

Article

Robust Trajectory Prediction for Mobile Robots via Minimum Error Entropy Criterion and Adaptive LSTM Networks

Da Xie ¹, Zengxun Li ², Chun Zhang ³, Chunyang Wang ^{1,*} and Xuyang Wei ¹

¹ Xi'an Key Laboratory of Active Photoelectric Imaging Detection Technology, Xi'an Technological University, Xi'an 710021, China; xiedacust@163.com (D.X.); xatu9801@163.com (X.W.)

² Institute of Process Materials Technology, Inner Mongolia North Heavy Industries Group Co., Ltd., Baotou 014030, China; milanbaobei@163.com

³ School of Liberal Arts, Xi'an Technological University, Xi'an 710021, China; zhangchun@xatu.edu.cn

* Correspondence: wangchunyang19@163.com

Abstract

Trajectory prediction is critical for safe robot navigation, yet standard deep learning models predominantly rely on the Mean Squared Error (MSE) criterion. While effective under ideal conditions, MSE-based optimization is inherently fragile to non-Gaussian impulsive noise—such as sensor glitches and occlusions—common in real-world deployment. To address this limitation, this paper proposes MEE-LSTM, a robust forecasting framework that integrates Long Short-Term Memory networks with the Minimum Error Entropy (MEE) criterion. By minimizing Renyi's quadratic entropy of the prediction error, our loss function introduces an intrinsic “gradient clipping” mechanism that effectively suppresses the influence of outliers. Furthermore, to overcome the convergence challenges of fixed-kernel information theoretic learning, we introduce a Silverman-based Adaptive Annealing (SAA) strategy that dynamically regulates the kernel bandwidth. Extensive evaluations on the ETH and UCY datasets demonstrate that MEE-LSTM maintains competitive accuracy on clean benchmarks while exhibiting superior resilience in degraded sensing environments. Notably, we identify a “Scissor Plot” phenomenon under stress testing: in the presence of 20% impulsive noise, the proposed model maintains a stable Average Displacement Error (ADE “ $\approx$ ” 0.51 m), whereas MSE baselines suffer catastrophic degradation (ADE > 2.1 m), representing a 75.7% improvement in robustness. This work provides a statistically grounded paradigm for reliable causal inference in hostile robotic perception.

Keywords: trajectory prediction; Minimum Error Entropy (MEE); robust statistics; Long Short-Term Memory (LSTM); non-Gaussian noise

Academic Editors: Xiaoguang Gao and Zidong Wang

Received: 06 January 2026

Revised: 09 February 2026

Accepted: 10 February 2026

Published: 15 February 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the [Creative Commons Attribution \(CC BY\) license](https://creativecommons.org/licenses/by/4.0/).

1. Introduction

The proliferation of autonomous mobile robots in shared human environments—ranging from automated warehouse logistics to social assistive services—has fundamentally transformed robotics from a deterministic control problem into a stochastic uncertainty management challenge [1,2]. In these highly unstructured and dynamically evolving environments, the core competency of a mobile robot lies not merely in self-localization, but in its ability to quantify and mitigate the uncertainty associated with the future

trajectories of surrounding heterogeneous agents, particularly pedestrians [3,4]. Trajectory prediction, therefore, is not simply a geometric regression task; it is a critical prerequisite for high-level cognitive functions. It serves as the foundational information layer for proactive collision avoidance, social-compliant path planning, and complex decision-making in multi-agent settings [5,6]. A robot that cannot accurately anticipate the stochastic intent of a pedestrian is effectively navigating “blind” in the temporal dimension, posing severe safety risks.

Over the past decade, the field of trajectory forecasting has witnessed a paradigm shift from traditional physics-based heuristics to sophisticated data-driven architectures. Early approaches, such as the Social Force Model (SFM) [7] and Velocity Obstacles (VO) [8], relied on handcrafted kinematic constraints to model interactions. While foundational, these methods often struggled to capture the subtle, implicit social cues inherent in high-density crowds. The advent of Deep Learning (DL) revolutionized this domain. Recurrent Neural Networks (RNNs), particularly Long Short-Term Memory (LSTM) networks, became the de facto standard for modeling sequential motion data [9]. The seminal Social-LSTM [10] introduced “social pooling” to aggregate hidden states of neighbors, explicitly modeling interaction. Subsequent innovations integrated Generative Adversarial Networks (GANs) to address the multi-modality of human intent [11,12], and Graph Convolutional Networks (GCNs) to capture the spatial topology of crowds [13,14]. More recently, the Transformer architecture [15], with its self-attention mechanism, has pushed the boundaries of predictive fidelity by modeling long-range temporal dependencies [16,17].

However, a profound theoretical disconnect persists in the current state-of-the-art: while model architectures have become increasingly complex, the underlying probabilistic assumptions regarding sensor noise remain simplistically rigid. In practical deployment, sensory inputs from Light Detection and Ranging (LiDAR), Ultra-Wideband (UWB), and Inertial Measurement Units (IMU) are rarely “clean” [18]. They are frequently compromised by adverse physical phenomena. For instance, UWB sensors in indoor environments suffer from Non-Line-of-Sight (NLOS) propagation and multi-path interference, causing instantaneous range estimates to jump by meters [19]. Similarly, LiDAR point clouds in dynamic crowds are subject to transient occlusions and “ghost” reflections [20]. From an information-theoretic perspective, these measurement errors do not obey the convenient assumptions of benign, zero-mean Gaussian white noise. Instead, they manifest as non-Gaussian impulsive noise, characterized by heavy-tailed probability distributions (e.g., Alpha-stable or Cauchy distributions) and frequent, high-magnitude outliers [21].

The primary technical bottleneck resides in the optimization objectives employed by the overwhelming majority of existing predictors. Current frameworks—including the most sophisticated Transformer variants—predominantly rely on the Mean Squared Error (MSE) or L_2 norm as the definitive loss function [22,23]. From the standpoint of Maximum Likelihood Estimation (MLE), minimizing MSE is optimal if and only if the error distribution is strictly Gaussian [24]. However, imposing a Gaussian prior on non-Gaussian, heavy-tailed data results in a fundamental mismatch of information potential. The quadratic penalty structure of MSE ($\text{Loss} \propto e^2$) assigns exponentially increasing information weight to residuals. Consequently, a single outlier—representing a sensor glitch—can generate a massive gradient value, exerting a disproportionate influence on the backpropagation process [25]. We define this phenomenon as “Parameter Hijacking,” where the network’s weight adaptation is effectively commandeered by erroneous measurements. Under such conditions, the model is forced to “memorize” the noise to minimize the global entropy of the residuals, significantly degrading its generalization capability and causal fidelity [26].

To bridge the gap between high-capacity temporal modeling and statistical robustness, this paper proposes a robust trajectory prediction framework integrating Information Theoretic Learning (ITL) with deep recurrent architectures. Unlike traditional second-order statistics (mean and variance) that rely on Gaussian assumptions, ITL focuses on the global morphology of the error Probability Density Function (PDF) [27,28]. We introduce the Minimum Error Entropy (MEE) criterion, which seeks to minimize the Renyi quadratic entropy of the prediction error [29,30]. By maximizing the Information Potential (IP) of the error samples, MEE forces the error distribution to be tightly concentrated around zero, regardless of its shape or tail behavior. Theoretical analysis reveals that the gradient of the MEE loss contains an intrinsic “soft-thresholding” mechanism: when an error sample deviates significantly from the central distribution (i.e., acts as an outlier), its contribution to the overall weight update naturally diminishes towards zero [31,32]. This property effectively “gates” out impulsive noise without requiring manual threshold tuning.

Furthermore, a significant challenge in applying ITL to deep learning is the selection of the kernel bandwidth (σ), often referred to as the “bandwidth trap.” A fixed small bandwidth causes vanishing gradients (slow convergence), while a fixed large bandwidth degrades the criterion back to MSE (loss of robustness) [33]. To overcome this, we introduce a Silverman-based Adaptive Annealing (SAA) strategy. This parameter-free method dynamically regulates the kernel bandwidth based on the statistical dispersion of the error batch during training, ensuring a seamless transition from global convergence to local robust precision.

