ISSN 2616-5775 Vol. 9, Issue 1: 33-40, DOI: 10.25236/AJCIS.2026.090104

Intention-Aware Trajectory Prediction with Class-Balanced Learning for Highway Scenarios

Xi Chen[1,a,*]

[1]*School of Transportation and Logistics, Southwest Jiaotong University, Chengdu, 610031, China*
[a]*chenxi61@my.swjtu.edu.cn*
[*]*Corresponding author*

Abstract: *In highway autonomous driving, data-driven trajectory prediction models suffer from Long-tailed Distributions, where straight-driving samples (>90%) dominate the expected gradient (termed Gradient Dominance), suppressing the learning of rare but critical intentions like lane changes. This leads to Intention Collapse, where models default to conservative straight trajectories. We propose an Intention-Aware Class-Balanced Framework to resolve this. Our approach introduces an Intention-Guided Distribution Rebalancing strategy using inverse-frequency weighting to break the gradient dominance, and an Intention-Conditioned Recurrent Decoder that maps discrete intentions to a continuous latent space for controllable generation. Experiments on the HighD dataset show our method reduces the Average Displacement Error (ADE) in safety-critical lane-changing scenarios by 21% (1.15m → 0.91m) compared to the Standard Encoder-Decoder, demonstrating superior robustness in tail events, and validating the efficacy of class-balancing in regression tasks.*

Keywords: *Trajectory Prediction, Long-tailed Distribution, Gradient Dominance, Intention Awareness, Class Balancing*

1. Introduction

Accurate trajectory prediction is vital for autonomous driving safety, serving as the cornerstone for downstream planning and decision-making [1, 2]. However, state-of-the-art models often fail in real-world highway scenarios due to Intention Collapse: they accurately predict majority behaviors (straight driving) but fail to anticipate rare maneuvers (lane changes, braking).

The root cause is Gradient Dominance in Long-tailed Distributions, a critical challenge in predictive modeling [3]. Since straight samples dominate (>90%), the expected gradient $\mathbb{E}[\nabla_\theta \mathcal{L}]$ is overwhelmed by the majority class. Consequently, the model actively suppresses responses to rare intentions to minimize global loss, collapsing complex interaction patterns into a single conservative mode.

To address this, we contend that merely scaling up model capacity (e.g., using large Transformers) does not resolve the underlying data imbalance; instead, the gradient dominance must be fundamentally addressed at the optimization level. We propose an Intention-Aware Class-Balanced Framework (Figure 1) with three contributions:

(1) Intention-Guided Rebalancing: We apply inverse-frequency weighting to sample and loss functions, restoring gradient contribution for rare classes.

(2) Intention-Conditioned Decoder: We design a decoder that uses intention as an explicit condition, maintaining independent feature subspaces for different maneuvers.

(3) Optimization-Centric Insight: We demonstrate that mitigating gradient dominance via rebalancing is more effective than structural complexity for restoring tail performance, offering a model-agnostic solution for long-tailed regression.

ISSN 2616-5775 Vol. 9, Issue 1: 33-40, DOI: 10.25236/AJCIS.2026.090104

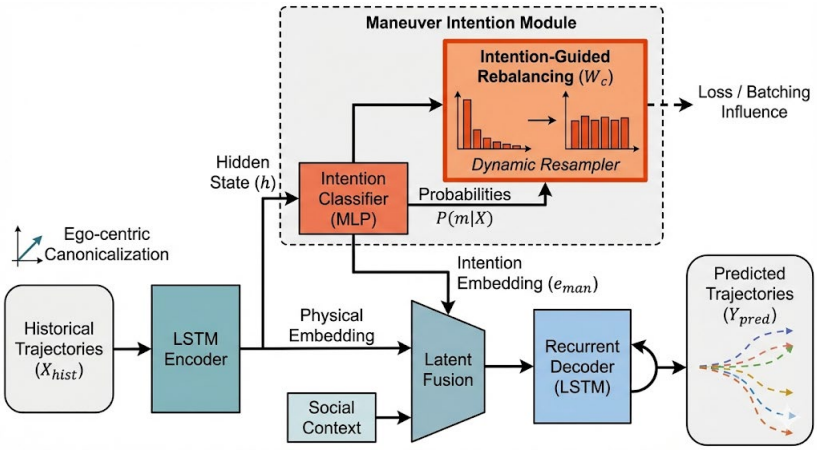

Figure 1: Overview of the Framework.

