GooDFlight: Goal-oriented Diffusion Model for Flight Trajectory Prediction

Songru Yang, Liqin Liu, Bowen Chen, Shaoyuan Cheng, Zhenwei Shi, *Senior Member, IEEE*, and Zhengxia Zou*, *Member, IEEE*

***Abstract*—Flight trajectory prediction is an essential task in the air traffic control field. Previous approaches to this problem usually follow a single-stage or short-term intention-guided prediction paradigm, which suffers from problems such as insufficient trajectory prediction diversity, limited accuracy and interpretability. Different from existing paradigms, in this paper, we present GooDFlight - A Goal-oriented Diffusion Model for Flight trajectory prediction. GooDFlight is a long-term intention-guided, diversity-emphasizing framework that decouples the flight trajectory prediction process into two stages: goal estimation and trajectory prediction. In the first stage, we propose a One-then-all goal estimation method to sufficiently cover the macro-uncertainty in flight patterns and then tailor the interaction-aware joint goal distribution, which extends the flight intention from a single, deterministic ground truth to the empirical intention distribution from the similar experience. In the second stage, we employ a transformer-based diffusion model to generate stochastic flight trajectories conditioned on the intention estimations, modeling the micro-uncertainty in flight operations under each pattern estimated in stage one. In terms of evaluation metrics, existing metrics have difficulties in accurately reflecting the model's ability to handle the natural uncertainty of trajectories. We further propose a simple yet effective Global-local endpoints Variance (GLeV) metric for evaluating the diversity of predicted trajectories under social acceptance. Our proposed method is validated in-depth on TrajAir, a large-scale dataset collected from the real-world air traffic control environment at the Pittsburgh-Butler Regional Airport, a non-towered general aviation airport. The experimental results demonstrate that the proposed method significantly outperforms other methods in terms of both accuracy and diversity with superior interpretability.**

***Index Terms*—Flight trajectory prediction, Intention estimation, Diffusion Model, Traffic scene understanding.**

I. Introduction

FLIGHT Trajectory Prediction is one of the fundamental research problems in the air traffic control system, which is crucial for various applications such as traffic flow prediction [1], delay prediction [2], conflict detection [3], [4], gathering increasing research interest. At the same time, general aviation (GA) is also experiencing rapid development. With the development and popularization of private aircraft and unmanned aerial vehicles, terminal airspace is becoming increasingly crowded and dangerous.

According to the data from General Aviation Safety Assurance (GASA) and Federal Aviation Administration (FAA) [5], more than 90% of the civil aircraft in the United States are GA aircraft, which caused approximately 36 accidents for every commercial airline accident in 2015, and 77% of non-fatal GA accidents are in terminal airspace where all flights typically begin and end in airspace surrounding airports [6]. One of the main reasons for such a high accident rate is that only 4% of active airports in the United States are towered, meaning that most general aircraft takeoffs and lands without a centralized authority ensuring separation [7]. Despite the fact that accidents continue to occur and increase, research on flight navigation in non-tower terminal airspace is far from sufficient, which has prompted the development of research related to trajectory prediction for non-tower airports.

Trajectory prediction is a long-lasting yet challenging task in commercial and general aviation. From kinetic and state estimation based methods to learning-based methods, researchers have made great efforts to achieve accurate trajectory prediction. Kinetic-based methods [8] establish a series of aerodynamics equations to model a deterministic flight pattern without diversity. State estimation-based methods [9], [10] are manually designed to estimate the state transition, which can only offer limited diversity when combined with uncertainty representation or multi-model aggregation. Recently, with the rapid development of data-driven methods, a large number of learning-based prediction methods have emerged [11], [12], leveraging deep neural network architecture to directly learn the implicit representations of the flight process from a large amount of historical flight trajectory data [13]–[15]. For GA navigation in non-tower terminal airspace, existing learning-based methods focus more on social-aware mechanisms and generative models for adapting to the complex interactions and the high dynamics [14], [16].

