

AMD: Adaptive Momentum and Decoupled Contrastive Learning Framework for Robust Long-Tail Trajectory Prediction

Bin Rao^{1*} Haicheng Liao^{1*} Yanchen Guan¹ Chengyue Wang¹ Bonan Wang¹
Jiaxun Zhang¹ Zhenning Li^{1†}

¹State Key Laboratory of Internet of Things for Smart City, University of Macau

{yc57416, yc27979, yc37976, yc47938, mc35002, yc47415, zhenningli}@um.edu.mo

Abstract

Accurately predicting the future trajectories of traffic agents is essential in autonomous driving. However, due to the inherent imbalance in trajectory distributions, tail data in natural datasets often represents more complex and hazardous scenarios. Existing studies typically rely solely on a base model’s prediction error, without considering the diversity and uncertainty of long-tail trajectory patterns. We propose an adaptive momentum and decoupled contrastive learning framework (AMD), which integrates unsupervised and supervised contrastive learning strategies. By leveraging an improved momentum contrast learning (MoCo-DT) and decoupled contrastive learning (DCL) module, our framework enhances the model’s ability to recognize rare and complex trajectories. Additionally, we design four types of trajectory random augmentation methods and introduce an online iterative clustering strategy, allowing the model to dynamically update pseudo-labels and better adapt to the distributional shifts in long-tail data. We propose three different criteria to define long-tail trajectories and conduct extensive comparative experiments on the nuScenes and ETH/UCY datasets. The results show that AMD not only achieves optimal performance in long-tail trajectory prediction but also demonstrates outstanding overall prediction accuracy.

1. Introduction

Achieving high-level autonomous driving relies heavily on the ability to accurately predict the future trajectories of surrounding traffic agents [26, 45]. Precise trajectory prediction enables autonomous vehicles to make informed decisions, ensuring safety and efficiency in complex traffic environments. Despite significant advancements, autonomous systems still face challenges in handling the vast diversity of

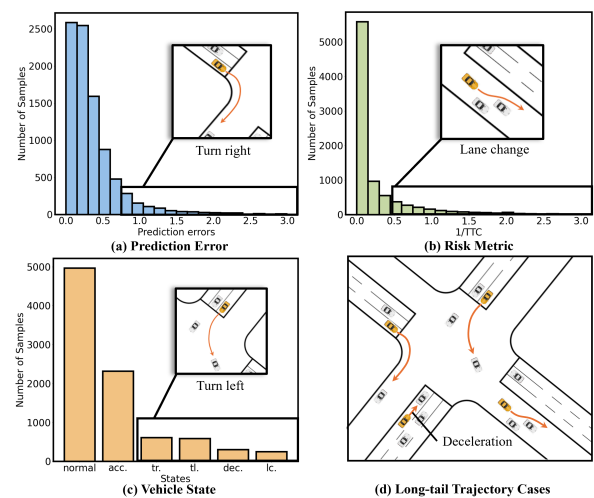


Figure 1. Long-tail trajectory distributions defined from multiple perspectives. Panels (a), (b), and (c) illustrate distributions based on prediction error, the risk metric (inverse time-to-collision, 1/TTC), and vehicle state, respectively. Panel (d) presents vehicle trajectories under various scenarios—such as turning, lane changing, acceleration (acc.), and deceleration (dec.) maneuvers—offering a visual representation to facilitate understanding of long-tail trajectories.

driving behaviors exhibited in real-world scenarios.

One critical challenge arises from the **long-tail distribution** of driving behaviors. In real-world traffic, a small number of common behaviors—such as steady cruising or standard lane changes—occur frequently and dominate datasets. In contrast, numerous rare but potentially hazardous behaviors, like abrupt maneuvers or interactions with erratic agents, are underrepresented. This imbalance poses significant difficulties for trajectory prediction models, which tend to perform well on frequent behaviors but struggle with rare ones. Addressing this issue is essential for the safe and reliable operation of autonomous vehicles.

To advance the field, it is crucial to address several fun-

*Equal contribution.

†Corresponding author: zhenningli@um.edu.mo

damental questions that have been inadequately explored:

Q1: What exactly constitutes the long-tail challenge in trajectory prediction? While data imbalance is a recognized issue, prior studies often lack a precise definition of the long-tail challenge in the context of trajectory prediction for autonomous driving. The long-tail phenomenon refers to a statistical distribution where a few common events (the “head”) comprise the majority of data, while many rare events (the “tail”) have few instances [21]. In trajectory prediction, this means datasets are rich in common driving behaviors but sparse in rare ones. These rare behaviors, however, are often critical for safety, involving complex maneuvers or high-risk situations [27, 41].

Q2: How can we effectively identify and characterize long-tail trajectories within datasets? Identifying which trajectories constitute the long tail is challenging due to the lack of explicit labels and the diversity of rare behaviors. Existing methods often rely on model-specific prediction errors to infer long-tail instances [19], which can be inconsistent across models and insufficient for capturing rare behaviors. A systematic approach is needed to identify long-tail trajectories based on intrinsic properties, such as risk levels, maneuver complexity, or other meaningful criteria. This would enable a more comprehensive understanding of underrepresented behaviors that are critical to safety.

Q3: How can we design models that accurately predict long-tail trajectories without compromising overall performance? Efforts to improve prediction accuracy on rare trajectories often face a trade-off with performance on common behaviors [23, 38, 42]. Focusing excessively on the tail can lead to overfitting or neglect of the head, reducing overall model effectiveness. Therefore, it is imperative to develop learning strategies that can enhance the prediction of rare trajectories while maintaining or even improving performance on common ones.

To address these critical questions, we propose a comprehensive approach to systematically tackle the long-tail challenge in trajectory prediction. We begin by formally defining the long-tail distribution in the context of autonomous driving by analyzing real-world driving data, illustrating how the imbalance manifests, and discussing its implications for model performance and safety. Next, we introduce a multi-criteria method for identifying long-tail trajectories based on intrinsic properties such as prediction error distribution, risk metrics like low time-to-collision (TTC), and complex vehicle states. By integrating these perspectives, as shown in Figure 1, we comprehensively characterize long-tail trajectories, providing a holistic foundation for developing targeted strategies to improve prediction accuracy in these challenging cases. To address the modeling challenges, we propose AMD, an Adaptive Momentum and Decoupled Contrastive Learning Framework designed to enhance prediction accuracy on long-

tail trajectories without compromising overall performance. AMD incorporates adaptive momentum updating to emphasize underrepresented samples, decoupled contrastive learning to balance optimization between head and tail classes, innovative data augmentation strategies to simulate real-world uncertainties and an online iterative clustering mechanism to adapt to distributional changes in the data.

Our work makes the following contributions:

1) We develop a multi-criteria method to identify and characterize long-tail trajectories based on intrinsic properties, enabling targeted improvements in prediction models. Defining long-tail trajectories by prediction error, risk metrics, and vehicle states ensures the model can effectively handle diverse and complex scenarios.