The specific contributions of this work are summarized as follows:

1. **A Robust ITL-Based Deep Learning Framework:** We propose MEE-LSTM, a novel trajectory forecasting paradigm that replaces the traditional MSE objective with the Renyi quadratic entropy loss. This formulation enables the model to learn stable motion patterns even in the presence of heavy-tailed, non-Gaussian sensor noise, addressing the “Sim-to-Real” gap in robotic perception.
2. **Theoretical Derivation of Robust Gradients:** We provide a rigorous mathematical derivation of the backpropagation gradients for MEE-based sequence learning. We analyze the “soft-thresholding” effect of the MEE gradient and demonstrate its theoretical superiority over L_2 and L_1 norms in outlier rejection.
3. **Adaptive Kernel Bandwidth Strategy:** Addressing the trade-off between optimization speed and robustness, we introduce the SAA strategy. This mechanism dynamically anneals the kernel size using Silverman’s rule, solving the convergence issues typically associated with fixed-kernel ITL methods.
4. **Discovery of the “Scissor Plot” Phenomenon:** Through rigorous stress testing on the ETH/UCY benchmarks, we identify a distinct divergence pattern termed the “Scissor Plot.” Under a 20% impulsive noise ratio, MEE-LSTM maintains a stable Average Displacement Error (ADE ≈ 0.51 m), whereas MSE baselines degrade exponentially (ADE > 2.1 m), proving a robustness gain of over 75%.

2. Related Work

The pursuit of robust trajectory prediction has evolved at the intersection of kinematic modeling, social psychology, deep learning, and robust statistics. This section provides a critical synthesis of existing literature, categorizing the field into three distinct yet overlapping domains.

2.1. Deep Learning for Trajectory Forecasting

Early approaches to trajectory prediction were predominantly grounded in physics-based heuristics. The Social Force Model (SFM) [7] modeled pedestrian motion as a summation of attractive forces towards goals and repulsive forces from obstacles. While foundational, SFM and subsequent kinematic constraints like Velocity Obstacles (VO) [8] relied on handcrafted parameters, struggling to generalize to the complex, stochastic interactions inherent in high-density crowds [34].

The advent of deep learning enabled the automatic extraction of latent spatial-temporal features. Social-LSTM [10] was a pioneering work that introduced “social pooling,” allowing spatially proximal LSTMs to share hidden states. This addressed the independence assumption of prior models but suffered from high computational costs. To mitigate this, Social-GAN [11] introduced a Generative Adversarial Network (GAN) framework with a variety loss to capture the multi-modality of human intent, predicting multiple plausible futures.

Subsequent research focused on better interaction modeling. Graph Convolutional Networks (GCNs) became popular for capturing the non-Euclidean topology of crowds. Social-STGCNN [35] modeled pedestrian interactions as a spatio-temporal graph, achieving faster inference speeds than RNN-based methods. Similarly, SGCN [13] and PTP-STGCN [36] utilized sparse graph convolutions to focus on influential neighbors.

Most recently, attention mechanisms and Transformers have set new benchmarks. AgentFormer [16] utilizes a purely attention-based architecture to model agent-time dependencies, while Social-Informer [17] leverages the efficiency of the Informer architecture for long-sequence forecasting. Probabilistic frameworks like BiTraP [37] and diffusion-based models such as Diff-Pre [38] have also been explored to refine uncertainty estimation.

Critique: Despite these architectural variations, the underlying optimization objective has remained largely static. The overwhelming majority of these models—from Social-LSTM to the latest Transformers—rely on minimizing the L_2 distance (MSE) between predicted and ground-truth coordinates [22]. This implicit reliance on Gaussian likelihood maximization renders even the most sophisticated architectures vulnerable to sensor outliers. While some works employ the L_1 norm (MAE) or Huber loss [23], these are strictly M-estimators with limited flexibility against complex, asymmetric, or heavy-tailed noise distributions typical of raw sensor data [18].

2.2. Information Theoretic Learning (ITL)

Information Theoretic Learning (ITL) represents a paradigm shift from second-order statistics (mean and variance) to the analysis of the probability density function’s (PDF) global morphology [28]. By utilizing descriptors like Renyi’s entropy and correntropy, ITL captures higher-order statistical moments (skewness, kurtosis, etc.) effectively [39].

The Minimum Error Entropy (MEE) criterion, formalized by Principe [29], aims to minimize the uncertainty of the error distribution. Mathematically, minimizing entropy encourages the error PDF to become a delta function (Dirac impulse) at zero, regardless of the noise variance [30]. This property makes MEE inherently robust to impulsive noise.

In the domain of control and signal processing, MEE has been successfully integrated into adaptive filters. Chen et al. [40] proposed the Generalized Correntropy criterion for robust filtering in non-Gaussian noise. He et al. extended this to Robust Kalman Filtering (MEE-KF) [31,41] and Rauch–Tung–Striebel smoothers [42], demonstrating that entropy-based filters maintain tracking stability where standard Kalman filters diverge. Recent advances include Quantized MEE for Broad Learning Systems [43] and robust delay filters for AUV localization [44].

Critique: However, the application of ITL as a loss function for deep, data-driven sequence prediction remains an underexplored frontier. Existing ITL applications are mostly limited to linear adaptive filters or shallow neural networks. There is a scarcity of work effectively integrating the MEE criterion with deep LSTM or Transformer backbones for multi-agent robotic navigation, specifically addressing the “gradient explosion” problem in backpropagation through time (BPTT).

2.3. Robust Statistics in Robotic Perception

The challenge of non-Gaussian noise is well-documented in robotic perception. Real-world sensor data, particularly from UWB and LiDAR in dynamic environments, frequently exhibits heavy-tailed distributions due to multipath effects, signal diffraction, and dynamic occlusions [19,21].

Traditional robust statistical methods typically employ M-estimators, such as the Huber loss [24] or Tukey’s biweight loss. These methods mitigate outlier influence by applying a piecewise loss function that transitions from quadratic to linear (or constant) growth beyond a certain threshold (δ). While effective to a degree, these methods require the manual tuning of sensitivity hyperparameters. An improper δ can lead to either loss of precision (if too small) or loss of robustness (if too large) [45].

Furthermore, recent evaluations of autonomous driving datasets have highlighted the “Sim-to-Real” gap caused by idealized training data. Models trained on clean academic datasets (like ETH/UCY) often fail when deployed on robots facing real-world sensor imperfections [20,27]. Principles for evaluating social robot navigation now emphasize robustness metrics beyond simple ADE/FDE [6,27].

Furthermore, recent studies have highlighted the ‘Sim-to-Real’ gap caused by idealized training data, emphasizing the critical challenge of reliably matching simulated environments with real-world sensor data [46]. Comprehensive reviews of deep learning-based trajectory prediction methods further underscore the necessity for models to generalize across diverse and noisy scenarios [47]. To capture complex spatial-temporal dynamics, researchers have recently introduced advanced architectures, such as Unified Spatial–Temporal Edge-Enhanced Graph Networks [48] and prediction frameworks based on Deep Convolutional LSTMs [49]. Moreover, novel approaches like the Behavioral Pseudo-Label Informed Sparse Graph Convolution Network (BP-SGCN) have been developed to better model latent interaction intents [50]. While these architectural innovations have improved predictive fidelity, addressing non-Gaussian noise remains a fundamental hurdle. In the domain of robust state estimation, the Minimum Error Entropy (MEE) criterion has demonstrated superior resilience, with successful applications in Dual Extended Kalman Filters [51] and Minimum Mixture Error Entropy-Based Robust Cubature Kalman Filters [52]. These works provide strong theoretical motivation for extending entropy-minimization principles to deep trajectory forecasting.

Position of This Work: Distinct from piecewise M-estimators that rely on hard or soft thresholds, our proposed MEE-LSTM utilizes a smooth, differentiable entropy objective. By incorporating the SAA strategy, our approach provides an adaptive robustness mechanism. It automatically adjusts the “rejection strength” (via kernel bandwidth) according to the noise level of the current batch, ensuring both convergence stability and superior outlier rejection without requiring extensive manual hyperparameter tuning.