The system comprises an Ego-centric Encoder, an Intention-Guided Rebalancing Module (highlighted), and an Intention-Conditioned Recurrent Decoder.

2. Related Work

Deep Learning Approaches: While early methods relied on kinematic rules like IDM [4], they struggle in complex interactions. Current state-of-the-art data-driven models, ranging from LSTM-based encoders [5] to Transformers (HiVT [6], Wayformer [7], Trajectron++ [8]) and graph-based representations (LaneGCN [9], LAformer [10]), excel in spatial-temporal feature extraction. However, they typically optimize a global Euclidean loss (e.g., L2). In long-tailed datasets, this objective inherently drives the model to fit the majority class to minimize total error, leading to feature representation collapse.

Modeling Uncertainty vs. Imbalance: To mitigate deterministic mode collapse, Generative models, such as GANs (Social-GAN [11]) and recent diffusion approaches [12], introduce latent noise, while Anchor-based methods (TNT [13], MultiPath++ [14]) utilize pre-defined trajectory templates. While improving output diversity, they do not address the underlying training data imbalance. Generative priors learned from skewed data still collapse to the majority mode, and anchor classifiers remain susceptible to gradient dominance. Unlike these methods which separate diversity from imbalance, we adapt Class-Balanced strategies [15, 16] to directly tackle the root cause—sample scarcity—within the regression framework.

3. Problem Formulation

We predict the future trajectory $Y_i = s_i^{t+1}, \ldots, s_i^{t+T}$ of a target vehicle i given its history X_i and neighbors. Driving intention m is defined as $m \in$ Keep Lane, Left/Right LC, Accel, Decel.

Label Generation (Onset-Detection Protocol): To maintain train-test consistency, we define intention labels using an onset-detection approach rather than relying on the full prediction horizon. Given an observed trajectory ending at t_{obs}, we compute the maximum lateral deviation within a short onset window:

$$\Delta y_{onset} = \max_{t \in [t_{obs}, t_{obs}+0.5s]} |y_t - y_{t_{obs}}| \quad (1)$$

We use a 0.5s onset window (10% of the total 5s horizon) rather than the full prediction window. This captures the initiation phase of maneuvers, ensuring labels reflect early behavioral signals observable from recent trajectory dynamics, rather than completed outcomes that would create train-test mismatch.

Labeling Rules (Prioritized): We assign intention labels as follows: (1) Lane Change if the onset lateral deviation $\Delta y_{onset} > 0.5$ m within the 0.5 s window $[t, t + 0.5s]$; (2) Speed Change if the absolute longitudinal acceleration $|a_x| > 0.3$ m/s² averaged over $[t - 0.5s, t]$; (3) Keep Lane otherwise. The 0.5 m threshold captures early lane-change drift, and the hierarchy (LC > Speed Change > Keep) resolves conflicts; temporal smoothing (τ=3) filters noise.

ISSN 2616-5775 Vol. 9, Issue 1: 33-40, DOI: 10.25236/AJCIS.2026.090104

From Global to Conditional Regression: Traditional models minimize global loss $\mathcal{L} = ||Y - \hat{Y}||^2$, which leads to mode collapse on imbalanced data. We instead learn a conditional distribution $P(Y|X, m)$, decomposing the global problem into K local sub-problems: $\hat{Y}^* = \text{argmax}_{Y \in \mathcal{T}_m} P(Y|X, m)$, where $\mathcal{T}_m$ is the behavior subspace for intention m.

4. Methodology

4.1. Ego-centric Coordinate System

We construct a dynamic ego-centric system where the ego vehicle's current position p_t is the origin. Historical positions are converted to relative displacements $\hat{p}_{t-k} = p_{t-k} - p_t$. This Input Canonicalization ensures spatial translation invariance.