Despite the research efforts made in this field, previous methods still suffer from insufficient coverage of flight patterns, although they have achieved accurate predictions on normal trajectories, they can hardly handle relatively rare or abnormal maneuvers, which are critical to flight safety. This lack of diversity is mainly rooted in the "accuracy-first" con-

The work was supported by the National Key Research and Development Program of China under Grant 2022ZD0160401, the National Natural Science Foundation of China under Grant 62471014 and 62125102, the Beijing Natural Science Foundation under Grant JL23005, and the Fundamental Research Funds for the Central Universities. *(Corresponding author: Zhengxia Zou (e-mail: zhengxiazou@buaa.edu.cn))*

Liqin Liu, and Zhenwei Shi are with the Image Processing Center, School of Astronautics, Beihang University, and with the Beijing Key Laboratory of Digital Media, Beihang University, and with the State Key Laboratory of Virtual Reality Technology and Systems Beihang University, Beijing 100191, China.

Shaoyuan Cheng is with the Beijing Institute of Space Mechanics and Electricity, Beijing 100076, China.

Songru Yang, Bowen Chen, Zhengxia Zou are with the Department of Guidance, Navigation and Control, School of Astronautics, Beihang University, Beijing 100191, China.

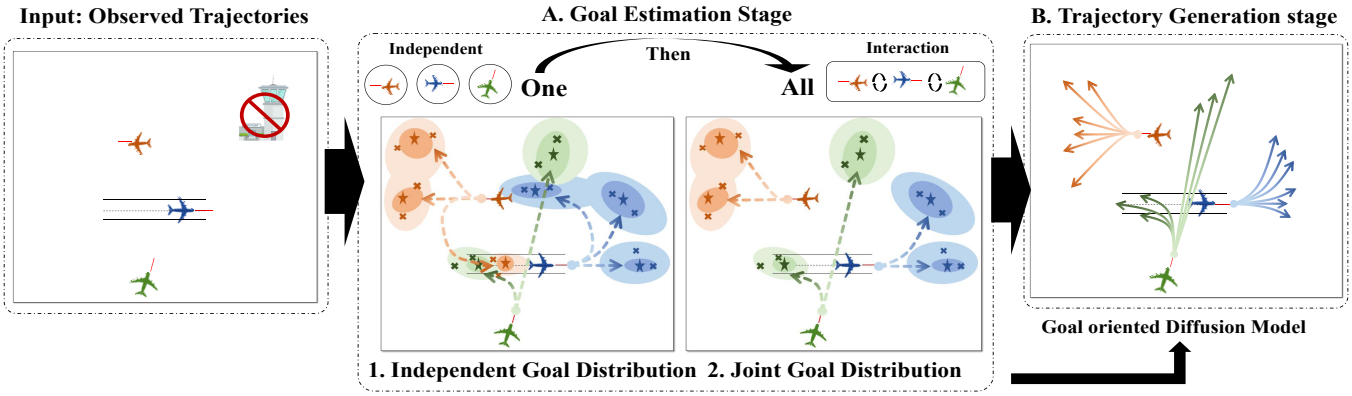

Fig. 1. The overall idea diagram of our proposed GooDFlight framework. The GooDFlight takes only observed trajectories as input, a two-stage framework including the goal estimation stage with a One-then-all paradigm (shown in A) and the trajectory generation stage based on the diffusion model (shown in B) are designed and composed in an end-to-end training pipeline to generate diverse and accurate flight trajectory predictions in terminal airspace of non-towered airports.

cept in the previous research. The trajectories possess inherent diversity, and the different movement patterns contained within them are imbalanced. However, previous methods relied solely on a single ground truth as the entire label for trajectory prediction or only used random noise in generative models to handle the uncertainty of trajectories, which limited the diversity of trajectory predictions to the most typical scenarios, and in turn eventually reduced the accuracy and interpretability of the predictions.