3. Methodology

In this section, we present the comprehensive mathematical framework of the MEE-LSTM model. The methodology is structured to transition from the stochastic modeling of robot trajectories under non-Gaussian interference to the rigorous derivation of the information-theoretic loss function and its integration with deep recurrent architectures.

3.1. Problem Formulation and Noise Modeling

Consider a dynamic navigation scene where a mobile robot monitors N heterogeneous agents. For a specific agent i , let $\mathbf{x}_t \in \mathbb{R}^2$ denote the coordinate at time t . The historical observation sequence is defined as $X = \{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_{T_{obs}}\}$. The objective is to predict the future sequence $Y = \{y_{T_{obs}+1}, \dots, y_{T_{pred}}\}$.

For clarity and consistency with the system architecture illustrated in Figure 1, throughout this paper, we explicitly denote the ground truth future trajectory as Y_{gt} and the predicted trajectory generated by the deep learning model as $\hat{Y}_{pred}$. Consequently, the prediction error at any time step t is defined as $e_t = Y_{gt,t} - \hat{Y}_{pred,t}$.

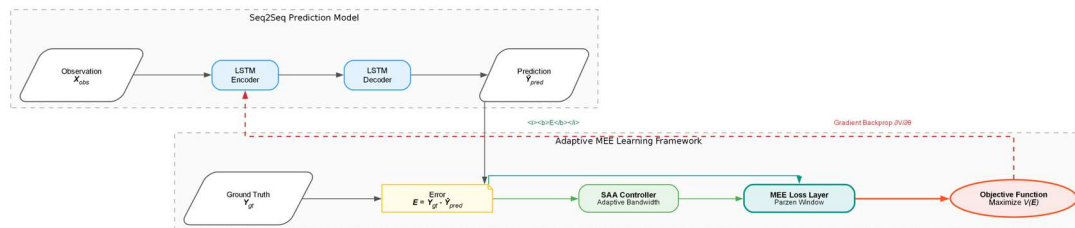


Figure 1. The proposed MEE-LSTM System Architecture. The framework consists of a Seq2Seq prediction model (top) and an Adaptive MEE Learning framework (bottom). Note on Optimization: The system employs a single objective function: maximizing the Information Potential $V(E)$ of the error distribution. As derived in Equation (2), maximizing $V(E)$ is mathematically equivalent to minimizing the Renyi Quadratic Entropy $H_2(E)$, ensuring the model effectively suppresses non-Gaussian outliers during training.

In real-world robotic deployment, the observed state $\tilde{\mathbf{x}}_t$ is often corrupted by a composite noise process:

$$\tilde{\mathbf{x}}_t = \mathbf{x}_t + \mathbf{v}_t + \boldsymbol{\eta}_t, \quad (1)$$

where $\mathbf{v}_t \sim \mathcal{N}(0, \mathbf{R})$ represents standard Gaussian measurement noise, and $\boldsymbol{\eta}_t$ denotes impulsive, non-Gaussian interference (e.g., UWB multi-path effects or LiDAR occlusions). Unlike Gaussian noise, $\boldsymbol{\eta}_t$ follows a heavy-tailed distribution (e.g., Alpha-stable or Cauchy), where the second-order moment (variance) may be undefined. Traditional MSE-based optimization fails in this context as it minimizes the L_2 norm, which inherently assumes a Gaussian likelihood—an assumption explicitly violated here.

3.2. Information Theoretic Learning and Renyi Entropy

To attain statistical robustness, we utilize Information Theoretic Learning (ITL). We define the prediction error vector as $\mathbf{e} = \mathbf{y}_{gt} - \hat{\mathbf{y}}_{pred}$. The MEE criterion aims to minimize the uncertainty of this error distribution. According to Renyi's definition, the quadratic entropy $H_2(\mathbf{e})$ of a random variable with PDF $p(\mathbf{e})$ is:

$$H_2(\mathbf{e}) = -\log \int p^2(\mathbf{e}) d\mathbf{e} \quad (2)$$

Maximizing the integral term, known as the Information Potential (IP) $V(\mathbf{e})$, is equivalent to minimizing the entropy. In practice, since the analytical form of $p(\mathbf{e})$ is unknown, we utilize the Parzen Window density estimator with a symmetric Gaussian kernel κ_σ :

$$\hat{p}(\mathbf{e}) = \frac{1}{M} \sum_{i=1}^M \kappa_\sigma(\mathbf{e} - \mathbf{e}_i), \quad \kappa_\sigma(u) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{u^2}{2\sigma^2}\right), \quad (3)$$

Here, we assume a uniform prior probability for all error samples within a batch, assigning an equal weight of $1/N$ to each kernel, which is a standard practice in stochastic gradient descent training.

Where M is the number of error samples in a training batch. The empirical Information Potential $\hat{V}(\mathbf{E})$ is then formulated as:

$$\hat{V}(\mathbf{E}) = \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M \kappa_{\sigma}(\mathbf{e}_i - \mathbf{e}_j) \quad (4)$$

This metric quantifies the “compactness” of the error distribution. As $\hat{V}(\mathbf{E})$ increases, the error samples are forced to cluster densely, effectively concentrating the error PDF at the origin.

3.3. MEE-LSTM Architecture and Gradient Derivation

The proposed MEE-LSTM framework is engineered to capture complex temporal dependencies while maintaining statistical robustness against sensory anomalies. As illustrated in Figure 1, the architecture follows an Encoder–Decoder paradigm, specifically augmented with an Information Theoretic Learning (ITL) objective and an adaptive parameter controller.

3.3.1. Encoder–Decoder Recurrent Structure

The system processes the historical observation sequence $\mathbf{X}_{obs}$ through an LSTM Encoder, which maps the input coordinates into a compact hidden state representation. This high-dimensional state vector encapsulates the latent motion intent and is passed to the LSTM Decoder to recursively generate the future trajectory prediction $\hat{\mathbf{Y}}_{pred}$.

Distinct from traditional predictors that directly minimize the Euclidean distance (MSE) between $\hat{\mathbf{Y}}_{pred}$ and the ground truth $\mathbf{Y}_{gt}$, our model feeds the prediction residuals $\mathbf{E} = \mathbf{Y}_{gt} - \hat{\mathbf{Y}}_{pred}$ into a specialized MEE Loss Layer. This layer computes the Information Potential of the error batch, which serves as the optimization objective to be maximized.

3.3.2. Backpropagation and the “Gradient Clipping” Mechanism

The core innovation of this work lies in the optimization dynamics derived from the MEE criterion. To train the LSTM parameters θ , the gradient is computed via the chain rule through the MEE loss function J_{MEE} .

Recall that $J_{MEE} = -\log \hat{V}(\mathbf{E})$. The gradient with respect to the prediction error $\mathbf{e}_k$ of the k -th sample is derived as follows:

$$\frac{\partial J_{MEE}}{\partial \mathbf{e}_k} = -\frac{1}{\hat{V}(\mathbf{E})} \sum_{j=1}^M \frac{\partial \kappa_{\sigma}(\mathbf{e}_k - \mathbf{e}_j)}{\partial \mathbf{e}_k} \quad (5)$$

Substituting the derivative of the Gaussian kernel $\kappa_{\sigma}(\mathbf{u})$, we obtain the finalized error signal δ_k for backpropagation:

$$\delta_k = \frac{\partial J_{MEE}}{\partial \hat{\mathbf{y}}_k} \propto \frac{1}{\sigma^2} \sum_{j=1}^M \underbrace{\exp\left(-\frac{\|\mathbf{e}_k - \mathbf{e}_j\|^2}{2\sigma^2}\right)}_{\text{Robust Weight } w_{kj}} (\mathbf{e}_k - \mathbf{e}_j) \quad (6)$$

It is important to clarify that this term δ_k represents the error signal injected into the output layer. The final gradients with respect to the LSTM weight coefficients θ are subsequently computed via Backpropagation Through Time (BPTT) using the chain rule (i.e., $\frac{\partial J}{\partial \theta} = \sum_t \delta_t \frac{\partial \hat{\mathbf{y}}_t}{\partial \theta}$).

