



HAL
open science

Interpretable Long Term Waypoint-Based Trajectory Prediction Model

Amina Ghoul, Itheri Yahiaoui, Fawzi Nashashibi

► **To cite this version:**

Amina Ghoul, Itheri Yahiaoui, Fawzi Nashashibi. Interpretable Long Term Waypoint-Based Trajectory Prediction Model. IEEE International Conference on Intelligent Transportation Systems (ITSC), 2023, Bilbao, Spain. hal-04329601

HAL Id: hal-04329601

<https://hal.science/hal-04329601>

Submitted on 7 Dec 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Interpretable Long Term Waypoint-Based Trajectory Prediction Model

Amina Ghoul¹, Itheri Yahiaoui², and Fawzi Nashashibi¹

Abstract—Predicting the future trajectories of dynamic agents in complex environments is crucial for a variety of applications, including autonomous driving, robotics, and human-computer interaction. It is a challenging task as the behavior of the agent is unknown and intrinsically multimodal. Our key insight is that the agents behaviors are influenced not only by their past trajectories and their interaction with their immediate environment but also largely with their long term waypoint (LTW). In this paper, we study the impact of adding a long-term goal on the performance of a trajectory prediction framework. We present an interpretable long term waypoint-driven prediction framework (WayDCM). WayDCM first predict an agent’s intermediate goal (IG) by encoding his interactions with the environment as well as his LTW using a combination of a Discrete choice Model (DCM) and a Neural Network model (NN). Then, our model predicts the corresponding trajectories. This is in contrast to previous work which does not consider the ultimate intent of the agent to predict his trajectory. We evaluate and show the effectiveness of our approach on the Waymo Open dataset.

I. Introduction

Predicting the future motion of a dynamic agent in an interactive environment is crucial in many fields and especially in autonomous driving.

A key challenge to future prediction is the high degree of uncertainty, in large part due to not knowing the behavior of the other agents. Because of these uncertainties, future motion of agents are inherently multimodal. Multimodality can be modeled by using implicit distributions from which samples can be drawn such as conditional variational autoencoders (CVAEs) [1] or generative adversarial networks (GANs) [2]. The uncertainties can also be captured by the prediction of possible intermediate goals of the agents before predicting the full trajectories [3], [4], [5].

Moreover, most works predict the trajectories of moving agents by considering their past trajectories and their dynamic and/or static environment [6].

However, we assume that before going anywhere, the agent knows its long-term waypoint, and its movement is therefore drawn to it. For example, we can consider the case of an agent entering his final destination on a GPS. The GPS then gives waypoints that need to be followed by the agent to reach its destination. In this paper, we study the impact of adding this waypoint information to our trajectory prediction model.

This work was carried out in the SAMBA collaborative project, co-funded by BpiFrance in the framework of the Investissement d’Avenir Program.
1: INRIA Paris, France `firstname.lastname@inria.fr` ;
2: CRESTIC, Université de Reims Champagne-Ardenne, Reims, France
`itheri.yahiaoui@univ-reims.fr`

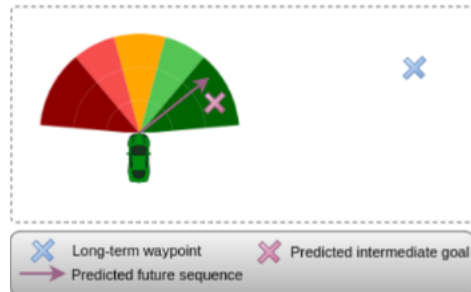


Fig. 1. The favorable directions are in green and the unfavorable directions in red. The cross in blue refers to the position of a long term waypoint of the vehicle. Knowing the position of this waypoint is an important cue for predicting its intermediate goal (the red cross). Here, the vehicle tends to go to right in order to get to reach the blue cross.

Furthermore, our model combines an interpretable discrete choice model with a neural network for the task of trajectory prediction. This combination allows for interpretable outputs. Our approach presents a way to easily validate NN models in safety critical applications, by using the interpretable pattern-based rules from the DCM.