To tackle the flaws above, in this paper, we propose a Goal-oriented Diffusion Model (GooDFlight), a new trajectory prediction framework aiming to capture the rich diversity of the flight trajectories and at the same time improve the prediction accuracy and interpretability. We decouple the flight trajectory prediction task into two stages: a goal estimation stage and a trajectory generation stage, which are integrated into an end-to-end training pipeline. Fig. 1 shows the overall idea of the proposed method. In GooDFlight, the Goal estimation stage first predicts diverse future goals of observed aircraft flight that cover all possible flying patterns, modeling the macro-diversity in flight intentions. Then, in the trajectory generation stage, a diffusion model based trajectory generator takes in both the observed trajectory to generate the future diverse trajectories conditioned on estimated goals, which capture the micro-diversity in flight trajectories.

The motivation for our approach stems from the fact that the flight patterns are usually severely imbalanced. This makes it difficult for existing learning-based methods to capture rare events in flight trajectory data, which are critical for safe flight. To fully cover the diversity of GA flights, in the goal estimation stage, we design a One-then-all goal estimation method composed of two consecutive modules: an Independent goal distribution predictor, and a Joint goal distribution cropper. The Independent goal distribution predictor takes in the observation trajectory of each aircraft as input and outputs a mixture of gaussians to fit the empirical intention distribution prior by gathering the previous similar flight goal distribution of these aircraft as pseudo-labels. After obtaining the independent goal distribution for every single aircraft, social-aware information is considered in the following Joint goal distribution cropper, which takes independent goal distributions as input, then adapt it to the current interaction through graph attention with dynamic weights learning, emphasizing the situation of each aircraft. This novel intention estimation method introduces the distribution of experienced intentions before considering interactions, ensuring that the model considers the interaction on a full set of possible intentions, which significantly improves its prediction diversity compared to intention prediction based solely on the observed situation.

Shouldering the task of outputting probabilistic flight trajectories for each goal condition, the trajectory generation stage is built on a transformer-based diffusion generator customized for trajectory prediction tasks. In this stage, we model the trajectory generation as a denoising diffusion process and predict future movements based on observed trajectories and goals. The predicted trajectories are then obtained by integrating the actions over the observed trajectories.

Compared with the direct use of generative models in existing methods to model all trajectory uncertainties based merely on noise input, we spread the burden of macroscopic uncertainty modeling to the goal estimation stage and only use the diffusion model to generate micro-uncertainties, which greatly improves the diversity and plausibility of the predicted flight patterns. At the same time, from the perspective of interpretability, unlike existing implicit modeling methods, our method explicitly constructs multiple distributions, including the independent target distribution of a single aircraft, the joint target distribution of multiple aircraft, and the trajectory of each estimated target. This macro-then-micro and One-then-all explicit modeling provides our framework with stronger interpretability.

In our experiment, to measure the diversity of candidate trajectories on the premise of social acceptance, we propose a new evaluation metric named Global-local endpoint Variance (GLeV). GLeV calculates the ratio of the variance of trajectory endpoints falling around the ground truth endpoint to the variance of the endpoint of all trajectories. Instead of merely focusing on similarity to single ground truth, we strive to take

the diversity of features within the candidate trajectories into consideration and encourage more diverse trajectory predictions while ensuring sufficient accuracy.

Extensive experiments are conducted on TrajAir - a large-scale trajectory dataset collected from the real-world non-towered terminal airspace at Pittsburgh-Butler Regional Airport. Quantitative and qualitative experimental results show that the proposed method significantly outperforms the baseline methods on prediction accuracy and diversity, which has achieved improvements of +48.21%/+45.83% on the Average Displacement Error (ADE) and the Final Displacement Error respectively, and an increase of +80.00% on our own proposed GLeV metric. At the same time, in terms of inference speed, although we use a diffusion model, our inference time on the 120s prediction horizon is less than 1s, which fully meets the real-time requirements.