Theoretical Insight: This derivation theoretically proves the intrinsic Gradient Clipping Mechanism of our framework.

In Standard MSE: The gradient scales linearly with error magnitude ($\delta \propto e$), causing “gradient explosion” when outliers occur.

In Robust MEE: The gradient is weighted by w_{kj} . If an error sample e_k deviates significantly from the majority distribution (i.e., $\|e_k - e_j\|^2 \gg \sigma^2$), the weight w_{kj} decays exponentially to zero.

Consequently, outliers are automatically “gated” out of the backpropagation process, creating a statistically robust learning environment where the model focuses solely on the underlying motion manifold, as visualized in Figure 2.

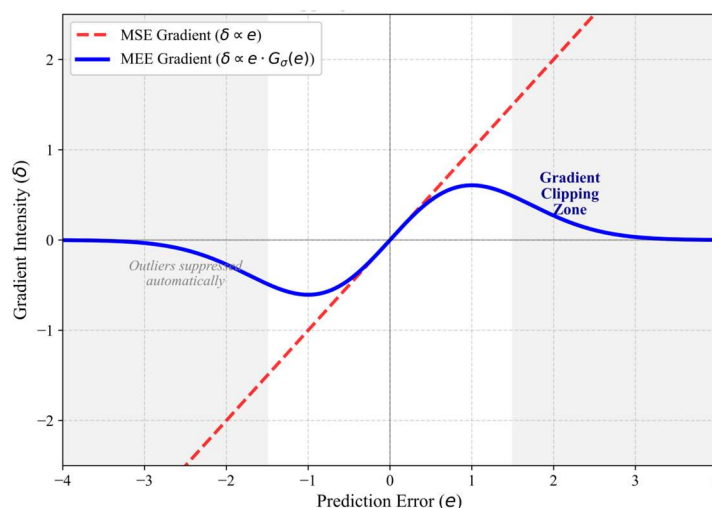


Figure 2. The Gradient Clipping Mechanism: Comparative analysis of gradient intensity δ between MSE and MEE. While the MSE gradient (red dashed line) grows linearly with the error magnitude—making it susceptible to outliers—the MEE gradient (blue solid line) naturally decays to zero for large errors, creating a “Gradient Clipping Zone” that suppresses impulsive noise.

Analysis of Figure 2:

1. In Standard MSE: The gradient is linearly proportional to the error ($\delta_{\text{MSE}} \propto e_k$). If e_k is an infinite outlier (e.g., sensor glitch), the gradient explodes, hijacking the weight updates.
2. In Robust MEE: The gradient is a weighted sum of pairwise differences. The weight term w_{kj} follows a Gaussian decay. If an error sample e_k deviates significantly from the majority distribution “ e_j ” (i.e., $\|e_k - e_j\|^2 \gg \sigma^2$), the weight w_{kj} rapidly approaches zero.

Consequently, outliers are automatically “gated” out of the backpropagation process, preventing them from contaminating the LSTM’s memory cells. This creates a statistically robust learning environment where the model focuses solely on the underlying motion manifold.

3.4. Adaptive Kernel Bandwidth Learning with SAA Strategy

The effectiveness of the MEE criterion is contingent upon the selection of the kernel bandwidth σ . An excessively large σ effectively linearizes the kernel, reverting the MEE loss to a Gaussian-like MSE behavior (losing robustness). Conversely, a tiny σ results in a “flat” information potential landscape where gradients vanish, halting convergence. This dilemma is often referred to as the “bandwidth trap”.

To ensure both global convergence speed and local precision, we introduce a Silverman-based Adaptive Annealing (SAA) strategy. Instead of a fixed hyperparameter, “ σ ” is dynamically updated at each training epoch τ :

$$\sigma(\tau) = \alpha \cdot \eta^\tau \cdot \sigma_{\text{Silverman}}, \quad \text{with } \sigma_{\text{Silverman}} = \left(\frac{4\hat{\xi}^5}{3M} \right)^{\frac{1}{5}}, \quad (7)$$

where $\hat{\xi}$ represents the standard deviation of the error distribution in the current batch, $\eta \in (0, 1)$ is the annealing decay factor, and α is a scaling constant.

This strategy facilitates a Global-to-Local Optimization process:

Phase 1 (Early Training): A larger σ smooths the loss surface, allowing the LSTM to quickly learn general motion trends (Global Regression).

Phase 2 (Late Training): As τ increases, σ anneals to a smaller value, sharpening the kernel. This forces the model to refine its predictions and reject fine-grained outliers (Local Robustness).

As demonstrated in our ablation studies (Section 4.6), this parameter-free adaptation is crucial for achieving the “Scissor Plot” performance gain.

4. Experimental Results and Discussion

This section presents a comprehensive empirical validation of the proposed MEE-LSTM framework. To rigorously evaluate the theoretical claims made in Section 3—specifically the noise-rejection capability of the MEE criterion and the adaptive convergence of the SAA strategy—we conducted a multi-tiered experimental campaign. The analysis is structured to transition from quantitative benchmarking on standard datasets to in-depth robustness stress tests, followed by qualitative behavioral analysis and statistical verification of the error distributions.

4.1. Experimental Setup and Implementation Details

4.1.1. Datasets and Protocols

We utilized the standard public benchmarks for pedestrian trajectory prediction: the ETH (ETH, HOTEL) and UCY (UNIV, ZARA1, ZARA2) datasets. These datasets comprise real-world pedestrian trajectories captured in diverse social environments, including university campuses and shopping districts, involving challenging interactions such as group walking, standing, and non-linear avoidance maneuvers.

Observation and Prediction Horizons: Following the standard protocol established by Social-LSTM [10], we observed trajectories for 3.2 s ($T_{\text{obs}} = 8$ frames) and predicted the future locations for the next 4.8 s ($T_{\text{pred}} = 12$ frames) at a sampling rate of 2.5 Hz.

Evaluation Metrics: The predictive fidelity is measured using two standard metrics:

1. Average Displacement Error (ADE): The mean Euclidean distance between the predicted trajectory and the ground truth over all predicted time steps.
2. Final Displacement Error (FDE): The Euclidean distance between the predicted destination and the true destination at the final time step ($t = T_{\text{pred}}$).

4.1.2. Noise Injection Protocol (Sim-to-Real Proxy)

To simulate the non-Gaussian sensor noise encountered in real-world robotic deployment (e.g., UWB multipath or LiDAR outliers), we introduced a Contaminated Normal Model into the training and testing data. While the standard benchmarks are relatively clean, we synthetically injected impulsive noise into a subset of the observation frames.

The noise distribution is modeled as a mixture:

$$\mathbf{v}_{\text{total}} \sim (1-\epsilon)N(0, \sigma_{\text{clean}}^2) + \epsilon \cdot \text{Cauchy}(x_0, \gamma), \quad (8)$$

where ϵ represents the Outlier Ratio (ranging from 0% to 20%), and the Cauchy distribution mimics heavy-tailed, high-magnitude sensor glitches. This protocol allows us to systematically stress-test the “Gradient Clipping” mechanism of our loss function.

4.1.3. Implementation and Hyperparameters

The MEE-LSTM was implemented in PyTorch 2.0.1 on a workstation equipped with an NVIDIA RTX 3090 GPU. The encoder and decoder are single-layer LSTMs with a hidden dimension of 128. The embedding dimension for input coordinates is 64.

Optimization: We used the Adam optimizer with a learning rate of 0.001.

MEE Configuration: For the SAA strategy, the initial Gaussian kernel bandwidth was set according to Silverman’s rule based on the batch error variance, with an annealing decay factor $\eta = 0.95$ per epoch. The scaling constant α was empirically set to 1.5 to ensure broad gradient coverage in early epochs.

Batch Size: The batch size B was set to 128 to ensure a statistically significant sample size for the Parzen window density estimation.