In this paper, we introduce **WayDCM**, which builds upon our last work TrajDCM [7], but taking into account a long term waypoint of the agents and proposing a first way to introduce it in a trajectory prediction framework. We conduct extensive experimentations on the real-world Waymo open dataset and we demonstrate the effectiveness of our method.

The content is organized as follows: Section II provides a review on the background and current state of the research fields. Section III introduces the method of the proposed model. Section IV presents the different experiments conducted. Section V presents the performance of the approach for vehicle trajectory prediction and the comparison to baseline models. Finally, a conclusion regarding the architecture and results is drawn.

II. Related Work

A. Input Representation

In order to predict its trajectory, state of the art works use different forms of input cues that can reveal information about the future motion of the moving target agent.

The first main cue consists of the target agent’s past state observations (positions, orientation, velocities, acceleration, etc.). Most studies use a sequence of past features to exploit the characteristics of the motion variation in the prediction task [8], [9].

Then, the dynamic scene states represented by the past state observations of the surrounding agents are an other important cue, as their interaction with each other and with

the target agent has an influence on the target agent’s future motion.

Also, the static scene elements such as lanes and crosswalks, for example, help to determine the reachable areas and therefore the possible patterns of the target agent motion. Different state of the art works exploit static scene features and/or the dynamic ones [10], [4] to infer the motion of the target agent.

In addition, some recent works such as [11] consider the historical trajectories previously passing through a location as a new type of input cue in order to help infer the future trajectory of an agent currently at this location.

However, these approaches do not consider the long term waypoints that lead to the agent’s final destination as an input cue.

In our work, we propose a lightweight representation based on agents states and intermediate goals definition. We also add a long-term waypoint of the agent in order to better predict and understand the agent’s behavior.

B. Conditioning on Intermediate Goal

Conditioning on intermediate goal (IG) enhances the ability of trajectory prediction models to capture the influence of explicit IGs on agent behavior and improve the accuracy of trajectory forecasts. Several methods such as TNT [8] or LaneRCNN [12] condition each prediction on the intermediate goal of the driver. Conditioning predictions on future IGs helps leverage the HD map by restricting those IGs to be in a certain space. MultiPath [13] and CoverNet [14] chose to quantize the trajectories into anchors, where the trajectory prediction task is reformulated into anchor selection and offset regression. In this paper, we use a radial grid similar to [7], [15] to predict the intermediate goal of the moving agent.

C. Interpretable Trajectory Prediction

Most state-of-the-art studies use neural networks for the task of trajectory prediction. However, neural networks models are not interpretable. In fact, their outputs cannot be explained by humans. In order to adress the lack of interpretability, recent studies focus on adding expert knowledge to deep learning models for trajectory prediction. A way to encourage interpretability in trajectory prediction architectures is through discrete modes. Brewitt et al. [16] propose a Goal Recognition method where the “goal” is defined by many behavioral intentions, such as “straight-on”, “turn left” for example. In [17], the authors propose a method combining a social force model, and a method comprised of deep learning approaches based on a Neural Differential Equation model. They use a deep neural network within which they use an explicit physics model with learnable parameters, for the task of pedestrian trajectory prediction. [7] learn a probability distribution over possibilities in an interpretable discrete choice model (DCM) for the task of vehicle trajectory prediction.

We use a similar approach as [7] for predicting the trajectories of diverse agents. However, this paper differs

from our previous work as we propose a method to include the long term waypoint information of the agent in a DCM model.

D. Contributions

To the best of our knowledge, we are the first to introduce the idea of a long term waypoint to a trajectory prediction framework. In fact, most studies do not take into account this information as they mostly model the interactions between agents [18], [19] or/and with the static environment [10], [9]. However, we argue that knowing where the agent is going, is a prior knowledge that should be considered in order to accurately predict its future trajectory. Indeed, while moving, an agent is not only influenced by its interactions with its environment, but also by its destination.