2) We propose an adaptive and robust framework that effectively balances learning between common and rare trajectories, enhancing prediction accuracy on long-tail data without degrading overall performance.

3) We conduct comprehensive experiments that consistently demonstrate AMD’s superior performance in terms of accuracy, adaptability, and reliability compared to existing state-of-the-art (SOTA) methods, validating its effectiveness across various challenging scenarios.

2. Related Work

Trajectory prediction is a key challenge in autonomous driving. Early methods based on kinematic and statistical models [29] are computationally efficient but struggle with complex environmental influences, limiting accuracy. Driven by data-centric approaches such as VectorNet [11], deep learning models have shown remarkable potential in trajectory prediction. Architectures such as Recurrent Neural Networks (RNNs) [1, 28], Graph Neural Networks (GNNs) [24, 43, 44], and Transformers [22, 23, 34, 38] have significant strengths in modeling temporal and spatial dependencies. However, capturing the inherent uncertainty in vehicle motion remains a major challenge in trajectory prediction.

2.1. Long-Tail Trajectory Prediction

Although existing trajectory prediction models perform well on benchmark datasets, they often struggle with rare or challenging scenarios—an issue known as the long-tail challenge in trajectory prediction [25]. In data-driven deep learning models, prediction performance heavily relies on data quality, and the inherent data imbalance exacerbates the long-tail problem [21]. This issue is not unique to trajectory prediction and is observed across various domains, such as image classification and natural language processing [31]. Numerous strategies, including data resampling [14] and loss re-weighting [37], have been proposed to address this problem. Recently, some studies have specifically targeted long-tail trajectory prediction. For example, FEND

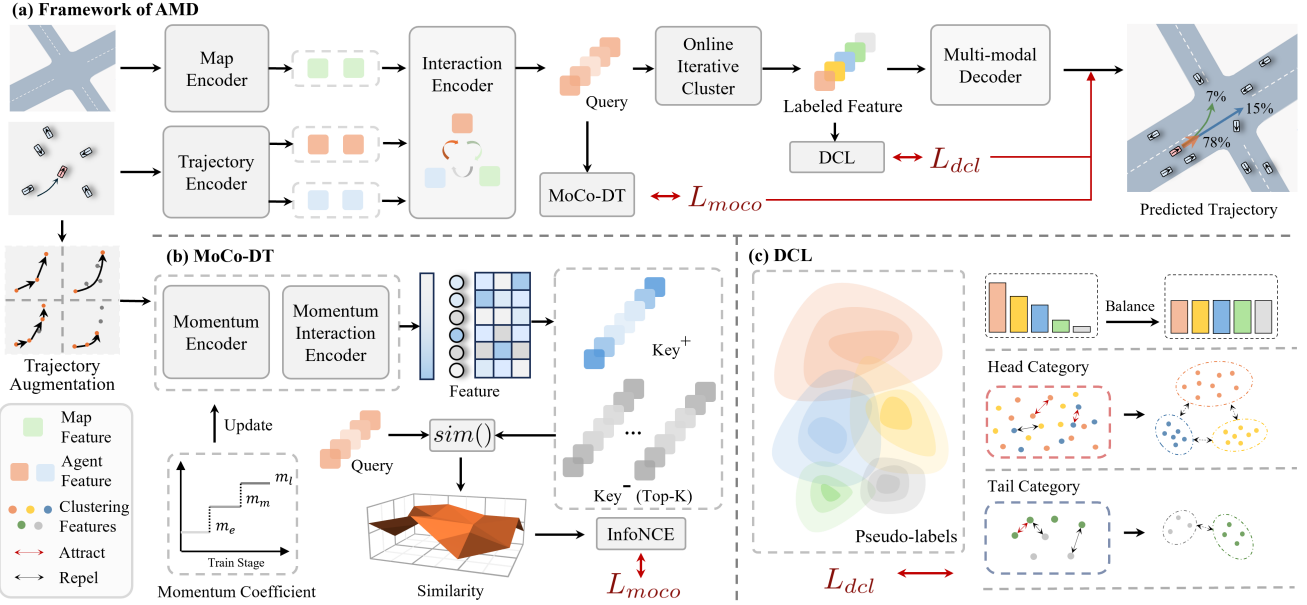


Figure 2. Overview of the proposed AMD framework. Panel (a) illustrates the structure of the model, including the Encoder, Interaction Module, and Predictor, which collectively enable multimodal trajectory prediction. This model takes as input the target agent, surrounding agents, and HD maps, ultimately outputting predicted multimodal trajectories. Panels (b) and (c) present details of the Adaptive Momentum Contrastive Learning (MoCo-DT) module design and the Decoupled Contrastive Learning (DCL) module.

[46] framework enhances long-tail prediction by augmenting future trajectories. TrACT [51] architecture identifies long-tail trajectories based on training curves. Hi-SCL [19] represents traffic scenarios as waveforms to improve feature extraction in long-tail trajectory prediction. These efforts underscore the increasing importance and interest in developing specialized techniques to effectively address the long-tail problem in trajectory prediction.

2.2. Contrastive Learning

Traditional contrastive learning [8] is an unsupervised strategy that compares different views of data to learn similar and distinct features, forming effective representations. Momentum Contrast (MoCo) [16] is an unsupervised contrastive learning framework that overcomes the issue of updating the contrastive sample pool by maintaining a dynamic memory, and enhancing representation learning.

Supervised contrastive learning [17, 49] adds label information to guide the construction of positive and negative pairs, improving the model’s ability to group similar samples and distinguish dissimilar ones. This method is more robust than unsupervised learning. In trajectory prediction, various frameworks have applied contrastive learning to enhance learning on underrepresented samples [6, 19, 46]. However, these methods often overlook the specific challenges of recognizing and predicting long-tail trajectories. Unlike previous methods, our study proposes a dual-layer

contrastive framework combining unsupervised and supervised strategies, enhancing long-tail trajectory recognition and enabling more accurate predictions.

3. Methodology

Problem Formulation. Trajectory prediction is a typical temporal sequence prediction problem. Given a traffic scenario containing $n + 1$ agents, the motion of all agents is represented as a series of state sequences, denoted as $\{X_i, Y_i\}$ ($i \in [0, n]$). Here, $X_i = \{X_i^t | t \in [0, T_h]\}$ represents the past observed trajectory of agent i , which includes information such as the agent’s position, speed, and heading angle of agent i and its surrounding agents, and $Y_i = \{Y_i^t | t \in [T_h, T_h + T_f]\}$ represents the future trajectory of agent i , also known as the ground truth. The road map information is represented as a series of feature vectors $M_N = \{m_1, \dots, m_N\}$. Therefore, the trajectory prediction task is defined as predicting the future trajectory Y_i of the target agent based on the observed past trajectory X_i and environmental context M_N .

