification to the trajectory prediction task, providing guaranteed robustness for models against adversarial attacks and imperfect inputs.
- We develop a randomized smoothed trajectory predictor tailored to the unique challenges of the task and propose an unconditional diffusion denoiser to enhance the performance.
- We introduce new certified performance metrics and through comprehensive experiments, demonstrate the accuracy and robustness of the smoothed models and highlight their advantages over non-certified models.

2. Related Works

Human trajectory prediction. In recent years, as autonomous driving systems and social robots have become more popular, the challenge of predicting human trajectories has caught much attention. The majority of the research revolves around enhancing accuracy by learning the interaction dynamics between humans more effectively. To this end, Social-LSTM [1] is the pioneering work employing neural networks. Subsequent studies propose different architectural solutions based on Convolutional Neural Networks [44, 63], Graph Neural Networks [9, 43], and Transformers [22, 25, 36, 50]. Additional approaches have incorporated the domain knowledge [35, 39], developed equivarient feature learning [24, 59] and explored various strategies for pooling social features [5, 31, 34].

Robustness for trajectory prediction. The vulnerability of trajectory predictors to adversarial attacks has been shown in several previous works [11, 48, 54]. To address this vulnerability, others proposed robustness defenses based on various heuristic approaches [4, 12, 32, 66]. However, none of these approaches are guaranteed robustness methods. Recently, Trajpac [65] proposed a verification approach for the robustness of trajectory predictors. They employ a probably approximately correct (PAC) strategy by approximating the prediction model locally with a linear model and use it as a proxy to determine the robustness of the model. However, their method has some limitations: (1) Their method is not agnostic to the input noise distribution due to the dependency of learned linear model on the noise distribution fed during learning. (2) Their method is inefficient in the number of required samples, with experiments often necessitating over $30,000$ random samples. (3) Their method is probabilistic, and does not provide a guaranteed robustness. In this work, we employ a randomized smoothing approach that provides a certified bound, requires significantly fewer samples, and generalizes to any noise distribution encountered during deployment.

Randomized smoothing certification. Certification is to guarantee that a model's outputs are within a bound around its initial output once the model's inputs are within a neighborhood of its initial input and is mainly used as a defense

against adversarial attacks. Various certification and verification methods have been proposed based on Satisfiability Modulo theories [20, 29], mixed integer linear programming [7, 21], solving optimization problems [19, 58] and layer by layer outer approximation of activations [53]. However, these methods are computationally expensive and cannot scale to common neural networks. Alternatively, randomized smoothing has been proposed as an efficient and model-agnostic approach and has achieved great success in the classification task [10, 14, 38]. More importantly, it imposes the least assumptions on the noise distribution, providing robustness against any confined noise. In randomized smoothing, the smoothed prediction for a given input is calculated by sampling some points around that input and aggregating their corresponding outputs. Cohen et al. [18] proved certified bounds for the smoothed predictors with a mean aggregator, particularly for the classification task. Moreover, it was shown that integrating a denoiser into a smoothed predictor can greatly enhance both accuracy and certified bounds [51]. Recently, randomized smoothing certification has been adapted for the detection task [17]. It introduces a median smoothing aggregator which is more appropriate for regression tasks. To the best of our knowledge, our work is the first randomized smoothing certification for the trajectory prediction problem, studying both mean and median smoothing.

Randomized smoothing is distinct from other methods that guarantee models' output such as conformal prediction [52] as conformal prediction provides the guarantee of ground truth coverage rather than the guarantee of the output region. Moreover, unlike randomized smoothing, conformal prediction is dependent on the input noise distribution (calibration set). Randomized smoothing is also different from uncertainty quantification approaches [30] and they serve distinct yet complementary purposes. Uncertainty estimation quantifies the model's uncertainty (aleatoric or epistemic) for a given input, while randomized smoothing transforms the original model's outputs into a new output with bounds through smoothing.

Denoising diffusion. Denoising diffusion probabilistic models [28] have achieved great success in image generation [23, 33, 47, 64], human pose prediction [49], GPS trajectory generation [67] and even trajectory prediction [3, 26, 42, 56]. However, the application of these models as denoisers remains largely underexplored with only a few studies investigating their use as denoisers in other domains such as image restoration [61, 68]. In training a diffusion model, Gaussian noise is progressively added to the input during the forward process. The model is then trained to reverse this process, recovering the input over several steps. This makes the diffusion model particularly suitable for denoising tasks on any noisy signal. We are the first to propose an unconditional diffusion-based denoiser for trajectory data and integrate it into our randomized smoothed predictor.

3. Method

In this section, we first explain the certification framework backgrounds and then describe our certification for the trajectory prediction task.

3.1. Certification

Randomized smoothing [18] is a technique initially introduced for certifying the robustness of models against ℓ_2-norm adversarial attacks in image classification. Given a prediction function f , randomized smoothing aims to bound the output of a smoothed function $\tilde{f} = \mathcal{A}(f)$ where $\mathcal{A}$ is an aggregation/smoothing operator. This bound is valid for a radius in the neighborhood of the input, named certification radius R.

We consider two choices for the smoothing operator: mean [18] and median [17] smoothing.

Mean smoothing. Given a function $f : \mathbb{R}^d \to [l, u]$, input $X \in \mathbb{R}^d$, and input certification radius R, mean smoothing computes the expected value of the predictor over a perturbed input, that is $\tilde{f}(X) = \mathbb{E}_\epsilon[f(X + \epsilon)]$, where $\epsilon \sim N\left(0, \sigma^2 I\right)$. It has been shown that given $\|r\|_2 < R$, the output of $\tilde{f}$ can be bounded as:

$$\begin{aligned} l + (u - l) \cdot \Phi\left(\frac{\eta(X) - R}{\sigma}\right) \leq \tilde{f}(X + r), \\ \tilde{f}(X + r) \leq l + (u - l) \cdot \Phi\left(\frac{\eta(X) + R}{\sigma}\right), \end{aligned} \tag{1}$$

where $\eta(X) = \sigma \cdot \Phi^{-1}\left(\frac{\tilde{f}(X)-l}{u-l}\right)$ and Φ is the cumulative distribution function of the standard Gaussian. We refer to the lower and upper certified bounds as LB and UB, respectively.