To summarize, we list the main contributions of our approach as follows:

- We introduce an interpretable DCM-based model which includes a long term waypoint of the agent, for trajectory prediction.
- We study the impact of adding such information on the predicted trajectories.
- We implement and evaluate our approach on the Waymo Open Dataset. Our goal is to open the discussion about adding the long term waypoints of agents in motion prediction datasets. And also to propose a first approach that considers this new information.

III. Method

A. Problem definition

The goal is to predict the future trajectories of a target agent $T : \hat{Y}_T = (\hat{x}_T^t, \hat{y}_T^t)$ from time $t = t_{obs} + 1$ to $t = t_f$. We have as input of our model the track history of the target agent and the n neighboring agents in a scene defined as $\mathbf{X} = [X_1, X_2, \dots, X_n]$. Each agent i is represented by a sequence of its states, from time $t = 1$ to $t = t_{obs}$. Each state is composed of a sequence of the agent relative coordinates x_i^t and y_i^t , velocity v_i^t , heading θ_i^t .

$$X_i^t = (x_i^t, y_i^t, v_i^t, \theta_i^t) \quad (1)$$

The positions of each agent i are expressed in a frame where the origin is the position of the target agent at t_{obs} . The x-axis is oriented toward the target agent’s direction of motion and y-axis points to the direction perpendicular to it.

B. Utility function

Similar to [7], we use a discrete choice model to help predict the intermediate goal of the agent among a discrete set of potential goals. The DCM allows to model the behavior of agents in their interactions with their surroundings.

The intermediate goals are extracted from a dynamic radial grid where the longitudinal size of the grid $maxl$ depends on the velocity of the target agent at t_{obs} . So that, we have $maxl = 1.5 \times v_T^{t_{obs}} \times t_f$.

We use the Random Utility Maximization (RUM) theory [20] that postulates that the decision-maker aims at

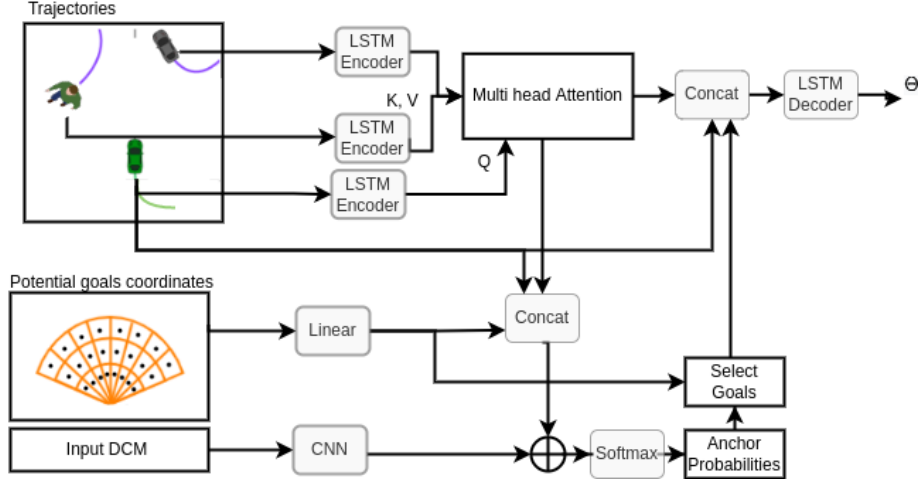


Fig. 2. Architecture of the method for trajectory prediction. The model takes as inputs the past trajectories of the agents in the scene, the target coordinates sampled from a radial grid, as well as the input of the DCM model. It outputs L trajectories. For more details see section III.

maximizing the utility relative to their choice. The utility that an agent i chooses an alternative k , is given as :

$$u_k(\mathbf{X}) = \beta_{dir}dir_k + \beta_{occ}occ_k + \beta_{col}col_k + \beta_{ddir}ddir_k + \beta_{ddist}ddist_k \quad (2)$$

where β_{dir} , β_{occ} , β_{col} , β_{ddir} and β_{ddist} are the parameters associated with the explanatory variables dir_k , occ_k , col_k , $ddir_k$ and $ddist_k$ that describe the observed attributes of the choice alternative.