Overview. The overall framework of AMD described in Figure 2, starts with four augmentation strategies that randomly enhance the target agent’s trajectory. Both the original and augmented trajectories are processed by a feature encoder, producing high-dimensional feature representations. A scene interaction module employing self-attention and cross-attention mechanisms integrates the target agent’s

features with those of surrounding agents and maps data to create a contextualized representation. These features are then used as positive samples in an improved momentum contrastive learning module (MoCo-DT). An online iterative clustering strategy generates pseudo-labels, and input into a decoupled contrastive learning (DCL) module. Finally, a multi-modal trajectory decoder predicts future trajectories across different modes.

3.1. Trajectory Augmentation

To address data imbalance and enhance model generalization, we draw inspiration from [6] and propose four novel augmentation methods for short-term trajectories: (1) Simplify, (2) Shift, (3) Mask, and (4) Subset. These methods are designed to improve the model’s robustness against real-world uncertainties in trajectory prediction and enhance accuracy for long-tail trajectories. Specifically: (1) Simplify reduces redundant points to emphasize primary movement patterns; (2) Shift applies random displacements to simulate external perturbations; (3) Mask randomly discards data points to mimic sensor failures; and (4) Subset selects consecutive subsequences to emulate incomplete temporal information. By generating diverse trajectory transformations, these methods enable the model to better adapt to varying patterns during training. Our experiments further confirm that such targeted augmentations lead to substantial performance gains in challenging scenarios. Detailed implementations are provided in **Appendix**.

3.2. Trajectory Feature Extractor

Feature Encoder. The trajectory encoder converts raw and augmented trajectory data into high-dimensional representations, providing quality feature inputs for subsequent tasks. We use a hierarchical embedding method combining Multi-Layer Perceptron (MLP), a Transformer Encoder (TrEnc), and Gated Recurrent Unit (GRU) to capture trajectory temporal features. To model driver memory decay, the final GRU hidden states are used as encoding features for the target agent F_{tar} and neighboring agents F_{nbr} . For the HD map M , lane nodes and lines are represented as discrete vectors [11], and hierarchical extraction yields the high-dimensional map encoding F_{lane} .

$$F_{tar} = \phi_G (\phi_T (\phi_M(t_{tar})) + \phi_M(t_{tar})) \quad (1)$$

$$F_{nbr} = \phi_G (\phi_T (\phi_M(t_{veh}, t_{ped})) + \phi_M(t_{veh}, t_{ped})) \quad (2)$$

$$F_{lane} = \phi_A (\phi_G (\phi_M(l_{node})), A) \quad (3)$$

where ϕ_M , ϕ_T , ϕ_G , and ϕ_A denote embedding functions implemented by the MLP, TrEnc, GRU and GAT. t_{tar} , t_{veh} , t_{ped} , and l_{node} represent the input features of the target agent, surrounding vehicles, pedestrians, and lane nodes, respectively. A is the adjacency matrix of the lane nodes.

Scene Interaction Encoder. To capture the dynamic interactions between the target agent and its surrounding environment, we designed a scene interaction encoder that incorporates a latent variable mechanism to generate diverse potential trajectory patterns. This module uses a unified cross-modal attention mechanism to fuse information from multiple modalities, including target features, surrounding agent features, and road node features. The result is a set of multimodal fused features F_{cross} , generated through positional encoding (pos) and multi-head attention (MHA).

$$F_{cross} = MHA(H_{mode} + pos, K_{t+n+l}, V_{t+n+l}) \quad (4)$$

where H_{mode} represents the pattern features generated by the latent variable mechanism, and K_{t+n+l} and V_{t+n+l} are the merged key and value, which include features of the target, surrounding agents, and road nodes. Further details are in the **Appendix**.

3.3. Momentum Contrastive Learning

In long-tail prediction tasks, rare samples are often overlooked due to their scarcity. To address this, we propose an improved Dynamic Momentum and Top-K Hard Negative Mining method (MoCo-DT), which improves the focus on long-tail samples in contrastive learning. Compared to the original MoCo approach [16], MoCo-DT dynamically adjusts the momentum coefficient m based on training progress t and total duration T , employing distinct coefficients m_e, m_m, m_l for the early, middle, and late training stages, respectively, to adapt to different training stages.

The Top-K Hard Negative Mining mechanism strengthens the model’s ability to distinguish challenging long-tail samples by selecting the most similar negative samples. Specifically, this mechanism computes the similarity between the query sample and both positive and negative samples, dynamically selecting the Top-K hardest negatives from the negative set to emphasize learning from difficult long-tail features in contrastive training.

$$l_{pos} = sim(q, k^+), \quad l_{neg} = sim(q, k^-) \quad (5)$$

$$\{k_1^-, \dots, k_K^-\} = TopK_{k^-}(sim(q, k^-))$$

where $sim()$ denotes the similarity function, q is the query encoding, and k^+ and k^- represent positive and negative encodings, respectively. K is the number of hard negative samples selected, and $TopK$ refers to selecting the top K samples with the highest similarity to the query sample among the negative samples.

The final contrastive loss is constructed by comparing the positive sample with the Top-K hard negative samples, calculated as follows:

$$L_{moco} = -\log \frac{\exp(sim(q, k^+)/\tau)}{\exp(sim(q, k^+)/\tau) + \sum_{i=1}^K \exp(sim(q, k_i^-)/\tau)} \quad (6)$$

where τ is a temperature parameter. This design allows the model to focus on challenging long-tail features.

3.4. Online Iterative Clustering Strategy

Traditional methods typically use static clustering on encoder-derived features with fixed labels, which struggle to adapt to dynamic feature distributions, especially in long-tail data. To address this, we propose an Online Iterative Clustering Strategy that dynamically updates pseudo-labels during training to improve recognition of long-tail patterns.

This strategy involves clustering sample features in each training epoch to generate adaptive pseudo-labels. After each mini-batch, target feature representations are stored in a feature set. At predefined intervals, K-means clustering [15] is applied to this set to create clusters representing distinct trajectory patterns, with clustering outcomes serving as pseudo-labels. By continuously updating these labels, our approach effectively captures subtle variations in rare trajectories, improving robustness and accuracy in long-tail trajectory identification.

3.5. Decoupled Contrastive Learning

Decoupled Contrastive Learning (DCL) [49] is a type of supervised contrastive learning approach. Compared with traditional contrastive methods, DCL mitigates the bias toward head classes with high frequency, thus enhancing the prediction performance on long-tail data. By assigning different weights to positive samples from two categories, DCL achieves a balanced representation of both head and tail classes. DCL employs an L2 regularization to prevent the optimization from being influenced by class sample sizes. This approach effectively maximizes inter-class distance while minimizing intra-class distance, ensuring robust performance across head and tail classes. The DCL loss function is defined as:

$$L_{dcl} = \frac{-1}{|P_i| + 1} \sum_{q_t \in \{q_i^+, P_i\}} \log \frac{\exp(w_r \cdot \langle q_t, q_i \rangle / \tau)}{\sum_{q_m \in \{q_i^+, U_i\}} \exp(\langle q_i, q_m \rangle / \tau)} \quad (7)$$

where q_i and q_i^+ are features of positive samples in the same category, and U_i is the set of all other category features. P_i denotes the set of features in a given category, τ is a temperature parameter, and w_r is the weight defined as:

$$w_r = \begin{cases} \alpha(|P_i| + 1), & \text{if } q_i = q_i^+ \\ (1 - \alpha)(|P_i| + 1)/|P_i|, & \text{if } q_i \in P_i \end{cases} \quad (8)$$

where $\alpha \in [0, 1]$ is a hyperparameter balancing the weighting between within-category and inter-category samples.