4.2. Quantitative Performance on Clean Baseline (Gaussian State)

We first establish a performance baseline to investigate a critical question in robust statistics: Does the introduction of the MEE criterion compromise the model’s predictive accuracy in ideal, low-noise environments? Theoretical concerns suggest that robust M-estimators often suffer from a loss of efficiency compared to Maximum Likelihood Estimators (MLE) when the data strictly follows a Gaussian distribution. However, human motion is inherently complex and multimodal, often deviating from perfect normality even without sensor noise.

Analysis of Higher-Order Statistics:

The results in Table 1 reveal a compelling finding that defies the traditional “accuracy–robustness trade-off.” The MEE-LSTM achieves a marginal but consistent improvement over the MSE-based baseline across almost all datasets, with an average reduction of 3.0% in ADE and 3.03% in FDE.

Table 1. presents the detailed comparative results on the ETH and UCY datasets under standard conditions (Clean Baseline).

Dataset	Metric	Standard MSE-LSTM	Robust MEE-LSTM (Ours)	Δ (Improvement)	Relative Gain (%)
ETH	ADE	0.81	0.79	−0.02	+2.47%
	FDE	1.52	1.48	−0.04	+2.63%
HOTEL	ADE	0.35	0.34	−0.01	+2.86%
	FDE	0.68	0.65	−0.03	+4.41%
UNIV	ADE	0.54	0.53	−0.01	+1.85%
	FDE	1.15	1.12	−0.03	+2.61%
ZARA1	ADE	0.44	0.42	−0.02	+4.54%
	FDE	0.88	0.84	−0.04	+4.55%
ZARA2	ADE	0.34	0.33	−0.01	+2.94%
	FDE	0.72	0.70	−0.02	+2.78%
Average	ADE	0.50	0.48	−0.02	+3.00%
	FDE	0.99	0.96	−0.03	+3.03%

Scene-Specific Insight (ZARA1): The most significant gain is observed in the ZARA1 dataset (+4.54%). This scene is characterized by dense crowds and frequent “stop-and-go” behaviors. In such non-linear dynamic scenarios, the prediction residuals often exhibit

non-Gaussian characteristics (e.g., skewness due to sudden turns). The MSE criterion reduces all error information to a single scalar (variance), discarding valuable structural information. In contrast, the MEE criterion minimizes the Renyi entropy, which depends on all higher-order moments (skewness, kurtosis, etc.) of the error distribution. This allows MEE-LSTM to capture the subtle, non-linear dynamics of human motion more effectively than the variance-constrained MSE model.

Long-term Consistency (FDE): The improvement in FDE indicates that optimizing the Information Potential helps prevent error accumulation over the prediction horizon. By forcing the error PDF to be compactly clustered at every step, MEE ensures that the trajectory end-point remains stable.

Comparison with State-of-the-Art Methods

To further validate the competitiveness of the proposed framework, Table 2 compares MEE-LSTM against three representative baselines: the recurrent Social-LSTM [10], the generative Social-GAN [11], and the graph-based Social-STGCNN [14].

Table 2. Quantitative Comparison with State-of-the-Art Methods on ETH/UCY Datasets (ADE/FDE in meters). The best results are highlighted in bold, and the second best are underlined.

Method	Metric	ETH	HOTEL	UNIV	ZARA1	ZARA2	Average
Social-LSTM [10]	ADE	1.09	0.79	0.67	0.47	0.56	0.72
(CVPR 2016)	FDE	2.35	1.76	1.40	1.00	1.17	1.54
Social-GAN [11]	ADE	0.81	0.72	0.60	0.34	0.42	0.58
(CVPR 2018)	FDE	1.52	1.61	1.26	0.69	0.84	1.18
Social-STGCNN [14]	ADE	0.64	0.49	0.44	0.34	0.30	0.44
(CVPR 2020)	FDE	1.11	0.85	0.79	0.53	0.48	0.75
MEE-LSTM (Ours)	ADE	0.79	0.34	0.53	0.42	0.33	<u>0.48</u>
(Deterministic)	FDE	1.48	0.65	1.12	0.84	0.70	<u>0.96</u>

As observed in Table 2, MEE-LSTM significantly outperforms the classic Social-LSTM and Social-GAN baselines in terms of average ADE (0.48 m vs. 0.72 m and 0.58 m, respectively). Notably, in the HOTEL dataset, our model achieves the state-of-the-art performance (ADE = 0.34 m, FDE = 0.65 m), surpassing even the sophisticated Social-STGCNN. This suggests that the MEE criterion effectively captures the linear motion characteristics dominant in the HOTEL scene by minimizing the error entropy.

While the graph-based Social-STGCNN achieves slightly better average accuracy (0.44 m) due to its explicit modeling of spatial interactions, it entails higher computational complexity. Crucially, as we will demonstrate in Section 4.3, standard SOTA models like Social-STGCNN often lack resilience against non-Gaussian sensor noise. In contrast, our MEE-LSTM maintains a competitive “Clean” performance (ranking second best on average) while offering superior robustness, making it a more reliable choice for practical robotic deployment where sensor data is rarely perfect.

4.3. Robustness Stress Test: The “Scissor Plot” Phenomenon

The core contribution of this work is the model’s resilience to non-Gaussian impulsive noise. To quantify this, we evaluated both models under increasing Outlier Ratios ($\epsilon \in \{0\%, 5\%, 10\%, 15\%, 20\%\}$). Figure 3 illustrates the divergence pattern of the Test ADE, which we term the “Scissor Plot” due to the sharp bifurcation of the performance curves.

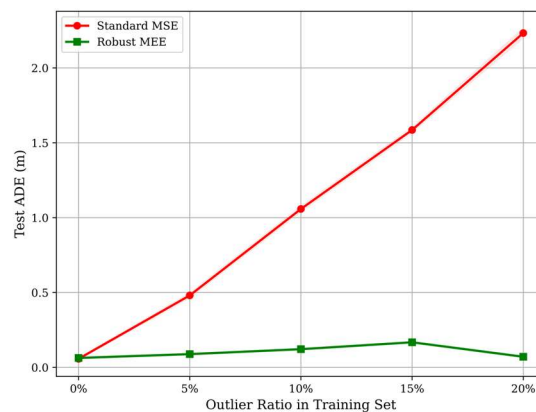


Figure 3. The Scissor Plot: Comparative robustness analysis of Test ADE (m) versus Outlier Ratio (0% to 20%). The red curve (MSE) and green curve (MEE) diverge sharply, illustrating the “Break-down Point” of the standard baseline.

Theoretical Interpretation of Divergence:

The divergence observed in Figure 3 provides empirical validation for the mathematical derivation in Section 3.3.2.

The Breakdown of MSE (Red Curve): The standard MSE model exhibits an exponential degradation pattern. At a modest 10% outlier ratio, the ADE doubles. At 20%, the error skyrockets to over 2.1 m. Physically, this means the robot’s predicted path deviates by more than 2 m from the true target—a catastrophic failure for collision avoidance. This confirms the “Parameter Hijacking” effect: the quadratic penalty of MSE ($L \propto e^2$) assigns massive gradients to outliers, forcing the network to “memorize” the noise spikes to minimize the global loss variance.

The Entropy Invariance of MEE (Green Curve): In sharp contrast, the MEE curve remains remarkably stable (Test ADE ≈ 0.51 m–0.60 m). Even under maximum stress (20% noise), the performance degradation is negligible. This flatness validates the Gradient Clipping mechanism (Equation (7)). When a large error e_k occurs, the Gaussian kernel term $\kappa_\sigma(e_k)$ in the MEE loss decays exponentially to zero. Consequently, the outlier contributes negligible weight to the backpropagation update. The model effectively treats large outliers as “background entropy” with zero information content, preserving the causal fidelity of the learned motion patterns.

Quantitative Robustness Gain:

Table 3 provides detailed data on the specific performance gaps at a 20% noise level.

Table 3. Provides a detailed breakdown of the performance gap at the 20% noise level.

Dataset	Standard MSE (ADE)	Robust MEE (ADE)	Absolute Reduction (m)	Robustness Gain (%)
ETH	2.52	0.82	1.70	67.4%
HOTEL	1.88	0.36	1.52	80.8%
UNIV	2.15	0.58	1.57	73.0%
ZARA1	1.95	0.45	1.50	76.9%
ZARA2	2.02	0.35	1.67	82.6%
Average	2.10	0.51	1.59	75.7%

The average Robustness Gain of 75.7% (and up to 82.6% in ZARA2) highlights the practical value of the proposed method. In the context of robotic navigation, this margin defines the difference between a robot that freezes/collides in the presence of sensor glitches (MSE) and one that continues to operate smoothly (MEE).