In the following, we describe each of these explanatory variables :

1) Keep direction dir

This part of the model captures the tendency of people to avoid frequent variation of direction. Agents choose their position in order to minimize the angular displacement from their current direction of movement.

The variable dir_k is defined as the angle in degrees between the direction of the alternative k and the direction dn , corresponding to the current direction of the agent.

2) Occupancy occ

We consider that alternatives containing neighbours in the vicinity are less desirable. It is defined as the weighted number of agents being in the direction of the alternative k , that is :

$$occ_k = \sum_{i=1}^N I_k e^{-dist_{ik}} \quad (3)$$

where N is the total number of agents in the environment, I_k is one if the distance $dist_{ik}$ between the agent i and the physical center of the alternative k is less than $maxl/3$.

3) Collision avoidance $coll$

The insight for this function is that when a neighbour agent's trajectory is head-on towards a potential intermediate goal, this IG becomes less desirable due to the chance of a collision.

For each direction d , we choose a collider based on the following indicator function :

$$I_{coll}^i = \begin{cases} 1 & \text{if } d_l < d_i < d_r, \\ & \text{and } 0 < D_i < 2 \times max_l, \\ & \text{and } \frac{\pi}{2} < |\theta_i - \theta_k| < \pi, \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

where d_l and d_r represent the bounding left and right directions of the cone in the choice set while d_i is the direction identifying the position of agent i . D_i is the distance between agent i and the target agent. $|\theta_i - \theta_k|$ is the difference between the movement direction of agent i , and the direction of the alternative k . Among the set of K_d potential colliders for each direction, a collider C is chosen in each cone as the agent that has the maximum value of $|\theta_i - \theta_k|$. Finally, we have :

$$coll_k = \alpha_C e^{\rho_C D_C}, \quad (5)$$

where α_C and ρ_C are defined in [21].

4) Long term waypoint direction $dangle$

We consider that an agent has the tendency to choose, for the intermediate goal, a spatial location that minimizes the angular displacement to the long term waypoint.

$dangle_k$ is defined as the angle in degrees between the destination and the long term waypoint of the alternative k .

5) Long term waypoint distance $ddist$

Here we consider that the agents tend to minimize the distance to the long term waypoint. The variable $ddist_k$ is defined as the distance (in meters) between the long term waypoint and the center of the alternative k .

We expect a negative sign for all the β parameters.

C. Architecture

For a target agent T at time t , X_T^t is embedded using a fully connected layer to a vector e_i^t and encoded using an LSTM encoder,

$$h_i^t = LSTM(h_i^{t-1}, e_i^t; W_{enc}), \quad (6)$$

W_{enc} are the weights to be learned. The weights are shared between all agents in the scene.

Then we build a social tensor similar to [6].

We use the multi-head attention mechanism [22] to model the social interactions, where the target vehicle $h_T^{t_{obs}}$ is processed by a fully connected layer to give the query and the social tensor is processed by 1×1 convolutional layer to give the keys and the values.

We consider K attention heads where K attention heads are specialized to the K potential intermediate goals.

For each attention head, we concatenate the output of the multi-head attention module A_k with the target vehicle trajectory encoder state $h_T^{t_{obs}}$ to give a context representation z_k for $k = 1, \dots, K$.

$$z_k = \text{Concat}(h_T^{t_{obs}}, A_k) \quad (7)$$

We then predict the intermediate goal by combining a DCM model with a NN model.

In order to help the knowledge-based model DCM capture the long term dependencies and the complex interactions, we use the Learning Multinomial Logit (L-MNL) [23] framework.

The IG selection probabilities is defined as :

$$\pi(a_k | \mathbf{X}) = \frac{e^{s_k(\mathbf{X})}}{\sum_{j \in K} e^{s_j(\mathbf{X})}}, \quad (8)$$

where

$$s_k(\mathbf{X}) = u_k(\mathbf{X}) + z_k(\mathbf{X}), \quad (9)$$

where $s_k(\mathbf{X})$ represents the IG function containing the NN encoded terms, $z_k(\mathbf{X})$, as well as utility function $u_k(\mathbf{X})$, following the L-MNL framework.