Trajectory predictors are commonly multi-output $f : \mathbb{R}^d \to [l_1, u_1] \times [l_2, u_2] \times \cdots \times [l_m, u_m]$, where $f(X) = (f_1(X), \ldots, f_m(X))$. In this case, the certification bounds are applicable individually to each coordinate. We will explain the estimation process for l_i's and u_i's in Section 3.2.

Note that this smoothing method is applicable to functions with initial lower and upper bounds. However, for functions that inherently lack those, an alternative option is to use median smoothing.

Median smoothing. Given a continuous function $f : \mathbb{R}^d \to \mathbb{R}^m$, and an input certification radius R, median smoothing aims to find a bound for the median of predictions, as given by $\tilde{f}(X) = q_{0.5}(X)$, where $q_p(X) = \sup\{y \in \mathbb{R} \mid \mathbb{P}[f(X+\epsilon) \leq y] \leq p\}$ is the quantile function with $q_{0.5}$ indicating the median and $\epsilon \sim N(0, \sigma^2 I)$. Then,

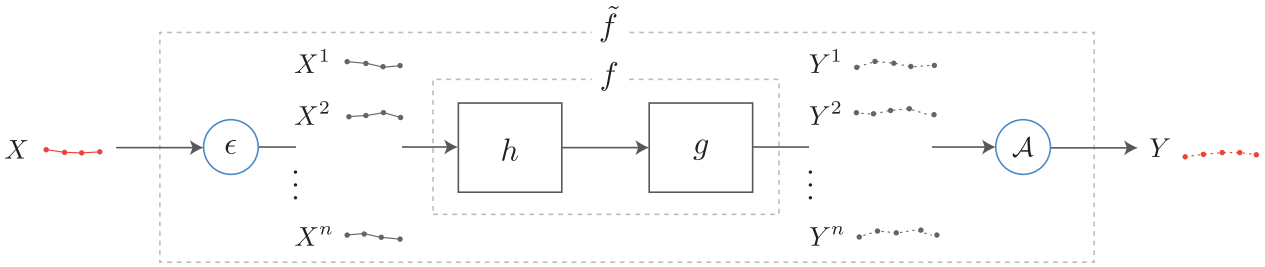

Figure 2. An outline of the proposed smoothed trajectory predictor: n different randomized input observations X^i are created by adding perturbation ϵ to the input X. The denoiser h processes these samples X^i, which are then fed into the trajectory predictor g to make the outputs Y^i. Applying an aggregation function $\mathcal{A}$ (median or mean) on Y^i, the final smoothed prediction Y is derived.

the certified bounds for $\|r\|_2 \leq R$ are as follows:

$$q_{\Phi(-\frac{R}{\sigma})}(X) \leq \tilde{f}(X+r) \leq q_{\Phi(\frac{R}{\sigma})}(X). \tag{2}$$

Similar to the mean smoothing, we refer to the lower and upper certified bounds as LB and UB, respectively.

> In simple words, with certification, we ensure that if an input to the smoothed predictor is perturbed within a radius R, the output remains within a certified range.

Note that, while R is a parameter determined by the application's requirements, σ serves as a hyperparameter. As we will see later, σ can be adjusted to balance between performance and the tightness of bounds. For instance, when $\sigma = 0$, the output aligns closely with the original predictor, and it yields trivial bounds $l \leq \tilde{f} \leq u$. As σ increases, the influence of the perturbation becomes more pronounced, causing the certified bounds to tighten, albeit with more smoothed/less accurate predictions. We will analyze the effect of σ in Section 4.

3.2. Certified Trajectory Prediction

Human trajectory prediction tackles a regression task with sequences as inputs and outputs. The position of an agent at any timestep t is represented by its xy-coordinates (x_t, y_t). Given an observation sequence for T_{obs} timesteps as $X = (x_{-T_{\text{obs}}+1}, y_{-T_{\text{obs}}+1}, \ldots, x_0, y_0)$, the model predicts the next T_{pred} positions $Y = g(X) = (x_1, y_1, \ldots, x_{T_{\text{pred}}}, y_{T_{\text{pred}}})$, aiming to be close to the ground-truth $\hat{Y}$. Notably, the trajectory predictor g can be construed as a function mapping $\mathbb{R}^{2T_{\text{obs}}} \rightarrow \mathbb{R}^{2T_{\text{pred}}}$, making it suitable for certification purposes with $d = 2T_{\text{obs}}$ and $m = 2T_{\text{pred}}$.

Figure 2 provides an overview of our approach. Initially, we acquire n Monte-Carlo samples from $\epsilon \sim N(0, \sigma^2 I)$, adding them to input X to get $X^1, \cdots, X^n$. They are then processed by our denoiser h. The certified bounds for $\tilde{f}(X) = \mathcal{A}(g(h(X+\epsilon)))$ are then computed according to Equation (1) and Equation (2) . Note that $\mathcal{A}$ represents the aggregation function (either median or mean) applied to $Y^1, \cdots, Y^n$ to yield the final smoothed prediction Y. In the followings, we explain the details of the method, and defer the full algorithm to the supplementary.