Our contributions can be summarized as follows:

- We propose a novel method named GooDFlight for trajectory prediction in non-towered terminal airspace. Our two-stage strategy decouples and models explicitly the intention diversity and the movement diversity of flight to accommodate unbalanced flight patterns, which enables the predicted trajectory with sufficient diversity, more accuracy, and higher interpretability.
- We propose a "One-then-all" goal estimation strategy, which models the complete macro pattern distribution of flights in the empirical distribution and dynamically adjusts according to the interaction between each aircraft. This strategy enables our method to fundamentally model complex interaction logic and diversity prediction.
- We propose GLeV, a new evaluation metric to evaluate the diversity of candidate trajectories on the premise of social acceptance. Extensive experiments on a real-world dataset (TrajAir) demonstrate that our approach outperforms previous methods in terms of both diversity and accuracy.

II. Related Work

A. Vehicle Trajectory Prediction

Vehicle Trajectory prediction is a fundamental research problem in autonomous driving. Earlier research of this topic mainly adopted the deterministic, precise-aware paradigm by exploring physics-based methods [17], [18], Kalman Filtering methods [19], [20], and RNNs [21]–[26]. However, deterministic predictions struggle to capture the inherent indeterminacy of future behavior, resulting in inaccurate predictions. To capture the diversity in the future distribution, recent work focused on stochastic trajectory prediction, which has derived two main frameworks: Deep generative framework and anchor conditioned framework. In deep generative framework, high-quality trajectory prediction is usually guaranteed by deep generative models, such as Generative Adversarial Network (GAN) [27]–[30] and Conditional Variational Autoencoder (CVAE) [31]–[34]. The Anchor-conditioned framework first predicts destinations as anchors and then plans the trajectories to them [35]. The methods in [36]–[39] generate scene-compliant endpoints by considering scene information such as center lanes and scene segmentation image.

The anchors as a form of condition, have another effective acquisition form: obtaining it from historical trajectories. Such methods not only encode information from the current scene but also introduce knowledge from historical scenes in the time dimension as conditions. [40]–[42] suggest that endpoints can be obtained by instance-based sampling from existing trajectory repositories. [43] clusters historical trajectories into template trajectories and then refine the template trajectories according to scene interactions. [44] further integrates historical behaviors into scene graph representations for learning and distills the models into a history-free form to adapt to the environment lacking historical information.

Unlike the above-mentioned methods for vehicle trajectory prediction, the flight trajectory prediction problem studied in this paper is more challenging due to the aircraft's high speed, complex motion patterns, and lack of environmental constraints.

B. Flight Trajectory Prediction

Existing aircraft trajectory prediction work is mainly divided into long-term predictions in high-altitude en-route airspace, and short-term predictions for terminal airspace. En-route trajectory prediction is usually single agent scenario, physics-based methods [45], [46], Kalman Filtering methods [47]–[49] and Learning-based methods [50]–[54], are employed to model flight patters under weather conditions. For terminal airspace, early work has focused on larger airports with commercial aviation (CA) traffic. Due to strict enforcement of FAA guidance, CA trajectory prediction methods implicitly learn determined takeoff and landing criteria using a simple deep-learning model. Recently, GA trajectory prediction in non-towered terminal airspace has become a frontier research field. Although GA traffic in terminal airspace is known to follow the FAA guidelines called "Airfield traffic pattern" [55] while having a much more complex socially compliant behavior. Patrikar et. al [14] published the first trajectory dataset: TrajAir and proposed the first baseline TrajAirNet on this task, introducing a graph attention for interaction modeling and a CVAE for trajectory generation. Further research in this field such as Social-PatteRNN [16] uses iterative short-term patterns series to estimate social influences and guide long-term trajectory predictions. More recently, for navigation with both pilot and controller, [56] adopted a context-aware (landing time, direction, etc) diffusion model for accuracy trajectory generation. Based on that, [57] self-adaptively retrieves closer historical local knowledge to implicitly assist the encoding of trajectories, significantly reducing the prediction uncertainty. While research progress has been made in determinacy, the inherent limitation of insufficient coverage of flight patterns and the natural uncertainty of human intention remain untouched. In this paper, we propose a One-then-all goal estimation method, leveraging a balanced pseudo label to model the empirical intention distribution and fit it to current interactions, ensuring the comprehensiveness of predictions.