3.6. Multi-modal Decoder

In long-tail trajectory prediction, the target agent may have multiple possible future paths. To address this, we design a Multi-modal Decoder that captures diverse trajectory modes using a latent variable mechanism and a Laplace Mixture Density Network (Laplace MDN). A single-layer GRU decoder generates varied trajectories, with the Laplace MDN

outputting position, scale parameters, and probabilities to assess each mode’s likelihood.

3.7. Training Loss

For multi-modal trajectory prediction, we employ a Laplace negative log-likelihood as the regression loss L_{reg} and a cross-entropy loss L_{cls} for mode classification, with direct training loss L_{target} as the task loss L_{task} . To further enable the model to capture long-tail trajectory characteristics, we incorporate momentum contrastive loss and decoupled contrastive loss to ensure the accuracy of trajectory prediction. The final total loss L is defined as follows:

$$L_{task} = L_{target} + \gamma_1 L_{reg} + \gamma_2 L_{cls} \quad (9)$$

$$L = L_{task} + \lambda_1 L_{moco} + \lambda_2 L_{dcl} \quad (10)$$

where γ_1 , γ_2 , λ_1 , and λ_2 are weighting parameters.

4. Experiments

4.1. Experimental Setup

Datasets. We evaluate our proposed method on the nuScenes [3] and ETH/UCY [20, 36] datasets, which contain real-world traffic data for vehicle and pedestrian scenarios, respectively, covering diverse trajectory patterns.

Long-tail Subset. To validate our model on long-tail data, we divide the dataset using three distinct criteria, differing from previous studies in long-tail trajectory prediction:

- **Prediction Error:** The dataset is divided into seven subsets based on prediction error: the Top 1%-5% with the highest errors, the remaining samples, and all samples.
- **Risk Metric:** We use (TTC) as a risk metric, identifying the Top 1%-3% of samples with the lowest TTC values, representing high-risk scenarios for target agents.
- **Vehicle State:** We categorize samples based on the target agent’s behavior, specifically labeling rapid acceleration, rapid deceleration, sharp lane changes, and sharp turns, creating four distinct long-tail subsets.

This multi-criteria approach avoids the limitations of single criteria definitions for long-tail trajectories and provides a more comprehensive evaluation of the model’s performance on diverse long-tail trajectories.

Metrics. We evaluate trajectory prediction performance using Average Displacement Error (ADE), Final Displacement Error (FDE), and Miss Rate (MR). For long-tail samples, we use minimum ADE (minADE) and minimum FDE (minFDE) to better assess performance on challenging samples. For overall multi-modal prediction, we use minADE_k and minFDE_k to evaluate the top-K predicted trajectories.

Implementation Details. In our experiments, the loss function weight parameters are set to $\gamma_1 = 1$, $\gamma_2 = 0.5$, $\lambda_1 = 1$, and $\lambda_2 = 0.1$. For MoCo-DT, the parameters m_e , m_m , and m_l are set to 0.95, 0.99, and 0.999, respectively. All models

Dataset	Model	Top 1%	Top 2%	Top 3%	Top 4%	Top 5%	Rest	ALL
nuScenes	Traj++ EWTA [32]	1.73/4.43	1.36/3.54	1.17/3.03	1.04/2.68	0.95/2.41	0.16/0.26	0.22/0.39
	Traj++ EWTA+contrastive [32]	1.28/2.85	0.97/2.15	0.83/1.83	0.76/1.64	0.70/1.48	0.15/0.24	0.18/0.30
	FEND [46]	<u>1.21/2.50</u>	<u>0.92/1.88</u>	<u>0.79/1.61</u>	<u>0.72/1.43</u>	<u>0.66/1.31</u>	0.14/0.20	0.17/0.26
	TrACT [51]	1.23/2.65	0.98/2.11	0.85/1.82	0.78/1.64	0.72/1.49	-	0.19/0.31
	AMD (Ours)	1.08/1.66	0.85/1.33	0.75/1.15	0.69/1.03	0.64/0.95	0.18/0.16	0.21/0.21
ETH/UCY	Traj++ EWTA [32]	0.98/2.54	0.79/2.07	0.71/1.81	0.65/1.63	0.60/1.50	0.14/0.26	0.17/0.32
	Traj++ EWTA+resample [39]	0.90/2.17	0.77/1.90	0.73/1.78	0.66/1.60	0.64/1.52	0.20/0.41	0.23/0.47
	Traj++ EWTA+reweighting [9]	0.97/2.47	0.78/2.03	0.68/1.73	0.62/1.55	0.56/1.40	0.15/0.26	0.18/0.32
	Traj++ EWTA+contrastive [32]	0.92/2.33	0.74/1.91	0.67/1.71	0.60/1.48	0.55/1.32	0.15/0.27	0.17/0.32
	LDAM [4]	0.92/2.35	0.76/1.96	0.68/1.71	0.62/1.53	0.57/1.37	0.15/0.27	0.17/0.33
	FEND [46]	0.84/2.13	0.68/1.68	0.61/1.46	0.56/1.30	0.52/1.19	<u>0.15/0.27</u>	0.17/0.32
	TrACT [51]	0.80/2.00	0.65/1.63	0.61/1.46	0.56/1.31	<u>0.52/1.18</u>	-	0.17/0.32
AMD (Ours)	0.76/1.75	0.66/1.59	0.58/1.37	0.54/1.25	0.51/1.16	0.16/0.24	0.18/0.27	

Table 1. Prediction errors (minADE/minFDE) for seven test samples from the nuScenes and ETH/UCY datasets, categorized by prediction error (FDE). For comparison with other methods, the nuScenes dataset uses a prediction horizon of 2s, while the ETH/UCY dataset uses a prediction horizon of 4.8s. The Top 1%-5% refers to the subset of samples with the largest prediction errors. Bold and underlined text represent the best and second-best results, respectively. Cases marked with ('-') indicate missing values.

are trained on an NVIDIA RTX 3090 GPU. For additional experimental setup details, please refer to the **Appendix**.

4.2. Comparisons to SOTA

(i) **Quantitative Comparison under Prediction Error.** To demonstrate the effectiveness of our method, we compared it with state-of-the-art long-tail trajectory prediction models. As shown in Table 1, our method outperforms other models on the Top 1%-5% most challenging long-tail samples. For the Top 1% hardest samples, our method achieves an error of 1.08/1.66, reducing error by 14.9% and 33.6% compared to the closest competing model. Additionally, for all samples (All), our method also demonstrates superior performance, reducing the minFDE by 19.2% compared to the closest competing model. These results highlight AMD’s advantages in long-tail trajectory prediction and its competitive edge in overall accuracy, with high precision and consistency across different levels of sample difficulty.