Diffusion denoiser. The denoised smoothing technique combines a classifier with a denoiser, by first passing perturbed inputs through the denoiser to pre-process them before being fed into the model [51]. Extending this technique to trajectory prediction, we combine the predictor g with h, making $f(X) = g(h(X))$. The denoiser suppresses the noise before feeding the data to the predictor, resulting in tighter certified bounds for the composed model f. In an optimal scenario, where the denoiser exhibits high efficacy $(h(X+\epsilon) \approx X)$, we obtain pseudo-clean data for g, leading to prediction performance closely resembling that of original data. When the denoiser is absent, we put $h = \text{id}$ and the certification is hold for $f(X) = g(X)$. We propose a simple autoencoder for h designed to unconditionally denoise the input. This model is trained independently from the predictor through multiple steps of a denoising diffusion process, enabling it to learn the distribution of trajectory data. At inference time, the diffusion process is repeated for the required number of steps in order to denoise the input trajectory. We explain more our diffusion model in the supplementary.

Adaptive clamping. As mentioned in Section 3.1, to establish certified bounds in case of mean aggregation for the multi-output human trajectory predictor with $m = 2T_{\text{pred}}$ outputs, one needs to compute l_j's and u_j's. However, the output of trajectory predictors inherently lacks bounds due to the unrestricted nature of the predicted positions. To address this for our certification equations, we propose adaptive clamping. The process involves computing the predictions given all X in the training dataset. By determining the maximum and minimum values from these computations, we establish $l_j = \min_X f_j(X)$ and $u_j = \max_X f_j(X)$ for each coordinate j. However, these bounds are not guaranteed. In other words, with new samples, the predictor may predict outside these estimated bounds. Therefore, we can-

Note that while ϵ has a Gaussian distribution, bounds are valid for any noise distribution within radius R.

not derive the certified bounds using the previous equations. To address this, all coordinates of the predicted trajectories, $f_j(X)$'s, are clamped with $\min(u_j, \max(l_j, .))$ operator to ensure conformity within the specified range. Note that one advantage of median smoothing is that the initial bounds are not required.

Certified metrics. The smoothed predictor generates a predicted trajectory with a certified bound around each predicted timestep. In order to assess them, we introduce the following metrics:

- **Average / Final Bound half-Diameter (ABD/FBD)**: ABD measures the distance of the farthest points within the bound from the predicted trajectory, averaged over all timesteps, and FBD measures this distance at the final timestep as:

$$\text{FBD} = \frac{1}{2}[(\text{UB}_{2T_{\text{pred}}-1} - \text{LB}_{2T_{\text{pred}}-1})^2 + (\text{UB}_{2T_{\text{pred}}} - \text{LB}_{2T_{\text{pred}}})^2]^{0.5}. \quad (3)$$

- **Certified-ADE / Certified-FDE:** The common Average/-Final Displacement Error (ADE/FDE) metrics are typically reported under the assumption of perfect inputs. However, in practical scenarios, various types of input noise can occur, which can significantly alter the performance of predictors. To address this gap, we propose these metrics that measure the highest ADE/FDE happening given noisy inputs. Specifically, they measure the distance of the farthest point inside the bounds to the ground-truth trajectory.
- **Certified Collision Rate (Certified-Col):** Collision rate has been previously introduced as a metric that quantifies the percentage of collisions between the predicted trajectory of an agent and the ground-truth trajectories of neighboring agents in the scene [34]. We introduce this metric as the percentage of examples in which at least one neighboring agent lies within the calculated certified bounds of the predicted trajectory.

Multi-modality. Unlike the classification task, the trajectory prediction is a multi-modal task *i.e.,* given an input trajectory, multiple plausible trajectories can be predicted as output. Nonetheless, it poses a unique challenge for certification in multi-modal predictors generating k modes on which mode to consider. To address this, we reformulate it into multi-output mapping $\mathbb{R}^{2T_{\text{obs}}} \rightarrow \mathbb{R}^{k \times 2T_{\text{pred}}}$, and leverage the fact that each mode captures a specific behavior. Consequently, we certify all k modes and choose the best mode based on the minimum of the Certified-FDE among all modes. The corresponding metrics are then computed for the selected mode.

4. Experiments

Datasets: ETH [45], UCY [37], and WildTrack [15] are well-established datasets containing annotations of human positions in crowded environments. We utilize the Trajnet++ [34] benchmark, which provides a fixed data split and unified pre-processing for these datasets. We used the common input and output lengths of $T_{\text{obs}} = 9$ and $T_{\text{pred}} = 12$.

Baselines: Up to our knowledge, our work is the first to introduce certification for trajectory prediction. Therefore, we conduct a comparative analysis between the mean and median smoothed predictors, representing the two certification techniques introduced. In order to obtain the smoothed predictors, we transform multiple existing trajectory predictors into smoothed models. We employ multiple state-of-the-art learning-based trajectory predictors, namely Directional-Pooling (D-Pool) [34], AutoBot [24], and EqMotion [59]. Moreover, we include Social-Force [27] as a rule-based trajectory predictor.

Metrics: We report the performances in terms of the Average / Final Displacement Error (ADE / FDE) between the model predictions and its ground-truth values, along with FBD, Certified-FDE, and Certified-Col, introduced in Section 3. The reported values are in meters and percentages. For the sake of space, we leave the results on ABD and Certified-ADE for the supplementary.

Implementation details: Throughout the experiments, the number of Monte-Carlo samples n is set to 100, R to 0.1, and σ range to $0.08 - 0.4$. Note that since σ serves as a hyperparameter, this range has been experimentally selected to ensure the models perform effectively.

4.1. Results

We initially report the performance of the predictors and their smoothed counterparts utilizing the median aggregation function in the left part of Figure 3. The figure shows the accuracy of models against the certified bounds, highlighting their accuracy with respect to robustness. Each point on the curves represents an instance of a smoothed predictor with a different hyperparameter σ. Therefore, changing σ allows selecting a model instance with the desired trade off between accuracy and robustness. While the original models (dashed lines in the figure) lack guarantees, the smoothed predictors provide certified bounds, albeit at the expense of a modest increase in FDE (ranging from 1% to 6% for different predictors with the smallest σ). The figure also provides a way to compare the guaranteed robustness of different predictors. Given an FDE value, predictors with a smaller bound (*i.e.,* smaller FBD) have better guaranteed robustness. The figures show that while smoothed EqMotion and smoothed Autobot have similar bounds, smoothed D-Pool has a smaller bound for FDEs below 1.25. Moreover, smoothed Social-Force has the largest bounds due to its significant sensitivity to input perturbations.