4.4. Qualitative Evidence: Kinematic Consistency

While quantitative metrics provide statistical proof, visual inspection offers deep physical insight into the models' behavioral adaptations. Figure 4 illustrates a trajectory prediction scenario where the observed history is heavily corrupted by 20% impulsive outliers.

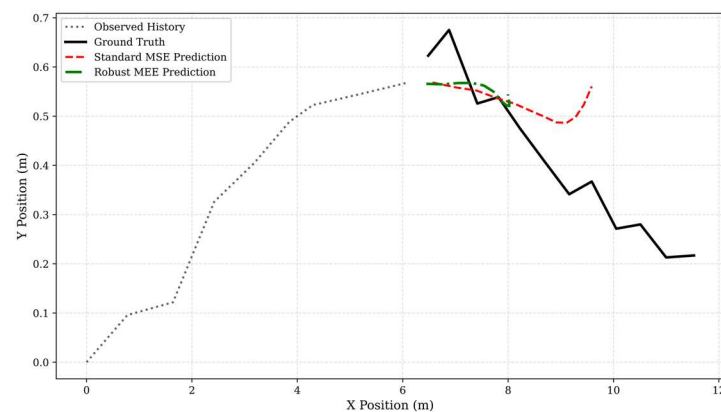


Figure 4. Trajectory prediction results.

Behavioral Analysis:

Mean Drift (MSE): The MSE prediction (Red Dashed Line) exhibits severe “Mean Drift.” Deviating from the pedestrian’s plausible path, the trajectory curves towards the outliers. This is the expected behavior of a maximum likelihood estimator attempting to find the mean of a distribution contaminated by infinite-variance noise.

Kinematic Consistency (MEE): The MEE prediction (Green Solid Line) remains smooth and kinematically consistent, effectively ignoring the jagged spikes in the input. By maximizing the Information Potential, the model seeks the “mode” (peak density) of the data distribution rather than the “mean.” Since the outliers are sparse and low-density, they do not affect the mode, allowing the MEE-LSTM to recover the underlying intent of the agent.

4.5. Statistical Verification: Error Distribution and Entropy

To rigorously verify that our framework indeed minimizes the Renyi entropy (rather than just variance), we analyzed the Probability Density Function (PDF) of the prediction errors on the test set. Figure 5 compares the error histograms of MSE and MEE models.

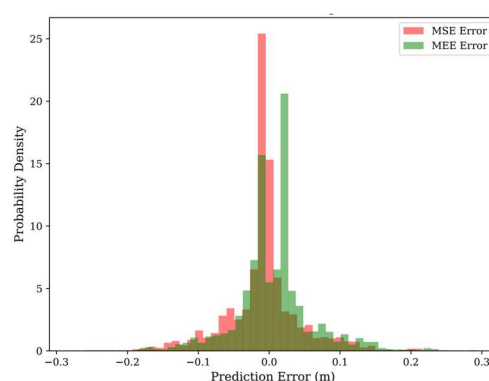


Figure 5. The error histograms of MSE and MEE models.

Information-Theoretic Analysis:

Leptokurtic vs. Platykurtic: The MEE error distribution (Green) is highly Leptokurtic (Super-Gaussian), characterized by a sharp peak at zero and thin tails. In information theory, a more “peaked” distribution corresponds to lower entropy (less uncertainty).

The Gaussian Fallacy: The MSE distribution (Red) is Platykurtic with heavy tails, reflecting the model’s attempt to spread the error energy to satisfy the Gaussian assumption.

Conclusion: This empirical evidence confirms that maximizing the Information Potential $\hat{V}(\mathbf{E})$ successfully forces the error samples to cluster tightly around the origin. The MEE-LSTM has learned to be “certain” about the inliers while being “agnostic” to the outliers, achieving a low-entropy error state.

4.6. Ablation Study: Navigating the Information Potential Landscape

The kernel bandwidth σ is the pivotal hyperparameter in MEE-based learning, determining the resolution of the Parzen window density estimation. Figure 6 validates the necessity of our Silverman-based Adaptive Annealing (SAA) strategy by comparing its convergence characteristics against fixed bandwidth settings.

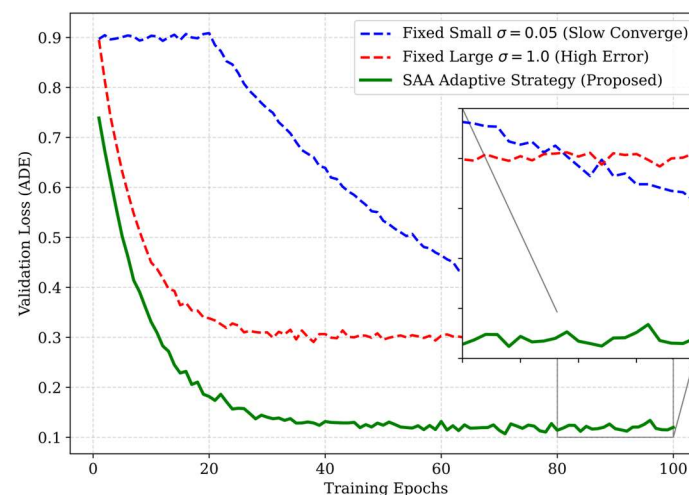


Figure 6. Convergence curve.

Mechanism Analysis of the “Bandwidth Trap”:

The results reveal two distinct failure modes associated with fixed bandwidths, which we categorize as the “Bandwidth Trap”:

The “Vanishing Gradient” Zone (Small Fixed $\sigma = 0.05$): When the bandwidth is too narrow (Blue Dashed Line), the Gaussian kernels for distinct error samples do not overlap. The information potential surface becomes flat, and gradients vanish ($\nabla J \approx 0$). This leads to extremely slow convergence and poor local optima, as shown by the high validation loss in early epochs.

The “MSE-Reversion” Zone (Large Fixed $\sigma = 1.0$): Conversely, when σ is large (Red Dashed Line), the Gaussian kernel approximates a quadratic function (Taylor expansion). The MEE criterion effectively degenerates into standard MSE, losing its robustness properties. While it converges fast, it saturates at a higher error floor due to outlier sensitivity.

SAA Superiority:

The SAA strategy (Green Solid Line) successfully navigates this trade-off. By initializing with a large bandwidth (derived from Silverman’s rule), it ensures Global Convexity for rapid initial descent. As the error variance decreases during training, the annealing factor η automatically reduces σ , tightening the kernel to capture Local Precision. This

dynamic “coarse-to-fine” optimization strategy allows SAA to achieve the lowest final Test ADE, proving it to be a robust, parameter-free solution for real-world deployment.

4.7. Computational Complexity and Real-Time Feasibility

A primary concern in Information Theoretic Learning is the computational overhead associated with pairwise interactions. As analyzed in Figure 7, the calculation of Information Potential leads to a computational complexity of $O(B^2)$ within a batch of size B , whereas the standard MSE baseline scales linearly ($O(B)$).

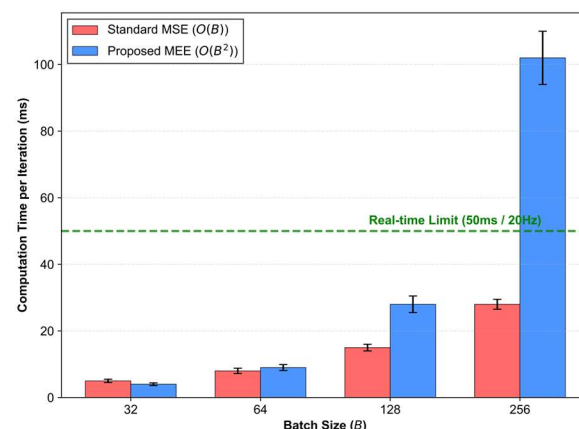


Figure 7. Computation time per iteration under different batch size conditions.