C. Diffusion Models

Diffusion models [58] are a powerful class of generative models, capable of generating high-quality samples through

an iterative denoising process. Diffusion Models have recently achieved remarkable results in the fields of image and video generation [59], [60], weather forecasting [61], remote sensing [62]–[66], etc. Diffusion models have also been introduced to trajectory prediction recently. MID [67] is the first diffusion-based method for trajectory prediction by modeling the indeterminacy variation process into a process from an ambiguous walkable region to the desired trajectory. Later, LeapFrog [68] distill many denoising steps into an initializer to reduce the time costs in MID frameworks. More recently, GIMTP [69] designed a Diffusion Graph Convolutional Network (DGCN) to capture various interactions effectively. IDM [70] modeled the original uncertainty and intention uncertainty with two sequential diffusion processes and revealed that introducing intention information is beneficial in modeling the diffusion process of fewer steps. Inspired by these insights, we used the estimated goal as the intention information to build a trajectory generator based on the diffusion models [71] to capture the uncertainty of flight operations towards each goal.

III. Methodology

A. Problem Formulation

In this paper, we focus on the short-term flight trajectory prediction task, predicting an aircraft's future trajectory from the observed flight trajectory of itself and surrounding aircraft. Table I summarizes the main variable symbols and their meanings in this paper. Let $\mathbf{X}$ denote the observed aircraft trajectories in terminal airspace and $\mathbf{X} = [X_1, X_2, \ldots, X_l] \in \mathbb{R}^{l\times T\times 3}$, where l is the total number of aircraft, X_i is the observed past flight of the ith aircraft over T timestamps, which can be noted as $X_i = [x^{-T+1}, x^{-T+2}, \ldots, x^0] \in \mathbb{R}^{T\times 3}$, where $x^t = [x, y, z]$ denotes the 3D spatial coordinate at timestamp t. The corresponding ground truth future trajectory for the to-be-predicted aircraft is $Y_i = [y^1, y^2, \ldots, y^{T_f}] \in \mathbb{R}^{T_f\times 3}$, where $y^t = [x, y, z]$ is the 3D coordinate at future timestamp t. Then the future trajectories in terminal airspace are denoted as $\mathbf{Y} = [Y_1, Y_2, \ldots, Y_l] \in \mathbb{R}^{l\times T_f\times 3}$.

Further, we conduct stochastic trajectory prediction via a two-stage structure. And we denote the goal estimations of all aircraft in terminal space as $\hat{\mathbf{G}} = [\hat{G}_1, \hat{G}_2, \ldots, \hat{G}_l] \in \mathbb{R}^{l\times k\times 3}$ where the estimated k goals for ith aircraft are denoted as $\hat{G}_i = [\hat{g}_0, \hat{g}_1, \ldots, \hat{g}_k] \in \mathbb{R}^{k\times 3}$, where $\hat{g}_i = [x, y, z]$ is the 3D coordinate of ith goal sample. Then for each $\hat{g}_i$ of the to-be-predicted aircraft, a trajectory is generated as $\hat{Y} = [\hat{y}^1, \hat{y}^2, \ldots, \hat{y}^{T_f}] \in \mathbb{R}^{T_f\times 3}$, where $\hat{y}^t = [x, y, z]$ is the 3D coordinate at future timestamp t. For all k estimated goals, k trajectories for ith aircraft are generated as $\hat{\mathcal{Y}}_i = [\hat{Y}_1, \hat{Y}_2, \ldots, \hat{Y}_k] \in \mathbb{R}^{k\times T_f\times 3}$ to make sure at least one prediction is close enough to the ground truth, and for all aircraft in the terminal space, the predicted trajectories are $\hat{\mathbf{Y}} = [\hat{\mathcal{Y}}_1, \hat{\mathcal{Y}}_2, \ldots, \hat{\mathcal{Y}}_l] \in \mathbb{R}^{l\times k\times T_f\times 3}$. Our overall aim is to jointly optimize the parameters W_g^* of goal estimation stage $\mathcal{F}_{ge}(\cdot)$ and the parameters W_t^* of trajectory generation stage $\mathcal{F}_{tg}(\cdot)$:

$$\begin{aligned} \hat{\mathbf{G}} &= \mathcal{F}_{ge}(\mathbf{X}; W_g^*), \\ \hat{\mathbf{Y}} &= \mathcal{F}_{tg}(\mathbf{X}, \hat{\mathbf{G}}; W_t^*) \end{aligned} \tag{1}$$

such that the average and endpoint prediction error for all aircraft in the current terminal airspace is minimized.

TABLE I
Summary of Main Notations

Notation	Description
$\mathbf{X}$	Observed trajectories
$\mathbf{Y}$	Future trajectories
$\hat{\mathbf{G}}$	Estimated goals
$\hat{\mathbf{Y}}$	Predicted trajectories
$\mathbf{G}$	Pseudo-labels of empirical intention
l	Number of agents in terminal space
k	Number of predicted goals and trajectories
N_f	Number of the pseudo-labels for an aircraft
$f_{Ie}(\cdot)$	Temporal encoder for the Independent goal distribution predictor
$\mathbf{V}_t$	Temporal feature of aircraft
$f_\mu(\cdot)$, $f_\sigma(\cdot)$, $f_{mr}(\cdot)$	Predictor for μ,σ and mixing rate of empirical intention
D_I	The distribution of independent goals
A_{ij}	Adjacency between ith and jth aircraft
$\mathbf{h}_m, \mathbf{h}_t, \mathbf{h}_s$	Meta information, temporal feature and interaction feature of graph attention networks
$\mathbf{u}_{temporal}, \mathbf{u}_{social}$	Temporal feature and interaction feature for trajectory generation stage
$f_\epsilon(\cdot)$	Noise estimation network for diffusion model
$\hat{\mathbf{a}}_k^\gamma$	The kth action series at denoising step γ
ϵ_θ^γ	Noise prediction at denoising step γ

B. The proposed GooDFlight Framework

In this section, we elaborate on our proposed Goal oriented Diffusion Model (GooDFlight) framework. As discussed in Section I, the uneven distribution of flight patterns makes existing methods almost "blind" to patterns other than hovering. To fill this flaw, by linking balanced previous intentions and future movement uncertainty, we design the One-then-all goal estimation method to capture the complete goal distribution for different possible future flight patterns. Further, based on the goal estimations, a novel framework, called GooDFlight is proposed to enhance the performance of flight trajectory prediction tasks in non-towered terminal airspace.

An overview of the GooDFlight framework is illustrated in Fig. 2, which is an anchor-conditioned two-stage model, with a goal estimation stage consisting of Independent goal distribution predictor (Part III-C and III-D) and Joint goal distribution cropper (Part III-E), and a trajectory generation stage consisting of a intention adjustable diffusion model (Part III-F). Specifically, inside the first stage, after the empirical intention distribution is processed as pseudo-labels for independent goal distribution, the Independent goal distribution predictor firstly projects the observed trajectory into a high-level trajectory representation to generate a mixture of gaussian distributions fitting the pseudo-labels, in which the independent goals are sampled. In succession, the Joint goal distribution cropper assigns probabilities for independent goal samples of each