In the right part of Figure 3, we show similar curves but

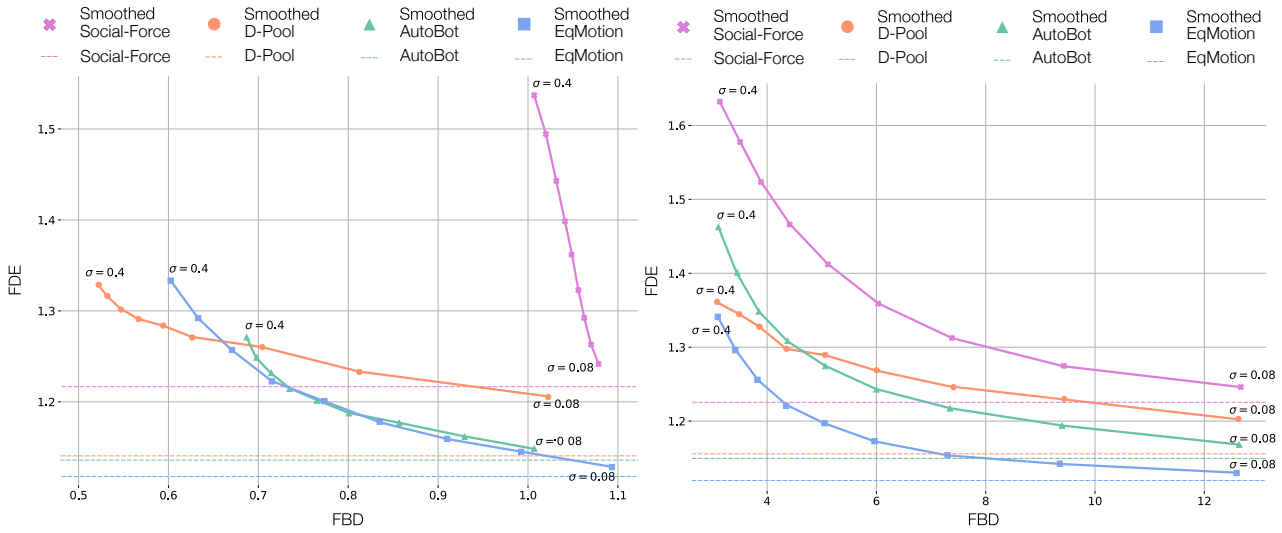

Figure 3. FDE against FBD for *median* aggregation on the left and *mean* aggregation on the right. The results are for different smoothed predictors with two aggregation functions and equally spaced σ within $[0.08, 0.4]$. The bottom left indicates the best performance. The figures show a trade-off between accuracy (represented in FDE) and robustness (represented as FBD). They also provide a comparison between models' guaranteed robustness.

using the mean aggregation function. Comparing the two sub-figures reveals that the median aggregator yields considerably smaller bounds compared to the mean aggregator, demonstrating the better alignment of the median with the trajectory prediction task. This is probably because trajectory predictors are sensitive to input noise, leading to diverse outputs. Consequently, mean aggregation is more susceptible to outliers, whereas the median is less affected. Therefore, for our subsequent experiments, we opt for median aggregation. Moreover, we have selected EqMotion as our main predictor due its superior performance. In the rest of this section, we report experimental results to answer the remaining research questions.

Is there a trade-off between accuracy and certified bound? As shown by Figure 3, by increasing σ, the bounds progressively tighten while the accuracy drops, indicating a trade-off between them (see Equations (1) and (2)). Note that the hyperparameter σ allows users to tailor the certified bound according to their needs. For instance, given a desired FBD of 0.72, we can choose $\sigma = 0.28$ for smoothed EqMotion.

Does the most accurate model have the best guaranteed robustness? We report the performance of models with both certified and non-certified metrics in Table 1. It shows that there is a large gap between FDE and Certified-FDE for all models, revealing models' lack of robustness. This shows the danger of solely relying on non-certified metrics. This analysis also uncovers a noteworthy observation: the model with the minimal FDE (EqMotion) is not the same as the model achieving the lowest Certified-FDE (D-Pool), indicating that a more accurate model is not necessarily more robust. It similarly shows a large gap between Col and Certified-Col for all models. As expected, Social-Force has the lowest collision rate due to its imposed collision-avoidance rules. However, the gap between Col and Certified-Col shows the model's sensitivity to input noise.

What practical advantages does the smoothed predictor offer compared to the original model? To study this question, we examine the robustness of the models in two scenarios: adversarial attacks and real-world perception noise. We first investigate the robustness of the models against adversarial attacks by performing PGD attacks [40]. We demonstrate a scenario in Figure 4 where the left figure shows the existence of an adversarially perturbed input for the original model, leading to large deviations from the original prediction (more than 2m). However, conducting PGD attacks on the smoothed predictor (the right figure) does not lead to predictions outside the certified bounds, demonstrating its guaranteed robustness. We provide a detailed quantitative analysis in the supplementary.