4.2. Intention-Conditioned Recurrent Decoder

To prevent intention forgetting, we inject the intention signal into every decoding step.

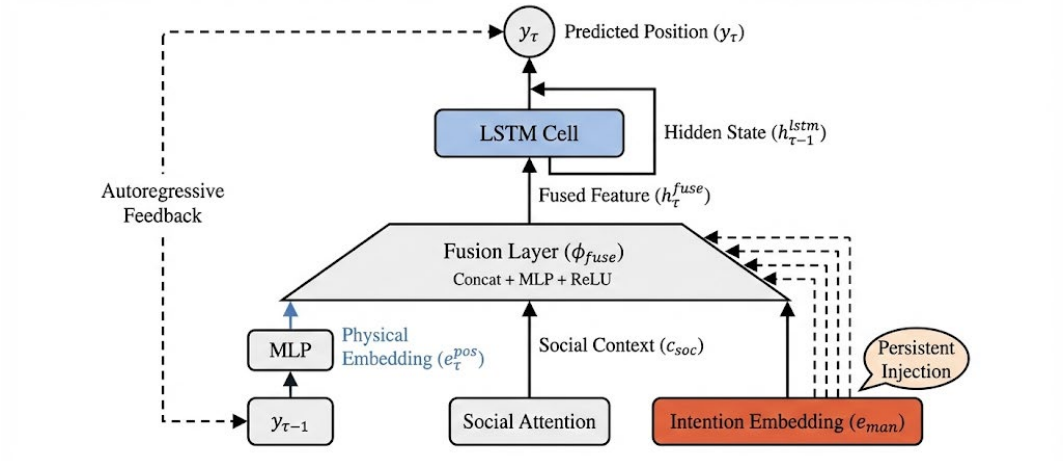

Figure 2: Intention-Conditioned Recurrent Decoder.

A Fusion Layer projects physical states, social context, and intention embeddings into a unified latent space for the LSTM generator. To prevent intention forgetting, we inject the intention signal into every decoding step (Figure 2).

Latent Fusion: At each step τ, we fuse the physical embedding e_τ^{pos}, social context c_{soc} (via Multi-Head Attention), and intention embedding e^{man} into a latent feature h_τ^{fuse}. This is fed into the LSTM: $h_\tau^{lstm} = \text{LSTM}\left(h_\tau^{fuse}, h_{\tau-1}^{lstm}\right)$. This ensures the "Lane Change" command persists throughout the horizon. During inference, we employ Confidence-Weighted Decoding: $e_{robust}^{man} = \sum_{m \in M} P(m|X) \cdot e_m^{man}$, softly interpolating intentions to smooth predictions in ambiguous scenarios.

4.3. Hierarchical Sampling

For each input we generate K = 6 trajectories. The three most probable intentions (sorted by the intention classifier) receive a fixed allocation of $3, 2, 1$ samples respectively. If fewer than three intentions have non-zero probability, the remaining slots are assigned to the highest-probability intention. Different trajectories for the same intention are obtained by sampling independent latent noise vectors, guaranteeing multimodality while keeping the majority of hypotheses focused on the most likely maneuver.

4.4. Intention-Guided Distribution Rebalancing

To mitigate long-tailed intention imbalance, we propose a dual-strategy rebalancing mechanism combining dynamic resampling and loss reweighting, targeting regression mode collapse.

ISSN 2616-5775 Vol. 9, Issue 1: 33-40, DOI: 10.25236/AJCIS.2026.090104

4.4.1. Intention-Driven Dynamic Resampling

Building upon the inverse frequency principle [15], we calculate the sampling weight W_c for each intention class c:

$$W_c = \frac{1}{\sqrt{N_c} + \epsilon} \tag{2}$$

where N_c is the total sample count for class c. We construct a Weighted Random Sampler S. In each training batch, samples are drawn according to probability $P(i) \propto W_{m_i}$. This compels the model to see a balanced ratio of straight and lane-change maneuvers, preventing the gradient updates from being dominated by the majority class.