(ii) **Quantitative Results under Risk Metric and Vehicle State.** To validate our model’s effectiveness in challenging long-tail scenarios, we compared it against a leading baseline [7] using the Risk and State subsets from nuScenes (Tables 2 and 3). These metrics provide deeper insights into performance based on collision risk and motion patterns. As shown in Table 2, our model excels in high-risk situations, improving upon the baseline by 8.5% in minADE and 25.8% in minFDE for the Top 1% risk group. This advantage is also evident in scenarios with aggressive maneuvers (Table 3). Notably, for the difficult Sharp Lane Change (SLC) subset, our model achieves a 2.7% lower minADE and a 14.0% lower minFDE, underscoring its superior performance and robustness in critical long-tail conditions.

(iii) **Quantitative Comparison on All Samples.** To validate the overall effectiveness of our method, we compared it with SOTA trajectory prediction models on the nuScenes

	Risk	Top 1%	Top 2%	Top 3%	ALL
Q-EANet [7]	0.71/0.97	0.71/0.96	0.78/1.06	0.70/0.99	
AMD (Ours)	0.65/0.72	0.72/0.83	0.70/0.87	0.69/0.88	

Table 2. Comparison of prediction errors (minADE/minFDE) on the nuScenes dataset, categorized by risk level. The Top 1%-3% subsets represent trajectories with the highest collision risk.

State	RA	RD	SLC	ST	Normal	ALL
Q-EANet [7]	0.86/1.14	0.96/1.15	1.13/1.64	0.97/1.37	0.61/0.89	0.70/0.99
AMD (Ours)	0.80/1.01	0.90/1.08	1.10/1.41	0.94/1.28	0.61/0.78	0.69/0.88

Table 3. Comparison of prediction errors (minADE/minFDE) on the nuScenes dataset, categorized by vehicle motion state for the six test samples. RA: Rapid Acceleration, RD: Rapid Deceleration, SLC: Sharp Lane Change, ST: Sharp Turn.

Model	minADE ₅	minADE ₁₀	minFDE ₁	MR ₅
Trajectron++ [38]	1.88	1.51	9.52	0.70
P2T [10]	1.45	1.16	-	0.64
LaPred [18]	1.47	1.12	8.12	0.53
GoHome [12]	1.42	1.15	<u>6.99</u>	<u>0.57</u>
ContextVAE [47]	1.59	-	8.24	-
SeFlow [52]	<u>1.38</u>	0.98	7.89	0.60
AFormer-FLN [48]	1.83	1.32	-	-
AMD (Ours)	1.23	<u>1.06</u>	6.99	0.50

Table 4. Comparison of the performance of various models across all samples on nuScenes dataset, using 6s trajectory predictions.

and ETH/UCY datasets. In Table 4, our model outperforms others in terms of minADE₅ and MR₅ metrics, achieving improvements of 10.9% and 5.7%, respectively, over the leading model. Results on ETH/UCY (Table 5) further confirm the AMD model’s superiority, surpassing others across all scenarios with average improvements of 5.3% in minADE and 12.9% in minFDE over the second-best model. These consistent and substantial performance gains clearly

Model	Venue	ETH	HOTEL	UNIV	ZARA1	ZARA2	AVG
PECNet [33]	ECCV	0.54/0.87	0.18/0.24	0.22/0.39	0.17/0.30	0.35/0.60	0.29/0.48
AgentFormer [50]	ICCV	0.45/0.75	0.14/0.22	0.25/0.45	0.18/0.30	0.14/0.24	0.23/0.39
Trajectron++ [38]	ECCV	0.39/0.83	0.12/0.21	<u>0.20/0.44</u>	0.15/0.33	0.11/0.25	<u>0.19/0.41</u>
NPSN [2]	CVPR	0.36/0.59	0.16/0.25	<u>0.23/0.39</u>	0.18/0.32	0.14/0.25	0.21/0.36
MID [13]	CVPR	0.39/0.66	0.13/0.22	0.22/0.45	0.17/0.30	0.13/0.27	0.21/0.38
TUTR [40]	ICCV	0.40/0.61	0.11/0.18	0.23/0.42	0.18/0.34	0.13/0.25	0.21/0.36
PPT [30]	ECCV	<u>0.36/0.51</u>	<u>0.11/0.15</u>	0.22/0.40	0.17/0.30	0.12/0.21	0.20/0.31
AMD (Ours)	-	0.32/0.42	0.09/0.13	0.20/0.34	<u>0.16/0.26</u>	<u>0.12/0.21</u>	0.18/0.27

Table 5. Comparison of the performance (minADE/minFDE) of various models across all samples on the ETH/UCY dataset.

Model	Components				Performance (minADE/minFDE)						
	TA	Moco-DT	IC	DCL	Top 1%	Top 2%	Top 3%	Top 4%	Top 5%	Rest	ALL
A	×	×	×	×	1.55/2.41	1.23/1.90	1.06/1.65	0.97/1.49	0.90/1.37	0.24/0.23	0.28/0.29
B	×	✓	✓	✓	1.47/2.05	1.10/1.57	0.95/1.35	0.85/1.21	0.79/1.10	0.20/0.18	0.23/0.22
C	✓	×	✓	✓	1.30/1.79	0.97/1.41	0.85/1.22	0.77/1.09	0.71/1.00	0.19/0.16	0.22/0.20
D	✓	✓	×	✓	1.38/1.95	1.05/1.51	0.90/1.30	0.81/1.16	0.75/1.06	0.20/0.17	0.23/0.21
E	✓	✓	✓	×	1.45/2.02	1.12/1.56	0.95/1.32	0.85/1.18	0.77/1.08	0.20/0.17	0.23/0.21
F	✓	✓	✓	✓	1.08/1.66	0.85/1.33	0.75/1.15	0.69/1.03	0.64/0.95	0.18/0.16	0.21/0.21

Table 6. Ablation results of different components on nuScenes dataset. TA means Trajectory Augmentation, IC means Iterative Clustering.

demonstrate our AMD model’s strong predictive capability in long-tail scenarios and robust generalization across diverse datasets and conditions.

4.3. Ablation Studies

Our model integrates key components that enhance its performance, evaluated through ablation studies (Table 6). The complete model (Model F) achieves SOTA performance across all metrics, demonstrating the strong synergistic effects among its components. Model A, with all modules removed, performs the worst, highlighting their collective importance. Model B (without trajectory augmentation) and Model E (without the decoupled contrastive learning) exhibit poor performance on top long-tail trajectories. The former struggles to capture rare trajectory patterns, while the latter increases prediction randomness due to insufficient class discrimination, indicating that trajectory augmentation and decoupled contrastive learning are crucial for long-tail learning. Other variants also show performance degradation when key components are removed, particularly on the Top 1% samples, confirming the collaborative contribution of each component to long-tail prediction.