In order to illustrate the impact of imperfect input data (*i.e.,* noisy inputs) coming from perception systems on the predictors, we employ an off-the-shelf joint detection and tracking model [62] to extract observation trajectories on the nuScenes dataset [8]. Figure 4 visualizes a real-world scenario with both the extracted imperfect observation sequence and the ground-truth. On the left, we observe that the imperfect observation influences the prediction of the model, leading to a large deviation from the prediction given ground-truth observation. This clearly shows that the performance of the model is sensitive to the input noise,

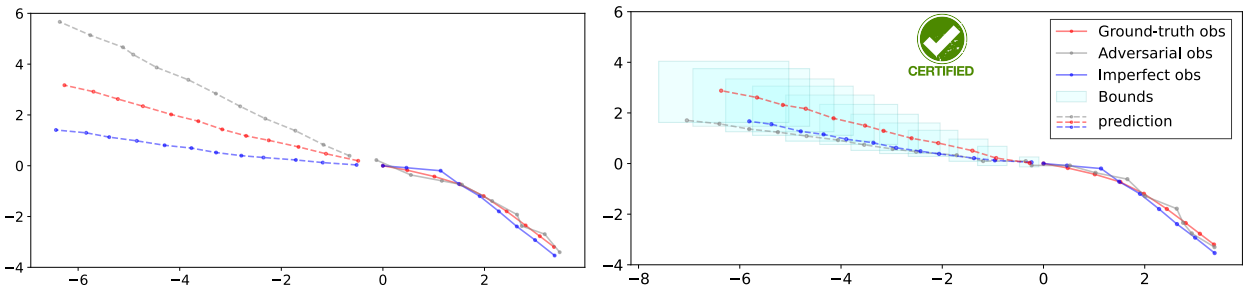

Figure 4. Comparing the performance of the original (on the left) and the smoothed predictor (on the right). The original predictor's outputs change drastically with adversarial and imperfect inputs. In contrast, the smoothed predictor consistently predicts within the bounds, demonstrating higher reliability. For easier comparison, the final predicted points of the original predictor for the adversarial and imperfect observations are marked on the right figure with gray and blue stars, respectively.

Model	FDE	Certified-FDE	Col	Certified-Col
Social-Force [27]	1.25	N/A	7.4	N/A
Smoothed Social-Force	1.26	2.27	8.0	46
D-Pool [34]	1.14	N/A	9.4	N/A
Smoothed D-Pool	1.23	2.0	9.0	49
AutoBot [24]	1.14	N/A	8.8	N/A
Smoothed AutoBot	1.17	2.05	9.3	53
EqMotion [59]	1.12	N/A	10.1	N/A
Smoothed EqMotion	1.14	2.07	10.6	57

Table 1. Comparing performances in terms of certified and non-certified metrics. Since non-smoothed models do not have any guarantee on their outputs, the certified metrics are not applicable (N/A) for them.

making the model unreliable for safety-critical applications. In contrast, on the right, the predictions of the smoothed predictor for both observation sequences remain within the certified bounds, providing a reliable model.

4.2. Discussions

Denoiser analysis. Our proposed trajectory denoiser acts as a pre-processing module and therefore, incorporating it in the smoothing operation improves the certified bounds. In this part, we first compare the performance of different denoisers for trajectory denoising without considering the predictors. To this end, we measure the magnitude of the remaining noise in their outputs when provided with noisy input trajectories at different levels. In Table 2, we report the performance of our proposed diffusion denoiser along with three established denoising methods for time-series data: the Wiener filter [57] as a statistical approach, a Moving Average filter [46] to filter high-frequency noise, and fitting a 4^{th} order polynomial that has been used previously to represent human trajectories [6]. The results demonstrate that the diffusion denoiser outperforms other approaches in noise reduction. Second, we evaluate the effect of the denoiser in the certified bounds of our smoothed predictor and compare it with not having a denoiser (assigning $h = \text{id}$) in Table 3. For similar FDE values, the smoothed predictor with diffusion denoiser has a significantly smaller certified bound, demonstrating the effectiveness of the denoiser in tightening the bounds.

Model	Noise = 0.08	0.24	0.40
No denoiser	0.08	0.24	0.40
Polynomial	0.08	0.22	0.36
Moving Average	0.07	0.18	0.29
Wiener Filter	0.06	0.16	0.26
Diffusion Denoiser (ours)	0.06	0.14	0.24

Table 2. Performance comparison of different denoisers. Noisy trajectories at three noise levels are fed to denoisers, and the remaining noise magnitude is reported.

Multi-modal settings. To analyze certification in multi-modal settings, we employed the multi-modal EqMotion with $k = 20$ [59]. We report the multi-modal metrics defined in Section 3.2 in Table 4. The multi-modal model is more accurate than its single-modal counterpart in terms of both FDE and Certified-FDE, as it captures diverse output

Model	FDE = 1.2	1.3	1.4
W/o denoiser	1.20	0.96	0.80
W/ denoiser	0.78	0.65	0.57

Table 3. Comparing FBD of the smoothed predictor with and without the denoiser across different FDE values.

Model	FDE	FBD	Certified-FDE
Single-modal	1.13	0.99	2.07
Multi-modal	0.39	0.64	1.38

Table 4. Comparison of single- vs. multi-**modal** settings for Smoothed EqMotion.

Model	FDE	FBD
Single-agent	1.13	0.99
Multi-agent	1.13	1.21

Table 5. Comparison of certified bounds in single- vs. multi-**agent** settings in similar FDE values.

modes. Moreover, it has a smaller FBD since each mode concentrates on a specific behavior, leading to a smaller certified bound.

Multi-agent settings. In the real world, the observed trajectories of all agents can be noisy. Our method can be extended to this setting, where instead of perturbing the trajectory of one agent, we consider perturbations added to all agents in the scene. Section 4.2 shows that with a similar FDE, the multi-agent model has a larger bound. Basically, when perturbing all agents, the interdependencies between agents make prediction change more, leading to larger certified bounds.