The alternative k corresponds to the target agent's intermediate goal at timestep t_f , extracted from a radial grid, similar to [7]. We consider L attention heads, for each attention head, we concatenate the output of the multi-head attention module A_l with the target agent trajectory encoder state $h_T^{t_{obs}}$ to give a context representation c_l for $l = 1, \dots, L$.

$$c_l = \text{Concat}(h_T^{t_{obs}}, A_l) \quad (10)$$

We select the L best scored targets, and we concatenate their embedding to the output of the context representation c_l for $l = 1, \dots, L$.

Finally, the context vector c_l is fed to an LSTM Decoder which generates the predicted parameters of the distributions over the target vehicle's estimated future positions of each possible trajectory for next t_f time steps,

$$\Theta_l^t = \Lambda(\text{LSTM}(h_l^{t-1}, z_l; W_{dec})), \quad (11)$$

where W_{dec} are the weights to be learned, and Λ is a fully connected layer. Similar to [7], we also output the probability P_l associated with each mixture component.

D. Loss function

Our proposed model outputs the means and variances $\Theta_l^t = (\mu_l^t, \Sigma_l^t)$ of the Gaussian distributions for each mixture component at each time step.

The loss for training the model is composed of a regression loss L_{reg} and two classification losses L_{score} and L_{cls} . L_{reg}

is the negative log-likelihood (NLL) similar to the one used in [6], L_{score} is a cross entropy loss and L_{cls} is also a cross entropy loss defined as :

$$L_{reg} = -\min_l \sum_{t=t_{obs}+1}^{t_{obs}+t_f} \log(\mathcal{N}(y^t | \mu_l^t, \Sigma_l^t)). \quad (12)$$

$$L_{score} = -\sum_{l=1}^L \delta_{l^*}(l) \log(P_l), \quad (13)$$

where δ is a function equal to 1 if $l = l^*$ and 0 otherwise.

$$L_{cls} = -\sum_{k=1}^K \delta_{k^*}(k) \log(p_k), \quad (14)$$

where p_k is the probability associated with the potential goal k , δ is a function equal to 1 if $k = k^*$ and 0 otherwise, k^* is the index of the potential intermediate goal most closely matching the endpoint of the ground truth trajectory.

Finally, the loss is given by :

$$L = L_{cls} + L_{reg} + L_{score}, \quad (15)$$

IV. Experiments

A. Dataset

Many datasets have been proposed for the task of motion prediction [24], [25], [26], [27]. We choose to evaluate our approach on the Waymo Open dataset [28] as it has a longer prediction horizon compared to the rest of the frequently used datasets (See Table. I). In fact, for the long term waypoint of the target agent, we consider the position at the longest time horizon. Moreover, in the Waymo dataset, we can predict the future trajectories of agents for multiple horizons of 3, 5 or 8 seconds unlike the other datasets. We predict the future trajectories of a target agent for a horizon of $t_f = 3$ seconds while considering the long term waypoint at 8 seconds. Therefore, we are able to compare our methods with other models only by using Waymo dataset.

The Waymo dataset provides 103,354, 20s 10Hz segments (over 20 million frames), mined for interesting interactions. The data is collected from different cities in the United States of America (San Francisco, Mountain View, Los Angeles, Detroit, Seattle and Phoenix).

TABLE I

PAST, FUTURE AND PREDICTION HORIZONS ON DIFFERENT MOTION PREDICTION DATASETS

Dataset	Past (s)	Future (s)	Prediction (s)
nuScenes	3	5	5
INTERACTION	1	3	3
Argoverse 1	2	3	3
Argoverse 2	5	6	6
Waymo	1	8	3, 5 and 8

B. Implementation details

We predict the future trajectory of the target agent T for a horizon of $t_f = 3$ seconds. We consider the final destination as the position of the target agent at $t = 8$ seconds. We use $K = 15$ number of potential intermediate goals. The radial grid representation is dynamic, i.e it depends on the current velocity of the target agent. The interaction space is 40 m ahead of the target vehicle, 10 m behind and 25 m on each side. We consider the neighbors situated in the interaction space at t_{obs} . We use $L + K = 6 + 15$ parallel attention operations. We use a batch size of 16 and Adam optimizer. The model is implemented using PyTorch [29].