4.4. Qualitative Comparison

Figure 3 presents visualization results of multimodal trajectory predictions on the nuScenes dataset under various long-tail scenarios, comparing others model [7] with ours (AMD model and its ablation variants Model B and Model E). Panels (a) and (b) depict high-curvature vehicle turning trajectories, while Panel (c) shows a trajectory with distinct deceleration actions. The results demonstrate that AMD accurately predicts these complex trajectories and gener-

ates additional plausible options. Compared to Model B and Model E, the trajectory augmentation (TA) strategy enhances generalization to complex dynamics by producing diverse samples, effectively capturing geometric features of maneuvers like turns. Meanwhile, decoupled contrastive learning (DCL) improves differentiation of rare trajectories by separating positive and negative sample representations, reducing prediction randomness. This mechanism enables AMD to maintain accuracy and model multimodal uncertainty effectively in long-tail distributions.

4.5. Inference Time Comparison

To demonstrate the efficiency of our AMD model, we conducted a comparative experiment on inference times using the nuScenes dataset, with VisionTrap tested on an RTX 3090 Ti GPU and the other models, including ours, evaluated on an RTX 3090 GPU. As shown in Table 7, AMD exhibits a clear advantage in inference speed. The results highlight that our model significantly reduces inference time while maintaining accuracy, making it well-suited for real-time autonomous driving.

Model	Inference Time (ms)	minADE ₁₀	minFDE ₁
Trajectron++ [38]	<u>38</u>	1.88	9.52
MultiPath [5]	87	1.50	<u>7.69</u>
P2T [50]	116	<u>1.16</u>	10.5
VisionTrap [35]	53	1.17	8.72
AMD (Ours)	14	1.06	6.99

Table 7. Inference time comparison on nuScenes dataset.

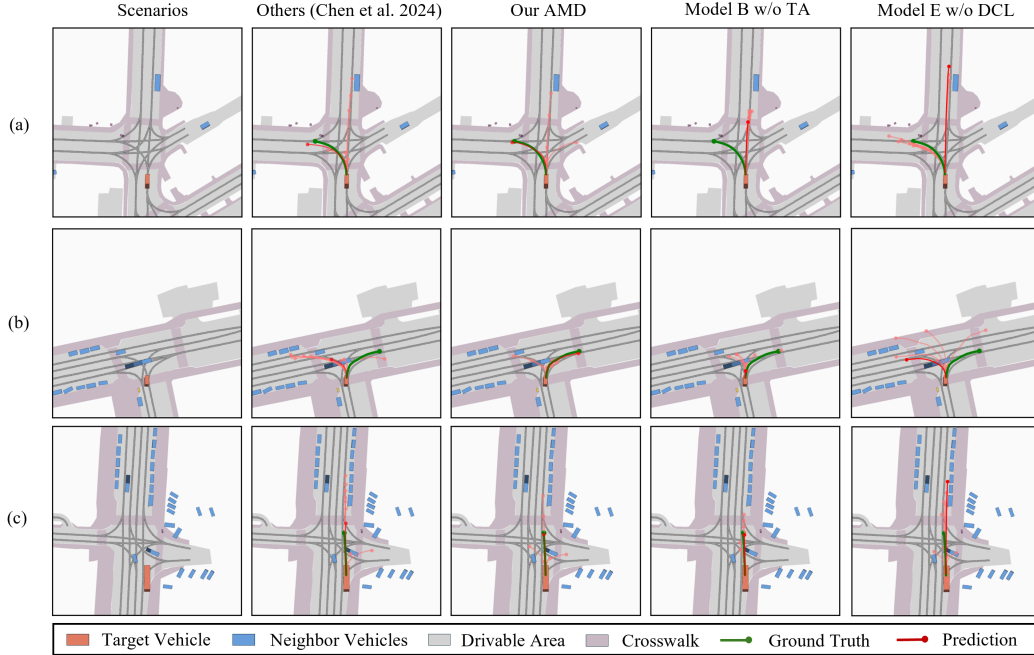


Figure 3. Qualitative results of long-tail trajectory predictions, covering various driving actions: (a) Turn left. (b) Turn right. (c) Deceleration. The red lines show the most probable trajectory, while the light red lines show the predicted multimodal trajectories.

4.6. Feature Space Visualisation

We conducted feature space visualization by applying t-SNE on nuScenes dataset to reduce the extracted features into a two-dimensional space for analysis. As shown in Figure 4a, our model enhances cluster compactness and tail separation, clearly separating head and tail patterns while forming distinct clusters for tail patterns. In contrast, removing MoCo-DT (Figure 4b) disperses hard samples, cluttering tail patterns, while removing DCL (Figure 4c) increases head-tail overlap. This confirms MoCo-DT boosts tail representation and DCL mitigates head-class dominance via balanced learning.

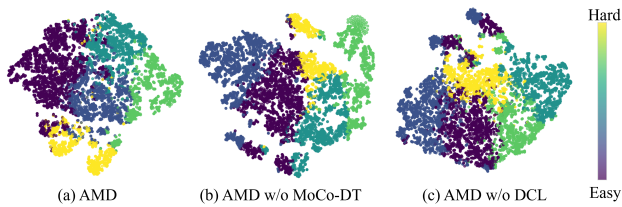


Figure 4. Visualization of feature spaces for different variants.

4.7. Hyperparameter Sensitivity Study

We conducted a sensitivity analysis of the hyperparameters λ_1 , λ_2 , m_e , m_m , and m_l on the nuScenes dataset, with results presented in Table 8. The parameters λ_1 and λ_2 influence the model’s ability to address long-tail trajectories by balancing loss contributions, while m_e , m_m , and m_l

primarily affect model stability. The optimal combination yielding the best performance was selected.

Hyperparameter Setting	Top 1%	Top 2%	Top 3%	ALL
λ_1, λ_2 1.0, 0.5	1.10/1.77	0.87/1.39	0.77/1.19	0.22/0.22
λ_1, λ_2 1.0, 0.1	1.08/1.66	0.85/1.33	0.75/1.15	0.21/0.21
λ_1, λ_2 0.5, 0.5	1.11/1.68	0.87/1.36	0.77/1.15	0.21/0.22
λ_1, λ_2 0.5, 0.1	1.26/2.00	0.97/1.53	0.87/1.33	0.22/0.24
m_e , 0.90, 0.95, 0.999	1.16/2.04	0.94/1.60	0.83/1.37	0.25/0.26
m_m , 0.90, 0.99, 0.999	1.14/1.79	0.91/1.38	0.77/1.16	0.22/0.22
m_l 0.95, 0.99, 0.999	1.08/1.66	0.85/1.33	0.75/1.15	0.21/0.21

Table 8. Hyperparameter sensitivity analysis on nuScenes dataset.