Downstream task. We also investigated whether improving trajectory prediction with the diffusion denoiser can enhance performance in a downstream task. We considered the task of robot navigation in a dense crowd scenario employing a crowd navigation simulator [16], where the objective for the robot is to navigate through a group of simulated pedestrians and reach a destination. At each timestep, the robot predicts the interactions, and determines its next action. We generated a dataset of 5,000 simulation episodes using the pre-trained SARL policy as the expert, and subsequently trained an imitation model on this data for 200 epochs [16]. We adhered to their protocol, which measures the effectiveness of a policy using the collision rate, and accumulated reward.

The results, summarized in Table 6, indicate that the learned policy is sensitive to input noise, which could be introduced in real-world scenarios. However, when our diffusion denoiser is incorporated, there is a consistent improvement in accumulated reward and a reduction in collision rate. This is due to the improved interaction prediction, resulting in better planning for the robot.

Method	Noise size	Reward ↑	Collision (%) ↓
Original	0	0.272	13.1
Original	0.2	0.230	21.0
Robust	0.2	0.263	15.1

Table 6. Quantitative results of crowd robot navigation [16] with different prediction methods.

4.3. Limitations

Randomized smoothing inevitably increases computational costs in order to provide guaranteed robustness. This is due to the fact that it requires evaluating the predictor n times to obtain Monte-Carlo samples. Nonetheless, this process can be parallelized. Using an NVIDIA GeForce RTX 3090, the evaluation time for predicting a trajectory of 4.8 seconds length in EqMotion is 0.07 seconds, while for smoothed EqMotion with $n = 100$, it is 0.1 seconds which is small enough for many real-world applications (only about 42% increase in the computational time). As future work, one can improve it *e.g.,* by better sampling strategies or optimized code structures.

5. Conclusions

In this work, we introduced a certified trajectory prediction approach that tackles the issue of lacking guaranteed robustness in the human trajectory prediction task. We also proposed a denoiser for trajectory data and introduced new certified metrics that ensure reliable performance assessments under noisy conditions. Throughout extensive experiments on various trajectory predictors, we found that the model with the highest accuracy is not always the most robust one. By adjusting certification parameters, one can prioritize either a tighter certified bound or higher accuracy. Moreover, our experiments demonstrated the edge of smoothed predictors over standard models in the presence of adversarial perturbations or input noise. We hope our work paves the way for more reliable trajectory predictors.

Acknowledgments

The authors would like to thank Brian Sifringer, Taylor Mordan, Ahmad Rahimi, and Yuejiang Liu for their helpful comments. This project was partially funded by Honda R&D Co., Ltd.

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)*, 2016. 2

[2] Anish Athalye, Nicholas Carlini, and David Wagner. Obfuscated gradients give a false sense of security: Circumventing defenses to adversarial examples. In *International Conference on Machine Learning (ICML)*. PMLR, 2018. 1

[3] Inhwan Bae, Young-Jae Park, and Hae-Gon Jeon. Singulartrajectory: Universal trajectory predictor using diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. 3

[4] Mohammadhossein Bahari, Saeed Saadatnejad, Ahmad Rahimi, Mohammad Shaverdikondori, Amir-Hossein Shahidzadeh, Seyed-Mohsen Moosavi-Dezfooli, and Alexandre Alahi. Vehicle trajectory prediction works, but not everywhere. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2

[5] Federico Bartoli, Giuseppe Lisanti, Lamberto Ballan, and Alberto Del Bimbo. Context-aware trajectory prediction. In *International Conference on Pattern Recognition (ICPR)*. IEEE, 2018. 2

[6] Stefan Becker, Ronny Hug, Wolfgang Hübner, and Michael Arens. An evaluation of trajectory prediction approaches and notes on the trajnet benchmark. *arXiv preprint arXiv:1805.07663*, 2018. 7

[7] Rudy R Bunel, Ilker Turkaslan, Philip Torr, Pushmeet Kohli, and Pawan K Mudigonda. A unified view of piecewise linear neural network verification. *Advances in Neural Information Processing Systems (NeurIPS)*, 31, 2018. 3

[8] Holger Caesar, Varun Bankiti, Alex H. Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. *Proceedings of IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 6

[9] Defu Cao, Jiachen Li, Hengbo Ma, and Masayoshi Tomizuka. Spectral temporal graph neural network for trajectory prediction. In *IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021. 2

[10] Xiaoyu Cao and Neil Zhenqiang Gong. Mitigating evasion attacks to deep neural networks via region-based classification. In *Proceedings of the 33rd Annual Computer Security Applications Conference*, pages 278–287, 2017. 3

[11] Yulong Cao, Chaowei Xiao, Anima Anandkumar, Danfei Xu, and Marco Pavone. Advdo: Realistic adversarial attacks for trajectory prediction. In *European Conference on Computer Vision (ECCV)*, pages 36–52. Springer, 2022. 1, 2

[12] Yulong Cao, Danfei Xu, Xinshuo Weng, Zhuoqing Mao, Anima Anandkumar, Chaowei Xiao, and Marco Pavone. Robust trajectory prediction against adversarial attacks. In *Conference on Robot Learning (CoRL)*, pages 128–137. PMLR, 2023. 1, 2

[13] Nicholas Carlini and David Wagner. Adversarial examples are not easily detected: Bypassing ten detection methods. In *Proceedings of the 10th ACM workshop on artificial intelligence and security*, pages 3–14, 2017. 1

[14] Nicholas Carlini, Florian Tramer, Krishnamurthy Dj Dvijotham, Leslie Rice, Mingjie Sun, and J Zico Kolter. (certified!!) adversarial robustness for free! *arXiv preprint arXiv:2206.10550*, 2022. 3

[15] Tatjana Chavdarova, Pierre Baqué, Stéphane Bouquet, Andrii Maksai, Cijo Jose, Timur M. Bagautdinov, Louis Lettry, Pascal Fua, Luc Van Gool, and François Fleuret. Wildtrack: A multi-camera hd dataset for dense unscripted pedestrian detection. *Prooceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 5