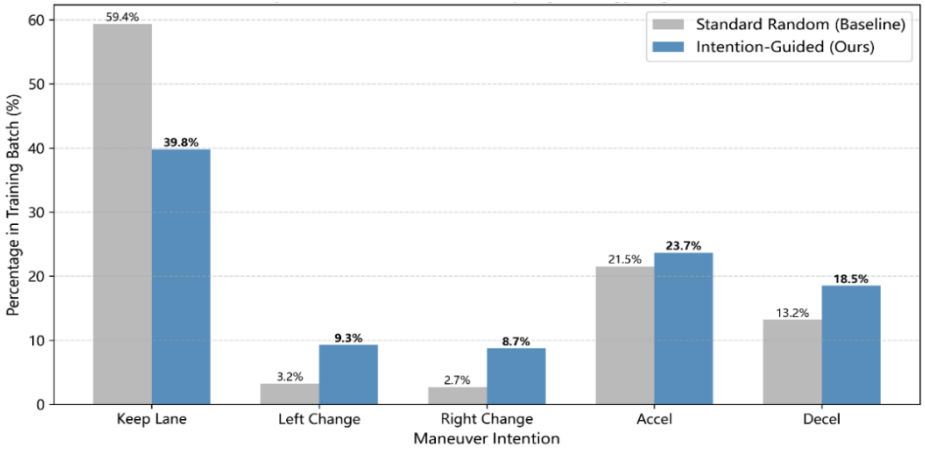

Figure 3: Impact of Class-Balanced Sampling Strategy.

Standard random sampling (Grey) results in straight-driving samples dominating the batch (>59%). Our Intention-Guided strategy (Blue) significantly boosts the presence of Lane Change (LC) classes (from ~3% to ~9%), effectively mitigating gradient dominance.

As shown in Figure 3, our sampler increases the exposure of lane-change samples by roughly 3× while still retaining a substantial portion of Keep Lane samples, preventing over-correction and preserving distribution realism.

4.4.2. Loss Reweighting

Complementarily, we apply weights to the classification loss:

$$L_{cls} = -\sum_{c=0}^{K-1} \alpha_c \, y_c \log(\hat{y}_c) \tag{3}$$

where α_c functions similarly to Focal Loss [17], penalizing misclassifications of rare intentions more heavily.

4.5. Intention-Conditioned Recurrent Decoding

Latent Fusion: During inference, we employ Confidence-Weighted Decoding or Hierarchical Sampling:

To cover multimodal uncertainties, we generate $K = 6$ hypotheses using a Hierarchical Sampling Strategy. We allocate samples based on intention probabilities (e.g., 3 samples for top-1, 2 for top-2), ensuring coverage of both the most likely maneuver and potential alternatives. e_{robust}^{man} is used for trajectory smoothing.

4.6. Training Objective

The multi-task loss is defined as:

$$\mathcal{L}_{total} = \lambda_{reg}\mathcal{L}_{ADE} + \lambda_{cls}\mathcal{L}_{intention} + \lambda_{phy}\mathcal{L}_{kinematics} \tag{4}$$

ISSN 2616-5775 Vol. 9, Issue 1: 33-40, DOI: 10.25236/AJCIS.2026.090104

where $\mathcal{L}_{kinematics} = ||v - \hat{v}||^2 + ||a - \hat{a}||^2$ enforces physical consistency.

Hyperparameters and Experimental Setup:

To ensure reproducibility, we detail the network configuration:

Encoder: 2-layer LSTM with Hidden Size=128. Input embedding dimension is 32.

Social Attention: Multi-Head Attention (Heads=4), Key/Value dimension 64.

Decoder: Single-layer LSTM, Hidden Size=128.

Training Strategy: Batch Size=64, using AdamW optimizer ($lr = 1e^{-3}$) with ReduceLROnPlateau scheduler (Patience=5, Factor=0.5). To prevent gradient explosion, Norm Clip=1.0 is applied.

Loss Weights: Through grid search, $\lambda_{reg} = 1.0$, $\lambda_{cls} = 0.5$, $\lambda_{phy} = 0.1$ yields the best balance.