5. Conclusion

In this paper, we propose an Adaptive Momentum and Decoupled Contrastive Learning framework (AMD) tailored for robust trajectory prediction in challenging long-tail scenarios. Leveraging a novel combination of unsupervised and supervised contrastive learning, AMD effectively enhances predictive performance on rare trajectory patterns while maintaining high accuracy on general trajectory distributions. Additionally, our random trajectory augmentation and online iterative learning strategies significantly boost the model’s adaptability, allowing it to robustly handle complex and diverse spatiotemporal dynamics. Experimental results demonstrate that AMD consistently surpasses SOTA methods on long-tail subsets and achieves competitive overall accuracy across multiple datasets.

Acknowledgements

This work was supported by the Science and Technology Development Fund of Macau [0122/2024/RIB2, 0215/2024/AGJ, 001/2024/SKL], the Research Services and Knowledge Transfer Office, University of Macau [SRG2023-00037-IOTSC, MYRG-GRG2024-00284-IOTSC], the Shenzhen-Hong Kong-Macau Science and Technology Program Category C [SGDX20230821095159012], the State Key Lab of Intelligent Transportation System [2024-B001], and the Jiangsu Provincial Science and Technology Program [BZ2024055].

References

- [1] Alexandre Alahi, Kratharth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social Istm: Human trajectory prediction in crowded spaces. In *CVPR*, pages 961–971, 2016. 2
- [2] Inhwon Bae, Jin-Hwi Park, and Hae-Gon Jeon. Non-probability sampling network for stochastic human trajectory prediction. In *CVPR*, pages 6477–6487, 2022. 7
- [3] 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. In *CVPR*, pages 11621–11631, 2020. 5
- [4] Kaidi Cao, Colin Wei, Adrien Gaidon, Nikos Arechiga, and Tengyu Ma. Learning imbalanced datasets with label-distribution-aware margin loss. *Advances in neural information processing systems*, 32, 2019. 6
- [5] Yuning Chai, Benjamin Sapp, Mayank Bansal, and Dragomir Anguelov. Multipath: Multiple probabilistic anchor trajectory hypotheses for behavior prediction. In *Conference on Robot Learning*, pages 86–99. PMLR, 2020. 7
- [6] Yanchuan Chang, Jianzhong Qi, Yuxuan Liang, and Egemen Tanin. Contrastive trajectory similarity learning with dual-feature attention. In *2023 IEEE 39th International conference on data engineering (ICDE)*, pages 2933–2945. IEEE, 2023. 3, 4
- [7] Jiuyu Chen, Zhongli Wang, Jian Wang, and Baigen Cai. Q-eonet: Implicit social modeling for trajectory prediction via experience-anchored queries. *IET Intelligent Transport Systems*, 18(6):1004–1015, 2024. 6, 7
- [8] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A simple framework for contrastive learning of visual representations. In *International conference on machine learning*, pages 1597–1607. PMLR, 2020. 3
- [9] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *CVPR*, pages 9268–9277, 2019. 6
- [10] Nachiket Deo and Mohan M Trivedi. Trajectory forecasts in unknown environments conditioned on grid-based plans. *arXiv preprint arXiv:2001.00735*, 2020. 6
- [11] Jiyang Gao, Chen Sun, Hang Zhao, Yi Shen, Dragomir Anguelov, Congcong Li, and Cordelia Schmid. Vectornet: Encoding hd maps and agent dynamics from vectorized representation. In *CVPR*, pages 11525–11533, 2020. 2, 4
- [12] Thomas Gilles, Stefano Sabatini, Dzmitry Tsishkou, Bogdan Stanculescu, and Fabien Moutarde. Gohome: Graph-oriented heatmap output for future motion estimation. In *2022 international conference on robotics and automation (ICRA)*, pages 9107–9114. IEEE, 2022. 6
- [13] Tianpei Gu, Guangyi Chen, Junlong Li, Chunze Lin, Yongming Rao, Jie Zhou, and Jiwen Lu. Stochastic trajectory prediction via motion indeterminacy diffusion. In *CVPR*, pages 17113–17122, 2022. 7
- [14] Hui Han, Wen-Yuan Wang, and Bing-Huan Mao. Borderline-smote: a new over-sampling method in imbalanced data sets learning. In *International conference on intelligent computing*, pages 878–887. Springer, 2005. 2
- [15] John A Hartigan and Manchek A Wong. Algorithm as 136: A k-means clustering algorithm. *Journal of the royal statistical society. series c (applied statistics)*, 28(1):100–108, 1979. 5
- [16] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum contrast for unsupervised visual representation learning. In *CVPR*, pages 9729–9738, 2020. 3, 4
- [17] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschiot, Ce Liu, and Dilip Krishnan. Supervised contrastive learning. *Advances in neural information processing systems*, 33:18661–18673, 2020. 3
- [18] ByeoungDo Kim, Seong Hyeon Park, Seokhwan Lee, Elbek Khoshimjonov, Dongsuk Kum, Junsoo Kim, Jeong Soo Kim, and Jun Won Choi. Lapred: Lane-aware prediction of multimodal future trajectories of dynamic agents. In *CVPR*, pages 14636–14645, 2021. 6
- [19] Zhengxing Lan, Yilong Ren, Haiyang Yu, Lingshan Liu, Zhenning Li, Yin Hai Wang, and Zhiyong Cui. Hi-scl: Fighting long-tailed challenges in trajectory prediction with hierarchical wave-semantic contrastive learning. *Transportation Research Part C: Emerging Technologies*, 165:104735, 2024. 2, 3
- [20] Laura Leal-Taixé, Michele Fenzi, Alina Kuznetsova, Bodo Rosenhahn, and Silvio Savarese. Learning an image-based motion context for multiple people tracking. In *CVPR*, pages 3542–3549, 2014. 5
- [21] Bolian Li, Zongbo Han, Haining Li, Huazhu Fu, and Changqing Zhang. Trustworthy long-tailed classification. In *CVPR*, pages 6970–6979, 2022. 2
- [22] Haicheng Liao, Xuelin Li, Yongkang Li, Hanlin Kong, Chengyue Wang, Bonan Wang, Yanchen Guan, K Tam, and Zhenning Li. Cdstraj: Characterized diffusion and spatial-temporal interaction network for trajectory prediction in autonomous driving. In *IJCAI*, pages 7331–7339, 2024. 2
- [23] Haicheng Liao, Zhenning Li, Huanming Shen, Wenxuan Zeng, Dongping Liao, Guofa Li, and Chengzhong Xu. Bat: Behavior-aware human-like trajectory prediction for autonomous driving. In *AAAI*, pages 10332–10340, 2024. 2
- [24] Haicheng Liao, Zhenning Li, Chengyue Wang, Bonan Wang, Hanlin Kong, Yanchen Guan, Guofa Li, and Zhiyong Cui. A