C. Compared Methods

The experiment includes a comparison of different models:

- LSTM : An LSTM model with past trajectory of the agents as input.
- MHA-LSTM [19]: This multi-head attention-based model takes as inputs the past trajectories of the agents in the scene and outputs L trajectories with their associated probabilities. We use $L = 6$ attention heads.
- MotionCNN [30] takes as input an image of the scene and uses a CNN followed by a fully connected layer to output multiple trajectories.
- MultiPath++ [31]: takes as input the agent state’s history and the road network and predicts a distribution of future behavior parameterized as a Gaussian Mixture Model (GMM).
- TrajDCM [7] : To predict the IG of the target agent, this model combines a DCM and a neural network model. It then predicts the corresponding full trajectories. The DCM model is composed of the three first functions described in III-B (*dir*, *occ* and *coll*).
- WayDCM (1) : The proposed model without the function *ddist* in the utility function.
- WayDCM (2) : The proposed model described in III.

V. Results

A. Evaluation metrics

In order to evaluate the overall performance of our method, we use the following two error metrics.

- **Minimum Average Displacement Error over k** ($minADE_k$) : The average of pointwise L2 distances between the predicted trajectory and ground truth over the k most likely predictions.
- **Minimum Final Displacement Error over k** ($minFDE_k$) : The final displacement error (FDE) is the L2 distance between the final points of the prediction and ground truth. We take the minimum FDE over the k most likely predictions and average over all agents.

B. Comparison of Methods

We compare the methods described in Section IV-C. We evaluated all the models on the validation set and we predicted the agent’s future motion for $t_f = 3$ seconds. We could not compare our approach with the models in the

Waymo leaderboard as our method needs the position at 8 seconds which is not given in the test set. Table II shows the obtained results. We can see that adding both the distance and the angle in the DCM model improve the original model TrajDCM. Furthermore, our method is more lightweight than Multipath++ and MotionCNN as it does not take into account any map information. We implemented these two models using their code available online. We can see that we outperform MotionCNN and we obtain comparable results for $MinADE_6$ against Multipath++.

TABLE II

COMPARISON OF DIFFERENT METHODS ON THE WAYMO VALIDATION SET (3 SECS HORIZON)

Model	$MinADE_6$	$MinFDE_6$
MotionCNN	0.3365	0.6145
MultiPath++	0.2692	0.4951
LSTM	0.4018	0.8029
MHA-LSTM	0.3141	0.7577
TrajDCM	0.3060	0.7201
WayDCM (1)	0.2779	0.6261
WayDCM (2)	0.2721	0.6037

C. Interpretable outputs

TABLE III

ESTIMATED PARAMETERS β

Model	β_{dir}	β_{col}	β_{occ}	β_{dangle}	β_{ddist}
TrajDCM	-2.70	-0.07	-0.06	-	-
WayDCM (1)	-3.48	-0.09	-0.04	-15.23	-
WayDCM (2)	-2.64	-0.05	-0.06	-10.83	-20.86

The estimated parameters of the utility functions Eq. 2 are reported in Table. III. We can see that the all of the coefficients β are negative. This is coherent with what we expected in Section III-B.

Moreover, we can see that the functions *ddist* and *dangle* contribute the most to the prediction of the IG as their absolute values are higher than the absolute values of the other parameters for the model WayDCM(2). The function *ddist* is the most significant. It means that agents tends to choose the IG which minimize the distance to the long term waypoint.

We can also notice that the functions *occ* and *coll* are not significant as their parameters for all the models are close to zero. This can be explained as a lack of interactions involving collisions for example. We plan to test our method on more interactive datasets in the future.