5. Experiments

We evaluated on the HighD Dataset [18](downsampled to 5Hz, 3s history, 5s prediction). The reported metrics are ADE, FDE and LC-ADE, where LC-ADE is the average displacement error over the full 5-second horizon for all samples whose ground-truth includes a lane-change (left or right); when a model produces multiple hypotheses we take the minimum ADE among the K samples, and if a lane-change is mis-predicted as Keep-Lane the error of that (incorrect) trajectory is still counted. For multimodal methods we use K = 6 hypotheses and allocate them hierarchically (3/2/1)—three samples for the most likely intention, two for the second-most likely, and one for the third—to balance diversity and latency, while deterministic baselines are evaluated with K = 1 (so min-ADE/min-FDE reduce to the single prediction). All results are presented as mean ± standard deviation.

5.1. Quantitative Analysis

Table 1 mainly compares our method against representative paradigms: Physics-driven (CS-LSTM), Generative (Social-GAN), and Memory-based (MANTRA). Our Hybrid Strategy outperforms baselines, achieving 0.85m ADE and 1.88m FDE. Crucially, the improvement stems from resolving the long-tail bias rather than backbone complexity.

Table 1: Performance Comparison (Horizon=5s).

Paradigm	K	ADE (m)	FDE (m)
CS-LSTM [19]	6	1.12	2.85
Social-GAN [4]	6	1.05	2.62
MANTRA [20]	6	0.92	2.15
Ours (Hybrid)	6	0.85	1.88

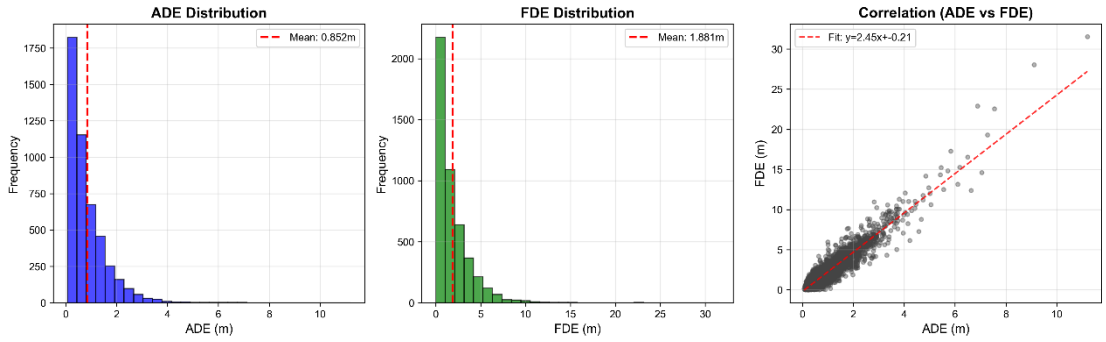

Figure 4: Trajectory Error Analysis.

As shown in Figure 4, the error histograms demonstrate that our model achieves tight convergence, with the vast majority of predictions concentrated in the low-error region (ADE < 2m, FDE < 5m). The mean ADE of 0.852m and mean FDE of 1.881m confirm the model's overall precision. Furthermore, the scatter plot (right) reveals a strong linear correlation ($R^2 > 0.9$) between ADE and FDE, indicating consistent prediction stability over the 5s horizon without significant divergence in long-term forecasting.

We intentionally chose these foundational baselines to isolate the impact of our Class-Balanced

ISSN 2616-5775 Vol. 9, Issue 1: 33-40, DOI: 10.25236/AJCIS.2026.090104

Strategy. While recent Transformer-based models [6, 21] achieve high performance via structural complexity, our goal is to demonstrate that distribution rebalancing alone can yield significant gains (e.g., 21% reduction in LC-ADE) even on lightweight backbones.

5.2. Ablation Study

We validate the contribution of each module in Table 2.

Baseline: Standard Encoder-Decoder. suffers in LC scenarios (1.15m LC-ADE).