[16] Changan Chen, Yuejiang Liu, Sven Kreiss, and Alexandre Alahi. Crowd-Robot Interaction: Crowd-Aware Robot Navigation With Attention-Based Deep Reinforcement Learning. In *IEEE International Conference on Robotics and Automation (ICRA)*, pages 6015–6022, 2019. ISSN: 2577-087X. 8

[17] Ping-yeh Chiang, Michael Curry, Ahmed Abdelkader, Aounon Kumar, John Dickerson, and Tom Goldstein. Detection as regression: Certified object detection with median smoothing. *Advances in Neural Information Processing Systems (NeurIPS)*, 2020. 2, 3

[18] Jeremy Cohen, Elan Rosenfeld, and Zico Kolter. Certified adversarial robustness via randomized smoothing. In *International Conference on Machine Learning (ICML)*, pages 1310–1320. PMLR, 2019. 2, 3

[19] Krishnamurthy Dvijotham, Robert Stanforth, Sven Gowal, Timothy A Mann, and Pushmeet Kohli. A dual approach to scalable verification of deep networks. In *UAI*, page 3, 2018. 3

[20] Ruediger Ehlers. Formal verification of piece-wise linear feed-forward neural networks. In *Automated Technology for Verification and Analysis: 15th International Symposium, ATVA 2017, Pune, India, October 3–6, 2017, Proceedings 15*, pages 269–286. Springer, 2017. 3

[21] Matteo Fischetti and Jason Jo. Deep neural networks and mixed integer linear optimization. *Constraints*, 23(3):296–309, 2018. 3

[22] Luca Franco, Leonardo Placidi, Francesco Giuliari, Irtiza Hasan, Marco Cristani, and Fabio Galasso. Under the hood of transformer networks for trajectory forecasting. *Pattern Recognition*, 138:109372, 2023. 2

[23] Rinon Gal, Yuval Alaluf, Yuval Atzmon, Or Patashnik, Amit H Bermano, Gal Chechik, and Daniel Cohen-Or. An image is worth one word: Personalizing text-to-image generation using textual inversion. *arXiv preprint arXiv:2208.01618*, 2022. 3

[24] Roger Girgis, Florian Golemo, Felipe Codevilla, Martin Weiss, Jim Aldon D'Souza, Samira Ebrahimi Kahou, Felix Heide, and Christopher Pal. Latent variable sequential set transformers for joint multi-agent motion prediction. In *International Conference on Learning Representations (ICLR)*, 2022. 1, 2, 5, 7

[25] Francesco Giuliari, Irtiza Hasan, Marco Cristani, and Fabio Galasso. Transformer networks for trajectory forecasting. In

2020 25th international conference on pattern recognition (ICPR). IEEE, 2021. 2

[26] Tianpei Gu, Guangyi Chen, Junlong Li, Chunze Lin, Yongming Rao, Jie Zhou, and Jiwen Lu. Stochastic trajectory prediction via motion indeterminacy diffusion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 17113–17122, 2022. 3

[27] Dirk Helbing and Peter Molnar. Social force model for pedestrian dynamics. *Physical Review E*, 51, 1998. 5, 7

[28] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *arXiv preprint arxiv:2006.11239*, 2020. 3

[29] Xiaowei Huang, Marta Kwiatkowska, Sen Wang, and Min Wu. Safety verification of deep neural networks. In *Computer Aided Verification: 29th International Conference, CAV 2017, Heidelberg, Germany, July 24-28, 2017, Proceedings, Part I 30*, pages 3–29. Springer, 2017. 3

[30] Eyke Hüllermeier and Willem Waegeman. Aleatoric and epistemic uncertainty in machine learning: An introduction to concepts and methods. *Machine Learning*, 110:457–506, 2021. 3

[31] Boris Ivanovic and Marco Pavone. The trajectron: Probabilistic multi-agent trajectory modeling with dynamic spatiotemporal graphs. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019. 2

[32] Ruochen Jiao, Xiangguo Liu, Takami Sato, Qi Alfred Chen, and Qi Zhu. Semi-supervised semantics-guided adversarial training for robust trajectory prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8207–8217, 2023. 1, 2

[33] Levon Khachatryan, Andranik Movsisyan, Vahram Tadevosyan, Roberto Henschel, Zhangyang Wang, Shant Navasardyan, and Humphrey Shi. Text2video-zero: Text-to-image diffusion models are zero-shot video generators. 2023. 3

[34] Parth Kothari, Sven Kreiss, and Alexandre Alahi. Human trajectory forecasting in crowds: A deep learning perspective. *IEEE Transactions on Intelligent Transportation Systems*, 2021. 2, 5, 7

[35] Parth Kothari, Brian Sifringer, and Alexandre Alahi. Interpretable social anchors for human trajectory forecasting in crowds. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 2

[36] Seongju Lee, Junseok Lee, Yeonguk Yu, Taeri Kim, and Kyoobin Lee. Mart: Multiscale relational transformer networks for multi-agent trajectory prediction. In *European Conference on Computer Vision (ECCV)*. Springer, 2024. 2

[37] Alon Lerner, Yiorgos Chrysanthou, and Dani Lischinski. Crowds by example. *Comput. Graph. Forum*, 26, 2007. 5

[38] Xuanqing Liu, Minhao Cheng, Huan Zhang, and Cho-Jui Hsieh. Towards robust neural networks via random self-ensemble. In *European Conference on Computer Vision (ECCV)*, 2018. 3

[39] Yuejiang Liu, Qi Yan, and Alexandre Alahi. Social nce: Contrastive learning of socially-aware motion representations. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 2

[40] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017. 6

[41] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. *arXiv preprint arXiv:1706.06083*, 2017. 3

[42] Weibo Mao, Chenxin Xu, Qi Zhu, Siheng Chen, and Yanfeng Wang. Leapfrog diffusion model for stochastic trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5517–5526, 2023. 3