VI. Conclusion and future work

The main scope of this paper was not to propose a model that outperforms the state-of-the art. Instead, we proposed a new approach to solve the trajectory prediction problem. We showed that using a long term waypoint can help improve the prediction performance. We argue that long term waypoints should be included in motion prediction datasets as they are important input cues for the model. In fact, agents tend to move towards them. The position of these waypoints can be known before predicting the future trajectory of an agent, if given by a GPS for example.

We proposed a first simple approach that includes this information through a DCM. However, we hope that this paper will encourage researchers to consider the long term waypoints for the task of trajectory prediction.

For future work, we plan to try and compare other ways to include this information using different datasets, in a motion prediction framework.

References

- [1] N. Lee, W. Choi, P. Vernaza, C. B. Choy, P. H. Torr, and M. Chandraker, "Desire: Distant future prediction in dynamic scenes with interacting agents," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 336–345.
- [2] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social gan: Socially acceptable trajectories with generative adversarial networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2255–2264.
- [3] T. Gilles, S. Sabatini, D. Tsishkou, B. Stanciulescu, and F. Moutarde, "GOHOME: graph-oriented heatmap output for future motion estimation," *arXiv preprint arXiv:2109.01827*, 2021.
- [4] —, "THOMAS: trajectory heatmap output with learned multi-agent sampling," in *ICLR*, 2022.
- [5] H. Zhao, J. Gao, T. Lan, C. Sun, B. Sapp, B. Varadarajan, Y. Shen, Y. Shen, Y. Chai, C. Schmid, C. Li, and D. Anguelov, "TNT: target-driven trajectory prediction," in *4th Conference on Robot Learning, CoRL 2020*, ser. Proceedings of Machine Learning Research, vol. 155, 2020, pp. 895–904.
- [6] K. Messaoud, N. Deo, M. M. Trivedi, and F. Nashashibi, "Trajectory prediction for autonomous driving based on multi-head attention with joint agent-map representation," in *IEEE Intelligent Vehicles Symposium, IV 2021, Nagoya, Japan, July 11-17, 2021*. IEEE, 2021, pp. 165–170.
- [7] A. Ghoul, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "Interpretable Goal-Based model for Vehicle Trajectory Prediction in Interactive Scenarios," May 2023, working paper or preprint. [Online]. Available: <https://hal.science/hal-04108657>
- [8] H. Zhao, J. Gao, T. Lan, C. Sun, B. Sapp, B. Varadarajan, Y. Shen, Y. Shen, Y. Chai, C. Schmid *et al.*, "Tnt: Target-driven trajectory prediction," in *Conference on Robot Learning*. PMLR, 2021, pp. 895–904.
- [9] N. Deo, E. Wolff, and O. Beijbom, "Multimodal trajectory prediction conditioned on lane-graph traversals," in *Conference on Robot Learning (CoRL)*, 2022, pp. 203–212.
- [10] K. Messaoud, N. Deo, M. M. Trivedi, and F. Nashashibi, "Trajectory prediction for autonomous driving based on multi-head attention with joint agent-map representation," in *2021 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2021, pp. 165–170.
- [11] Y. Zhong, Z. Ni, S. Chen, and U. Neumann, "Aware of the history: Trajectory forecasting with the local behavior data," in *European Conference on Computer Vision*. Springer, 2022, pp. 393–409.
- [12] W. Zeng, M. Liang, R. Liao, and R. Urtasun, "LaneRCNN: Distributed representations for graph-centric motion forecasting," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 532–539.
- [13] Y. Chai, B. Sapp, M. Bansal, and D. Anguelov, "MultiPath: Multiple probabilistic anchor trajectory hypotheses for behavior prediction," in *CoRL*, 2019.
- [14] T. Phan-Minh, E. C. Grigore, F. A. Boulton, O. Beijbom, and E. M. Wolff, "CoverNet: Multimodal behavior prediction using trajectory sets," *CoRR*, vol. abs/1911.10298, 2019. [Online]. Available: <http://arxiv.org/abs/1911.10298>