- cognitive-driven trajectory prediction model for autonomous driving in mixed autonomy environments. In *IJCAI*, 2024. 2
- [25] Haicheng Liao, Chengyue Wang, Zhenning Li, Yongkang Li, Bonan Wang, Guofa Li, and Chengzhong Xu. Physics-informed trajectory prediction for autonomous driving under missing observation. In *IJCAI*, 2024. 2
- [26] Haicheng Liao, Hanlin Kong, Bonan Wang, Chengyue Wang, Wang Ye, Zhengbing He, Chengzhong Xu, and Zhenning Li. Cot-drive: Efficient motion forecasting for autonomous driving with llms and chain-of-thought prompting. *IEEE Transactions on Artificial Intelligence*, 2025. 1
- [27] Haicheng Liao, Zhenning Li, Guohui Zhang, Keqiang Li, and Chengzhong Xu. Toward human-like trajectory prediction for autonomous driving: A behavior-centric approach. *Transportation Science*, 2025. 2
- [28] Haicheng Liao, Chengyue Wang, Kaiqun Zhu, Yilong Ren, Bolin Gao, Shengbo Eben Li, Chengzhong Xu, and Zhenning Li. Minds on the move: Decoding trajectory prediction in autonomous driving with cognitive insights. *IEEE Transactions on Intelligent Transportation Systems*, 2025. 2
- [29] Chiu-Feng Lin, A Galip Ulsoy, and David J LeBlanc. Vehicle dynamics and external disturbance estimation for vehicle path prediction. *IEEE Transactions on Control Systems Technology*, 8(3):508–518, 2000. 2
- [30] Xiaotong Lin, Tianming Liang, Jianhuang Lai, and Jianfang Hu. Progressive pretext task learning for human trajectory prediction. In *ECCV*, pages 197–214. Springer, 2024. 7
- [31] Ziwei Liu, Zhongqi Miao, Xiaohang Zhan, Jiayun Wang, Boqing Gong, and X Yu Stella. Open long-tailed recognition in a dynamic world. *IEEE TPAMI*, 46(3):1836–1851, 2022. 2
- [32] Osama Makansi, Özgün Cicek, Yassine Marrakchi, and Thomas Brox. On exposing the challenging long tail in future prediction of traffic actors. In *ICCV*, pages 13147–13157, 2021. 6
- [33] Karttikeya Mangalam, Harshayu Girase, Shreyas Agarwal, Kuan-Hui Lee, Ehsan Adeli, Jitendra Malik, and Adrien Gaidon. It is not the journey but the destination: Endpoint conditioned trajectory prediction. In *ECCV*, pages 759–776. Springer, 2020. 7
- [34] Abdullah Mohamed, Kun Qian, Mohamed Elhoseiny, and Christian Claudel. Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction. In *CVPR*, pages 14424–14432, 2020. 2
- [35] Seokha Moon, Hyun Woo, Hongbeen Park, Haeji Jung, Reza Mahjourian, Hyung-gun Chi, Hyerin Lim, Sangpil Kim, and Jinkyu Kim. Visiontrap: Vision-augmented trajectory prediction guided by textual descriptions. In *ECCV*, pages 361–379. Springer, 2024. 7
- [36] Stefano Pellegrini, Andreas Ess, Konrad Schindler, and Luc Van Gool. You’ll never walk alone: Modeling social behavior for multi-target tracking. In *ICCV*, pages 261–268. IEEE, 2009. 5
- [37] T-YLPG Ross and GKHP Dollár. Focal loss for dense object detection. In *CVPR*, pages 2980–2988, 2017. 2
- [38] Tim Salzmann, Boris Ivanovic, Punarjay Chakravarty, and Marco Pavone. Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data. In *ECCV*, pages 683–700. Springer, 2020. 2, 6, 7
- [39] Li Shen, Zhouchen Lin, and Qingming Huang. Relay back-propagation for effective learning of deep convolutional neural networks. In *ECCV*, pages 467–482. Springer, 2016. 6
- [40] Liushuai Shi, Le Wang, Sanping Zhou, and Gang Hua. Trajectory unified transformer for pedestrian trajectory prediction. In *ICCV*, pages 9675–9684, 2023. 7
- [41] Divya Thuremella, Lewis Ince, and Lars Kunze. Risk-aware trajectory prediction by incorporating spatio-temporal traffic interaction analysis. In *2024 IEEE International Conference on Robotics and Automation (ICRA)*, pages 14421–14427. IEEE, 2024. 2
- [42] Bonan Wang, Haicheng Liao, Chengyue Wang, Bin Rao, Yanchen Guan, Guyang Yu, Jiaxun Zhang, Songning Lai, Chengzhong Xu, and Zhenning Li. Beyond patterns: Harnessing causal logic for autonomous driving trajectory prediction. *arXiv preprint arXiv:2505.06856*, 2025. 2
- [43] Chengyue Wang, Haicheng Liao, Zhenning Li, and Chengzhong Xu. Wake: Towards robust and physically feasible trajectory prediction for autonomous vehicles with wavelet and kinematics synergy. *PAMI*, 2025. 2
- [44] Chengyue Wang, Haicheng Liao, Bonan Wang, Yanchen Guan, Bin Rao, Ziyuan Pu, Zhiyong Cui, Cheng-Zhong Xu, and Zhenning Li. Nest: A neuromodulated small-world hypergraph trajectory prediction model for autonomous driving. In *AAAI*, pages 808–816, 2025. 2
- [45] Chengyue Wang, Haicheng Liao, Kaiqun Zhu, Guohui Zhang, and Zhenning Li. A dynamics-enhanced learning model for multi-horizon trajectory prediction in autonomous vehicles. *Information Fusion*, 118:102924, 2025. 1
- [46] Yuning Wang, Pu Zhang, Lei Bai, and Jianru Xue. Fend: A future enhanced distribution-aware contrastive learning framework for long-tail trajectory prediction. In *CVPR*, pages 1400–1409, 2023. 3, 6
- [47] Pei Xu, Jean-Bernard Hayet, and Ioannis Karamouzas. Context-aware timewise vaes for real-time vehicle trajectory prediction. *IEEE Robotics and Automation Letters*, 2023. 6
- [48] Yi Xu and Yun Fu. Adapting to length shift: Flexilength network for trajectory prediction. In *CVPR*, pages 15226–15237, 2024. 6
- [49] Shiyu Xuan and Shiliang Zhang. Decoupled contrastive learning for long-tailed recognition. In *AAAI*, pages 6396–6403, 2024. 3, 5
- [50] Ye Yuan, Xinhao Weng, Yanglan Ou, and Kris M Kitani. Agentformer: Agent-aware transformers for socio-temporal multi-agent forecasting. In *ICCV*, pages 9813–9823, 2021. 7
- [51] Junrui Zhang, Mozghan Pourkeshavarz, and Amir Rasouli. Tract: A training dynamics aware contrastive learning framework for long-tail trajectory prediction. *arXiv preprint arXiv:2404.12538*, 2024. 3, 6
- [52] Qingwen Zhang, Yi Yang, Peizheng Li, Olov Andersson, and Patric Jensfelt. Seflow: A self-supervised scene flow method in autonomous driving. In *ECCV*, pages 353–369. Springer, 2025. 6