Ours (Hybrid): Adding both Balancing and Intention Fusion reduces LC-ADE to 0.82m, confirming that rebalancing gradients is essential for learning tail dynamics. The significant jump (Variant 4 vs. 2/3) suggests a synergistic effect, where rebalancing unlocks the potential of intention conditioning by providing properly distributed training signals.

Table 2: Module Contribution.

Model Variant	k	Resampling	Intention Fusion	ADE (m)	LC-ADE (m)	Balanced Acc
Baseline	6	-	-	1.05	1.15	62.1%
Only Balancing	6	✓	-	0.94	1.05	78.6%
Only Intention	6	-	✓	0.92	0.95	65.4%
Ours (Hybrid)	6	✓	✓	0.85	0.91	81.5%

5.3. Intention Classification Performance

To further validate that our model effectively mitigates "Intention Collapse," we present the detailed classification metrics in Table 3.

A direct consequence of gradient dominance in previous works is the extremely low recall for tail classes (e.g., Lane Change). In contrast, our method achieves 91.7% Recall for Left Lane Change and 84.0% for Right Lane Change, proving that the rebalancing strategy successfully forces the model to learn these rare but critical distinctive features.

Table 3: Intention Classification Performance

Intention Class	Precision	Recall	F1-Score	Support
Keep Lane	0.957	0.744	0.837	2951
Left Lane Change	0.816	0.917	0.864	145
Right Lane Change	0.817	0.840	0.828	106
Accelerate	0.640	0.912	0.752	1008
Decelerate	0.63	0.870	0.734	562
Weighted Avg	0.845	0.802	0.808	4772

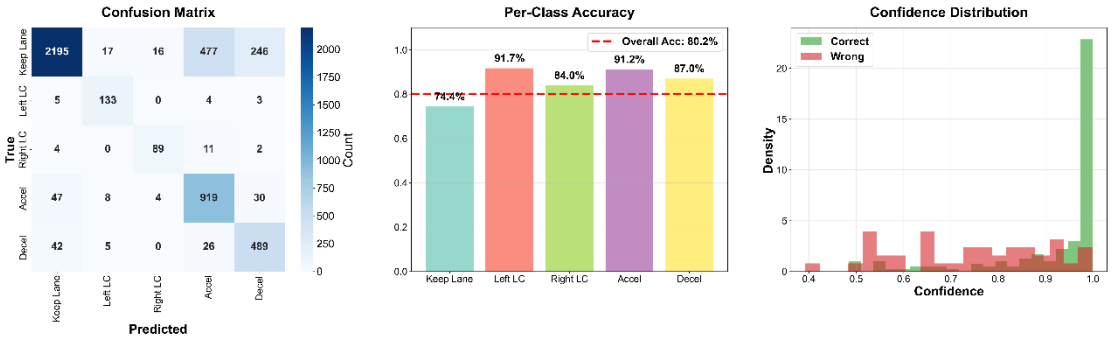

Figure 5: Maneuver Classification Analysis.

Left: confusion matrix over five maneuver classes (Keep Lane, Left Lane Change, Right Lane Change, Accel, Decel).

Middle: per-class accuracy with overall accuracy shown as a dashed line (80.2%).

Right: confidence distribution for correct vs. incorrect predictions. Correct predictions concentrate

ISSN 2616-5775 Vol. 9, Issue 1: 33-40, DOI: 10.25236/AJCIS.2026.090104

near high confidence (≈0.95–1.0), whereas misclassified samples exhibit significantly lower and more dispersed confidence, indicating that the classifier's probability estimates can be used as a reliability signal for downstream trajectory fusion.

As visualized in the Confusion Matrix (Figure 5), we observe that Keep Lane recall is lower than lane-change classes, mainly due to confusion with subtle accel/decel patterns. Therefore, we do not rely on hard intention decisions; instead, we leverage the confidence-weighted decoding to down-weight low-confidence predictions, improving robustness under ambiguous longitudinal behaviors.