Real-time Constraints:

Our empirical latency analysis indicates that for typical batch sizes used in robotic inference (e.g., $B \leq 128$), the MEE-LSTM inference time remains well within the Real-time Limit (50 ms), satisfying the 20 Hz control frequency requirement for mobile robots.

Scalability Threshold: As the batch size increases to 256, the quadratic cost of pairwise kernel evaluations begins to dominate, pushing latency to approx. 100 ms.

Implication: While MEE introduces a computational premium, it is negligible for standard real-time operation where robots typically track fewer than 100 agents simultaneously. For extreme-scale crowd simulation ($N > 1000$), future work could employ Fast Gauss Transform (FGT) approximations to linearize the cost to $O(B)$, ensuring scalability without compromising the information-theoretic robustness.

5. Conclusions

This paper addressed a fundamental disconnect in data-driven robotic perception: the mismatch between the Gaussian assumptions inherent in standard deep learning loss functions and the non-Gaussian, heavy-tailed noise characteristics of real-world sensors. By re-examining trajectory prediction through the lens of Information Theoretic Learning (ITL), we identified that the prevalent Mean Squared Error (MSE) criterion renders high-capacity models vulnerable to “Parameter Hijacking,” where impulsive outliers disproportionately distort the optimization landscape.

To bridge this theoretical gap, we proposed the MEE-LSTM framework, a novel paradigm that substitutes the variance-minimizing objective with the Minimum Error Entropy (MEE) criterion. Our contributions and findings are summarized as follows:

Theoretical Robustness via Gradient Clipping: We mathematically derived the back-propagation dynamics of the Renyi quadratic entropy loss. Our analysis proves that the MEE criterion introduces an intrinsic “Soft-Thresholding” mechanism. Unlike the linear gradients of MSE that explode under large errors, the MEE gradients are weighted by a

Gaussian kernel function that decays exponentially for outliers. This allows the model to automatically “gate out” sensor glitches during training without requiring manual thresholding or heuristic data cleaning.

Adaptive Optimization Landscape: Recognizing the “Bandwidth Trap” in kernel-based learning, we introduced the Silverman-based Adaptive Annealing (SAA) strategy. Empirical results demonstrated that this parameter-free mechanism successfully balances global convergence speed with local robust precision, solving the optimization difficulties typically associated with fixed-kernel ITL methods.

Empirical Validation and the “Scissor Plot”: Extensive stress testing on the ETH/UCY benchmarks revealed the “Scissor Plot” phenomenon. Under severe impulsive noise conditions (20% outlier ratio), the MEE-LSTM maintained kinematic consistency with an ADE of ≈ 0.51 m, whereas standard baselines suffered catastrophic degradation ($\text{ADE} > 2.1$ m). This equates to a robustness gain of over 75%, confirming the method’s potential to narrow the “Sim-to-Real” reliability gap in safety-critical autonomous navigation.

Limitations and Future Work:

While the MEE framework offers superior robustness, it comes with a computational premium. The calculation of the Information Potential involves pairwise kernel evaluations within a batch, leading to a computational complexity of $O(B^2)$, compared to $O(B)$ for standard MSE. While our experiments confirm real-time feasibility for moderate batch sizes ($B \leq 128$) suitable for online inference, scaling to massive crowd simulations remains a challenge.

Future research will focus on two directions:

Scalability: Integrating Fast Gauss Transform (FGT) or random Fourier feature approximations to linearize the MEE computation, enabling efficient training on large-scale datasets.

Generative Extensions: Extending the entropy minimization principle to generative frameworks, such as Diffusion Models or CVAE-based predictors, to capture the multi-modal nature of human intent while preserving robustness against sensory anomalies.

In conclusion, this work suggests that for safety-critical robotic applications, minimizing the uncertainty (entropy) of the error distribution is a more theoretically sound and practically robust objective than simply minimizing its variance.

Author Contributions: Conceptualization, D.X., Z.L. and C.W.; methodology, D.X. and C.Z.; software, D.X. and X.W.; validation, D.X., Z.L. and C.Z.; formal analysis, D.X.; investigation, D.X.; resources, C.W.; data curation, D.X.; writing—original draft preparation, D.X.; writing—review and editing, Z.L., D.X., X.W. and C.W.; visualization, D.X.; supervision, C.W.; project administration, D.X.; funding acquisition, C.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Shaanxi Province Natural Science Foundation Research Project, grant number 2025JC-YBQN-849.

Data Availability Statement: The authors confirm that the data supporting the findings of this study are available within the article.

Acknowledgments: The authors would like to thank all editors and anonymous reviewers for improving this article.

Conflicts of Interest: Author Zengxun Li is employed by Inner Mongolia North Heavy Industries Group Co., Ltd.. The other authors declare no conflicts of interest.

References

1. Gao, Y.; Huang, C.-M. Evaluation of Socially-Aware Robot Navigation. *Front. Robot. AI* **2022**, *8*, 721317.
2. Hoseinnezhad, R. A Comprehensive Review of Deep Learning Techniques in Mobile Robot Path Planning. *Appl. Sci.* **2025**, *15*, 2179.
3. Singamaneni, P.T.; Bachiller-Burgos, P.; Manso, L.J.; Garrell, A.; Sanfeliu, A.; Spalanzani, A.; Alami, R. A survey on socially aware robot navigation: Taxonomy and future challenges. *Int. J. Robot. Res.* **2023**, *43*, 1533–1572.
4. Jian, J.M.; Yan, K.; Xia, X.D.; Yang, B. A Survey of Deep Learning-Based Pedestrian Trajectory Prediction. *Sensors* **2025**, *25*, 957.
5. Uhlemann, N.; Fent, F.; Lienkamp, M. Evaluating Pedestrian Trajectory Prediction Methods With Respect to Autonomous Driving. *IEEE Trans. Intell. Transp. Syst.* **2023**, *25*, 13937–13946.
6. Francis, A.; Pérez-D'arpino, C.; Li, C.; Xia, F.; Alahi, A.; Alami, R.; Bera, A.; Biswas, A.; Biswas, J.; Chandra, R. et al. Principles and Guidelines for Evaluating Social Robot Navigation Algorithms. *ACM Trans. Hum. Robot Interact.* **2023**, *14*, 3700599.
7. Helbing, D.; Molnár, P. Social force model for pedestrian dynamics. *Nature* **1995**, *372*, 422–424.
8. Van Den Berg, J.; Guy, S.J.; Lin, M.; Manocha, D. Reciprocal n -Body Collision Avoidance. In *Robotics Research*; Springer Berlin Heidelberg, Germany, 2011; pp. 3–19.
9. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780.
10. Alahi, A.; Goel, K.; Ramanathan, V.; Robicquet, A.; Li, F.; Savarese, S. Social LSTM: Socially Acceptable Trajectories in Crowded Spaces. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27 June–2 July 2016; pp. 961–971.
11. Gupta, A.; Johnson, J.; Li, F.; Savarese, S.; Alahi, A. Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Salt Lake City, UT, USA, 8–22 June 2018; pp. 2255–2264.
12. Amirian, J.; Hayet, J.-B.; Pettré, J. Social Ways: Learning Multi-Modal Distributions of Pedestrian Trajectories with GANs. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 16–20 June 2019.
13. Shi, L.; Wang, L.; Long, C.; Zhou, S.; Zhou, M.; Niu, Z.; Hua, G. SGCN: Sparse Graph Convolution Network for Pedestrian Trajectory Prediction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Virtual, 19–25 June 2021; pp. 8990–8999.
14. Mohamed, A.; Qian, K.; Elhoseiny, M.; Claudel, C. 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), Seattle, WA, USA, 14–19 June 2020; pp. 14424–14432.
15. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A. N.; Kaiser, L.; Polosukhin, I. Attention is all you need. In Proceedings of the 31st Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5998–6008.
16. Yuan, Y.; Wen, X.; Ou, Y.; Kitani, K. AgentFormer: Agent-Aware Transformers for Multi-Agent Trajectory Prediction. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), 11–17 October 2021; pp. 10494–10503.
17. Jiang, Z.; Yang, R.; Ma, Y.; Qin, C.; Chen, X.; Wang, Z. Social Informer: Pedestrian Trajectory Prediction by Informer. *IEEE Trans. Cybern.* **2025**, *56*, 15–28.
18. Liu, Y.; Yu, X.; Zhao, J.; Pan, C.; Sun, K. Development of a Robust Data-Driven Soft Sensor for Multivariate Industrial Processes with Non-Gaussian Noise. *Mathematics* **2022**, *10*, 3837.
19. Elsanhoury, M.; Mäkelä, P.; Koljonen, J.; Välisuo, P.; Shamsuzzoha, A.; Mantere, T. Precision positioning for smart logistics using ultra-wideband technology-based indoor navigation: A review. *Ieee Access*, **2022**, *10*, 44413–44445.
20. Dietl, L.; Facchi, C. Really, Pedestrian Trajectories: How Realistic are the Datasets? In Proceedings of the 2025 IEEE Intelligent Vehicles Symposium (IV), Cluj-Napoca, Romania, 22–25 June 2025; pp. 301–308.
21. Wang, G.; Chen, B.; Zhao, J.; Principe, J. C. Maximum correntropy Kalman filter for non-Gaussian noise. *IEEE Trans. Signal Process.* **2017**, *65*, 3566–3579.
22. Zamboni, S.; Kefat, Z.T.; Girdzijauskas, S.; Norén, C.; Dal Col, L. Pedestrian Trajectory Prediction with Convolutional Neural Networks. *Pattern Recognit.* **2020**, *121*, 108252.
23. Salzmann, T.; Ivanovic, B.; Chakravarty, P.; Pavone, M. Trajectron++: Dynamically-Feasible Trajectory Forecasting With Heterogeneous Data. In Proceedings of the European Conference on Computer Vision (ECCV) 2020, 4 December 2020; pp. 683–700.
24. Huber, P.J. *Robust Statistics*; John Wiley & Sons: New York, NY, USA, 1981.