- [15] P. Kothari, B. Siffringer, and A. Alahi, "Interpretable social anchors for human trajectory forecasting in crowds," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021, pp. 15 556–15 566.
- [16] C. Brewitt, B. Gjevvar, S. Garcin, and S. V. Albrecht, "Grit: Fast, interpretable, and verifiable goal recognition with learned decision trees for autonomous driving," in *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, pp. 1023–1030.
- [17] J. Yue, D. Manocha, and H. Wang, "Human trajectory prediction via neural social physics," in *Computer Vision—ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XXXIV*. Springer, 2022, pp. 376–394.
- [18] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social LSTM: Human trajectory prediction in crowded spaces," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016, pp. 961–971.
- [19] K. Messaoud, I. Yahiaoui, A. Verroust-Blondet, and F. Nashashibi, "Attention based vehicle trajectory prediction," *IEEE Transactions on Intelligent Vehicles*, vol. 6, no. 1, pp. 175–185, 2020.
- [20] C. F. Manski, "The structure of random utility models," *Theory and decision*, vol. 8, no. 3, p. 229, 1977.
- [21] G. Antonini, M. Sorci, M. Bierlaire, and J.-P. Thiran, "Discrete choice models for static facial expression recognition," in *Advanced Concepts for Intelligent Vision Systems: 8th International Conference, ACIVS 2006, Antwerp, Belgium, September 18-21, 2006. Proceedings 8*. Springer, 2006, pp. 710–721.
- [22] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, and I. Polosukhin, "Attention is all you need," *Advances in neural information processing systems*, vol. 30, 2017.
- [23] B. Siffringer, V. Lurkin, and A. Alahi, "Enhancing discrete choice models with representation learning," *Transportation Research Part B: Methodological*, vol. 140, pp. 236–261, 2020.
- [24] H. Caesar, V. Bankiti, A. H. Lang, S. Vora, V. E. Liong, Q. Xu, A. Krishnan, Y. Pan, G. Baldan, and O. Beijbom, "nuScenes: A multimodal dataset for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR)*, 2020, pp. 11 621–11 631.
- [25] W. Zhan, L. Sun, D. Wang, H. Shi, A. Clausse, M. Naumann, J. Kummerle, H. Königshof, C. Stiller, A. de La Fortelle *et al.*, "Interaction dataset: An international, adversarial and cooperative motion dataset in interactive driving scenarios with semantic maps," *arXiv preprint arXiv:1910.03088*, 2019.
- [26] M.-F. Chang, J. W. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan, and J. Hays, "Argoverse: 3d tracking and forecasting with rich maps," in *Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.
- [27] B. Wilson, W. Qi, T. Agarwal, J. Lambert, J. Singh, S. Khandelwal, B. Pan, R. Kumar, A. Hartnett, J. K. Pontes, D. Ramanan, P. Carr, and J. Hays, "Argoverse 2: Next generation datasets for self-driving perception and forecasting," in *Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks (NeurIPS Datasets and Benchmarks 2021)*, 2021.
- [28] S. Ettinger, S. Cheng, B. Caine, C. Liu, H. Zhao, S. Pradhan, Y. Chai, B. Sapp, C. R. Qi, Y. Zhou, Z. Yang, A. Chouard, P. Sun, J. Ngiam, V. Vasudevan, A. McCauley, J. Shlens, and D. Anguelov, "Large scale interactive motion forecasting for autonomous driving: The waymo open motion dataset," in *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, October 2021, pp. 9710–9719.
- [29] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga *et al.*, "Pytorch: An imperative style, high-performance deep learning library," *Advances in neural information processing systems*, vol. 32, 2019.
- [30] S. Konev, K. Brodt, and A. Sanakoyeu, "Motioncnn: a strong baseline for motion prediction in autonomous driving," *arXiv preprint arXiv:2206.02163*, 2022.
- [31] S. Konev, "Mpa: Multipath++ based architecture for motion prediction," 2022. [Online]. Available: <https://arxiv.org/abs/2206.10041>