5.4. Qualitative Results

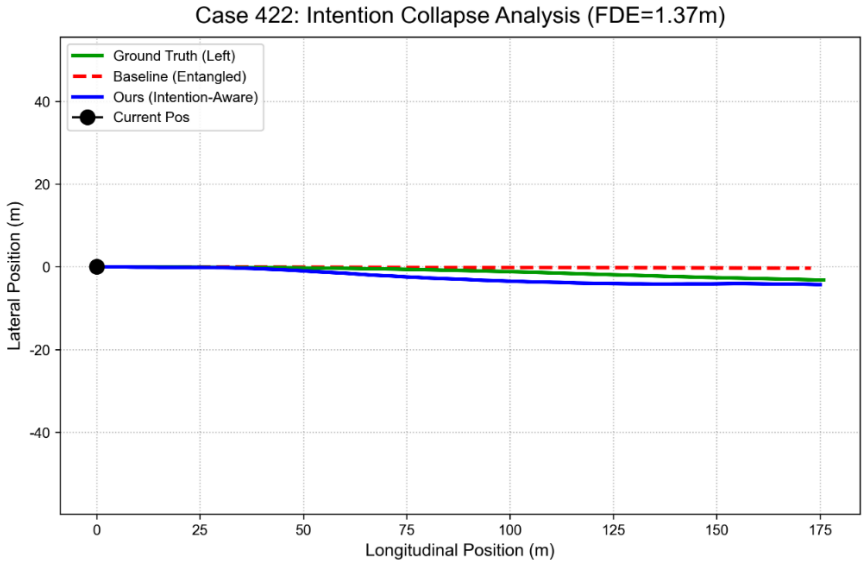

Figure 6: Intention Awareness vs. Baseline.

Baseline (Red) predicts straight due to intention collapse. Ours (Blue) correctly identifies the early cut-in maneuver.

As illustrated in Figure 6, we analyze a scenario demonstrating the "Intention Collapse" phenomenon.

Scenario Description: The target vehicle initiates a Left Lane Change (indicated by the Green Ground Truth line) with a distinct lateral displacement. The Baseline model (Red), dominated by the majority straight-driving distribution, fails to diverge from the lane center, exhibiting a typical "regression to the mean" behavior.

Intention-Aware Prediction: Conversely, our Intention-Aware model (Blue) successfully detects the onset of the Left Lane Change maneuver. Despite the strong prior for going straight, the intention-conditioned decoder leverages the detected signal to effectively steer the predicted trajectory towards the target lane (Left), achieving a significantly lower Final Displacement Error (FDE = 1.37m) compared to the Baseline.

Consistency Analysis: The result highlights the framework's ability to maintain consistency between the predicted intention and the executed trajectory. Unlike the Baseline which remains "entangled" in the straight mode, our model's generation is explicitly conditioned on the detected Left intention, ensuring robustness even in long-tail cut-in scenarios.

6. Conclusion

In this work, we demonstrated that optimization-level rebalancing is a potent alternative to increasing model complexity for long-tailed trajectory prediction. By effectively countering gradient dominance, our framework achieves a 91.7% recall in lane-change detection, proving that 'tail' safety is recoverable even with lightweight backbones.