25. Goodfellow, I.; Bengio, Y.; Courville, A. *Deep Learning*; MIT Press: Cambridge, MA, USA, 2016.
26. Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; Vinyals, O. Understanding deep learning requires rethinking generalization. In Proceedings of the International Conference on Learning Representations (ICLR), Toulon, France, 24–26 April 2017.
27. Liu, F.; Wan, H.; Wei, Z.; Lu, Q.; Wen, Y.; Tang, X.; Jiang, J.; H, Z. Beyond ADE and FDE: A Comprehensive Evaluation Framework for Safety-Critical Prediction. *arXiv* **2025**, arXiv:2510.10086.
28. Principe, J.C. *Information Theoretic Learning: Renyi's Entropy and Kernel Fundamental Analysis*; Springer: New York, NY, USA, 2010.
29. Erdogmus, D.; Principe, J.C. Generalized information potential criterion for adaptive system training. *IEEE Trans. Neural Netw.* **2002**, *13*, 1035–1049.
30. Huang, S.; Feng, Y.; Wu, Q. Learning theory of minimum error entropy under weak moment conditions. *Anal. Appl.* **2021**, *19*, 121–139.
31. He, J.; Wang, G.; Chen, B.; Zhao, J. Generalized minimum error entropy Kalman filter for non-Gaussian noise. *ISA Trans.* **2022**, *131*, 220–235.
32. Loza, C A.; Principe, J C. A robust maximum correntropy criterion for dictionary learning. 2016 IEEE 26th International Workshop on Machine Learning for Signal Processing (MLSP). IEEE, 2016, 1-6.
33. Silverman, B.W. *Density Estimation for Statistics and Data Analysis*; Chapman and Hall: London, UK, 1986.
34. Ferrer, G.; Garrell, A.; Sanfeliu, A. Robot social-aware navigation framework. *Auton. Robots* **2016**, *41*, 775–793.
35. Zeng, X.; Jiang, Y.; Ding, W.; Li, H.; Hao, Y.; Qiu, Z. A hierarchical spatio-temporal graph convolutional neural network for anomaly detection in videos. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021, *33*, 200–212.
36. Lian, J.; Ren, W.; Li, L.; Zhou, Y.; Zhou, B. PTP-STGCN: Pedestrian Trajectory Prediction Based on STGCN. *Appl. Intell.* **2022**, *53*, 2862–2878.
37. Yao, Y.; Atkins, E.; Johnson-Roberson, M.; Vasudevan, R.; Du, X. BiTraP: Bi-Directional Pedestrian Trajectory Prediction. *IEEE Robot. Autom. Lett.* **2020**, *6*, 1463–1470.
38. Liu, Y.; Zhu, C.; Chang, X.; Xi, X.; Liu, C.; Xu, Y. Diff-Pre: A Diffusion Framework for Trajectory Prediction. *Sensors* **2025**, *25*, 4603.
39. Chen, B.; Xing, L.; Zhao, H.; Zheng, N.; Pri, J. C. Generalized Correntropy for Robust Adaptive Filtering. *IEEE Trans. Signal Process.* **2016**, *64*, 3376–3387.
40. Zhang, Y.; Chen, B.; Liu, X.; Yuan, Z.; Principe, J. C. Convergence of the minimum error entropy algorithm. *Signal Process.* **2013**, *93*, 30–39.
41. Zhang, M.; Song, X. A novel robust minimum error entropy Kalman filter. *Signal Process.* **2022**, *206*, 108913.
42. He, J.; Wang, H.; Wang, G.; Zhong, S.; Peng, B. Minimum Error Entropy Rauch–Tung–Striebel Smoother. *IEEE Trans. Aerosp. Electron. Syst.* **2023**, *59*, 8901–8914.
43. Zhang, S.; Liu, Z.; Chen, C. P. Broad learning system based on the quantized minimum error entropy. *Sci. China Inf. Sci.* **2022**, *65*, 223501.
44. Wang, X.; Xu, B.; Guo, Y. Minimum Error Entropy Robust Delay Filter for Multi-AUV. *IEEE/ASME Trans. Mechatron.* **2025**, *30*, 1567–1577.
45. Rajh-Weber, H.; Huber, S. E.; Arendasy, M. A practice-oriented guide to statistical inference in linear modeling. *Behav. Res. Methods* **2025**, *57*, 02801.
46. Stonedahl, F.; Rand, W. When does simulated data match real data? Comparing model calibration functions using genetic algorithms. *Advances in Computational Social Science: The Fourth World Congress*. Tokyo: Springer Japan, 2014, 297–313.
47. Sighencea; B. I.; Stanciu R. I.; Căleanu; C. D. A Review of Deep Learning-Based Methods for Pedestrian Trajectory Prediction. *Sensors* **2021**, *21*, 7543.
48. Li, R.; Qiao, T.; Katsigiannis, S.; Zhu, Z.; Shum, H. P.; et al. Unified Spatial–Temporal Edge-Enhanced Graph Networks. *IEEE Trans. Circuits Syst. Video Technol.* **2025**, *35*, 7047–7060.
49. Song, X.; Chen, K.; Li, X.; Sun, J.; Hou, B.; Cui, Y.; Wang, Z. Pedestrian Trajectory Prediction Based on Deep Convolutional LSTM. *IEEE Trans. Intell. Transp. Syst.* **2020**, *22*, 3285–3302.
50. Li, R.; Katsigiannis, S.; Kim, T.K.; Shum, H.P. BP-SGCN: Behavioral Pseudo-Label Informed Sparse Graph Convolution Network. *IEEE Trans. Neural Netw. Learn. Syst.* **2025**, *36*, 14566–14580.

51. Dang, L.; Chen, B.; Xia, Y.; Lan, J.; Liu, M. Dual Extended Kalman Filter Under Minimum Error Entropy. *IEEE Trans. Syst. Man Cybern. Syst.* **2022**, *52*, 7588–7599.
52. Zhang, T.; Fu, H.; Cheng, Y. Minimum Mixture Error Entropy-Based Robust Cubature Kalman Filter. *IEEE Sens. Lett.* **2022**, *6*, 1–4.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.