[43] Abduallah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14424–14432, 2020. 2

[44] Nishant Nikhil and Brendan Tran Morris. Convolutional neural network for trajectory prediction. In *European Conference on Computer Vision (ECCV) Workshops*, 2018. 2

[45] Stefano Pellegrini, Andreas Ess, and Luc Van Gool. Improving data association by joint modeling of pedestrian trajectories and groupings. In *European Conference on Computer Vision (ECCV)*. Springer, 2010. 5

[46] Lawrence R Rabiner and Bernard Gold. Theory and application of digital signal processing. *Englewood Cliffs: Prentice-Hall*, 1975. 7

[47] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10684–10695, 2022. 3

[48] Saeed Saadatnejad, Mohammadhossein Bahari, Pedram Khorsandi, Mohammad Saneian, Seyed-Mohsen Moosavi-Dezfooli, and Alexandre Alahi. Are socially-aware trajectory prediction models really socially-aware? *arXiv preprint arXiv:2108.10879*, 2021. 1, 2

[49] Saeed Saadatnejad, Ali Rasekh, Mohammadreza Mofayezi, Yasamin Medghalchi, Sara Rajabzadeh, Taylor Mordan, and Alexandre Alahi. A generic diffusion-based approach for 3d human pose prediction in the wild. In *International Conference on Robotics and Automation (ICRA)*, 2023. 3

[50] Saeed Saadatnejad, Yang Gao, Kaouther Messaoud, and Alexandre Alahi. Social-transmotion: Promptable human trajectory prediction. In *International Conference on Learning Representations (ICLR)*, 2024. 1, 2

[51] Hadi Salman, Mingjie Sun, Greg Yang, Ashish Kapoor, and J Zico Kolter. Denoised smoothing: A provable defense for pretrained classifiers. *Advances in Neural Information Processing Systems (NeurIPS)*, 33:21945–21957, 2020. 3, 4

[52] Glenn Shafer and Vladimir Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9(3), 2008. 3

[53] Gagandeep Singh, Timon Gehr, Matthew Mirman, Markus Püschel, and Martin Vechev. Fast and effective robustness

certification. *Advances in Neural Information Processing Systems (NeurIPS)*, 31, 2018. 3

[54] Kaiyuan Tan, Jun Wang, and Yiannis Kantaros. Targeted adversarial attacks against neural network trajectory predictors. In *Learning for Dynamics and Control Conference*, pages 431–444. PMLR, 2023. 1, 2

[55] Jonathan Uesato, Brendan O'donoghue, Pushmeet Kohli, and Aaron Oord. Adversarial risk and the dangers of evaluating against weak attacks. In *International Conference on Machine Learning (ICML)*. PMLR, 2018. 1

[56] Yixiao Wang, Chen Tang, Lingfeng Sun, Simone Rossi, Yichen Xie, Chensheng Peng, Thomas Hannagan, Stefano Sabatini, Nicola Poerio, Masayoshi Tomizuka, et al. Optimizing diffusion models for joint trajectory prediction and controllable generation. In *European Conference on Computer Vision (ECCV)*. Springer, 2024. 3

[57] Norbert Wiener. *Extrapolation, interpolation, and smoothing of stationary time series: with engineering applications*. The MIT press, 1949. 7

[58] Eric Wong and Zico Kolter. Provable defenses against adversarial examples via the convex outer adversarial polytope. In *International Conference on Machine Learning (ICML)*, pages 5286–5295. PMLR, 2018. 3

[59] Chenxin Xu, Robby T Tan, Yuhong Tan, Siheng Chen, Yu Guang Wang, Xinchao Wang, and Yanfeng Wang. Eqmotion: Equivariant multi-agent motion prediction with invariant interaction reasoning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 1, 2, 5, 7

[60] Yihong Xu, Loick Chambon, Mickaël Chen, Alexandre Alahi, Matthieu Cord, Patrick Perez, et al. Towards motion forecasting with real-world perception inputs: Are end-to-end approaches competitive? In *International Conference on Robotics and Automation (ICRA)*, 2024. 1

[61] Cheng Yang, Lijing Liang, and Zhixun Su. Real-world denoising via diffusion model. *arXiv preprint arXiv:2305.04457*, 2023. 3

[62] Tianwei Yin, Xingyi Zhou, and Philipp Krahenbuhl. Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 11784–11793, 2021. 6

[63] Simone Zamboni, Zekarias Tilahun Kefato, Sarunas Girdzijauskas, Christoffer Norén, and Laura Dal Col. Pedestrian trajectory prediction with convolutional neural networks. *Pattern Recognition*, 121:108252, 2022. 2

[64] Lvmin Zhang, Anyi Rao, and Maneesh Agrawala. Adding conditional control to text-to-image diffusion models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3836–3847, 2023. 3

[65] Liang Zhang, Nathaniel Xu, Pengfei Yang, Gaojie Jin, Cheng-Chao Huang, and Lijun Zhang. Trajpac: Towards robustness verification of pedestrian trajectory prediction models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 8327–8339, 2023. 2

[66] Qingzhao Zhang, Shengtuo Hu, Jiachen Sun, Qi Alfred Chen, and Z Morley Mao. On adversarial robustness of trajectory prediction for autonomous vehicles. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 15159–15168, 2022. 1, 2

[67] Yuanshao Zhu, Yongchao Ye, Shiyao Zhang, Xiangyu Zhao, and James Yu. Difftraj: Generating GPS trajectory with diffusion probabilistic model. In *Thirty-seventh Conference on Neural Information Processing Systems*, 2023. 3

[68] Yuanzhi Zhu, Kai Zhang, Jingyun Liang, Jiezhang Cao, Bihan Wen, Radu Timofte, and Luc Van Gool. Denoising diffusion models for plug-and-play image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1219–1229, 2023. 3