ISSN 2616-5775 Vol. 9, Issue 1: 33-40, DOI: 10.25236/AJCIS.2026.090104

References

[1] Rudenko, A., Palmieri, L., Herman, M., Kitani, K.M., Gavrila, D.M. and Arras, K.O. (2020) Human Motion Trajectory Prediction: A Survey. The International Journal of Robotics Research, 39, 895-935.
[2] Mozaffari, S., Al-Jarrah, O.Y., Dianati, M., Jennings, P. and Mouzakitis, A. (2020) Deep Learning-Based Vehicle Behavior Prediction for Autonomous Driving Applications: A Review. IEEE Transactions on Intelligent Transportation Systems, 23, 33-47.
[3] Makansi, O., Çiçek, Ö., Marrakchi, Y. and Brox, T. (2021) On Exposing the Challenging Long Tail in Future Prediction of Traffic Actors. IEEE/CVF International Conference on Computer Vision (ICCV), 13147-13157.
[4] Treiber, M., Hennecke, A. and Helbing, D. (2000) Congested Traffic States in Empirical Observations and Microscopic Simulations. Physical Review E, 62, 1805-1824.
[5] Alahi, A., Goel, K., Ramanathan, V., Robicquet, A., Fei-Fei, L. and Savarese, S. (2016) Social LSTM: Human Trajectory Prediction in Crowded Spaces. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 961-971.
[6] Zhou, Z., Ye, L., Wang, J., Wu, K. and Lu, K. (2022) HiVT: Hierarchical Vector Transformer for Multi-Agent Motion Prediction. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 8823-8833.
[7] Nayakanti, N., Al-Rfou, R., Zhou, A., Goel, K., Refaat, K.S. and Sapp, B. (2022) Wayformer: Motion Forecasting via Simple & Efficient Attention Networks. International Conference on Robotics and Automation (ICRA), 2980-2987.
[8] Salzmann, T., Ivanovic, B., Chakravarty, P. and Pavone, M. (2020) Trajectron++: Dynamically-Feasible Trajectory Forecasting with Heterogeneous Data. European Conference on Computer Vision (ECCV), 683-700.
[9] Liang, M., Yang, B., Hu, R., Chen, Y., Liao, R., Feng, S. and Urtasun, R. (2020) Learning Lane Graph Representations for Motion Forecasting. European Conference on Computer Vision (ECCV), 541-556.
[10] Liu, M., et al. (2024) LAformer: Trajectory Prediction for Autonomous Driving with Lane-Aware Scene Constraints. IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2039-2049.
[11] Gupta, A., Johnson, J., Fei-Fei, L., Savarese, S. and Alahi, A. (2018) Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2255-2264.
[12] Bronars, M., Cheng, S. and Xu, D. (2024) Legibility Diffuser: Offline Imitation for Intent Expressive Motion. IEEE Robotics and Automation Letters, 9, 10161-10168.
[13] Zhao, H., Gao, J., Lan, T., Sun, C., Sapp, B., Varadarajan, B., Shen, Y., Shen, Y., Chai, Y., Schmid, C. and Li, C. (2020) TNT: Target-Driven Trajectory Prediction. Conference on Robot Learning (CoRL), 1359-1368.
[14] Varadarajan, B., Hefny, A., Srivastava, A., Kshetramade, K.S., Voelz, J., Covert, J., Kim, E., Hu, C., Bronstein, A. and Anguelov, D. (2022) MultiPath++: Efficient Information Fusion and Trajectory Aggregation for Behavior Prediction. International Conference on Robotics and Automation (ICRA), 7814-7821.
[15] Cui, Y., Jia, M., Lin, T.Y., Song, Y. and Belongie, S. (2019) Class-Balanced Loss Based on Effective Number of Samples. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 9268-9277.
[16] Kang, B., Xie, S., Rohrbach, M., Yan, Z., Gordo, A., Feng, J. and Kalantidis, Y. (2020) Decoupling Representation and Classifier for Long-Tailed Recognition. International Conference on Learning Representations (ICLR).
[17] Lin, T.Y., Goyal, P., Girshick, R., He, K. and Dollár, P. (2017) Focal Loss for Dense Object Detection. IEEE International Conference on Computer Vision (ICCV), 2980-2988.
[18] Krajewski, R., Bock, J., Kloeker, L. and Eckstein, L. (2018) The HighD Dataset: A Drone Dataset of Naturalistic Vehicle Trajectories on German Highways for Validation of Highly Automated Driving Systems. 21st International Conference on Intelligent Transportation Systems (ITSC), 2118-2125.
[19] Deo, N. and Trivedi, M.M. (2018) Convolutional Social Pooling for Vehicle Trajectory Prediction. IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 1549-1558.
[20] Marchetti, F., Becattini, F., Seidenari, L. and Del Bimbo, A. (2020) MANTRA: Memory Augmented Networks for Multiple Trajectory Prediction. IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7141-7150.
[21] Gu, J., Sun, C. and Zhao, H. (2021) DenseTNT: End-to-End Trajectory Prediction from Dense Goal Sets. IEEE/CVF International Conference on Computer Vision (ICCV), 15303-15312.