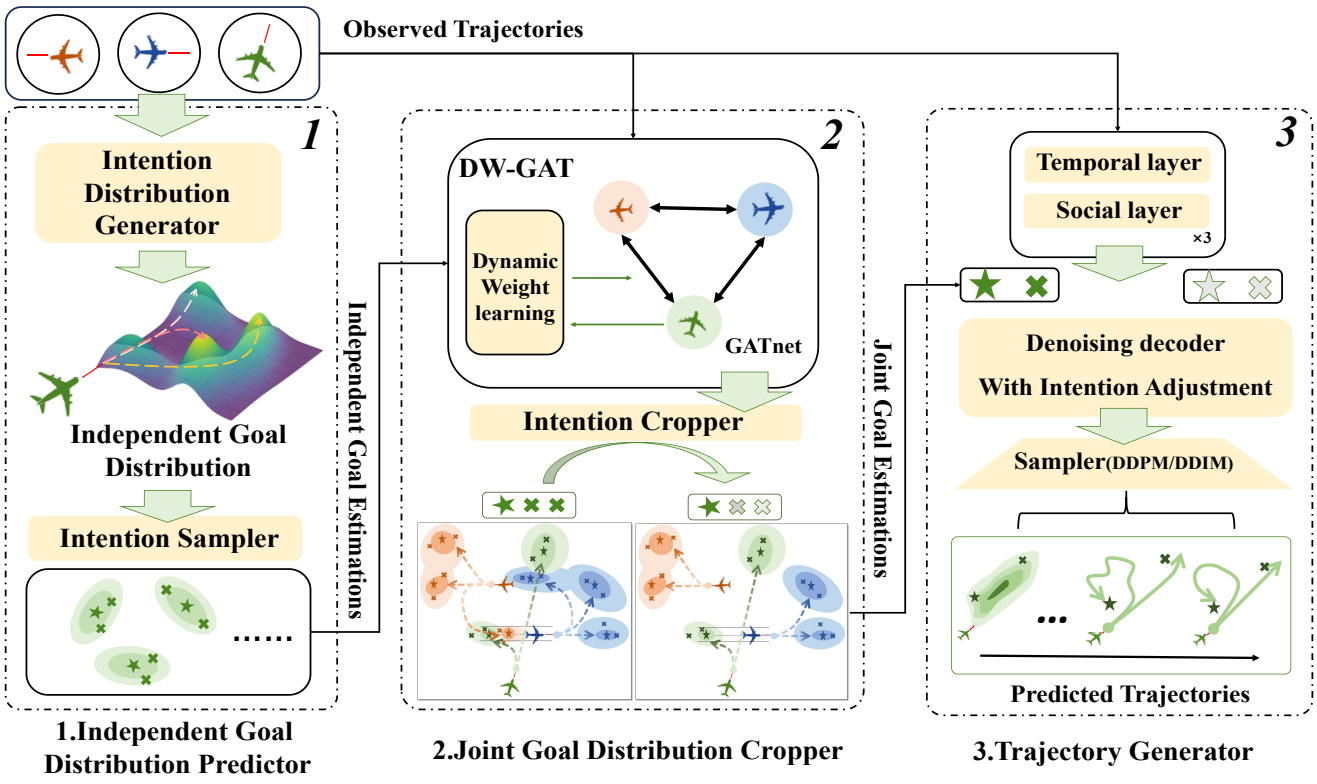

Fig. 2. The overall structure of proposed GooDFlight framework. Independent goal distribution predictor (1), Joint goal distribution cropper (2), and Trajectory Generator (3) are displayed from left to right. The Independent goal distribution predictor generates all possible future flight patterns in the form of goals for each aircraft according to previous flights. Then the independent goal estimations are tailored to the joint goal distribution, filtering out the unreachable goals. Finally, the trajectory generator equipped with a denoising decoder and Intention adjustment mechanism takes the joint goal estimation as pattern conditions and generates multimodal flight trajectory predictions in the reverse diffusion process.

aircraft from independent goal predictions and observed trajectory. By pruning low-probability goals, independent goal distribution is adjusted into joint goal distribution. Social-committing goals are then input with the observed trajectory into a condition-adjustable diffusion model to generate future trajectories through a 10-step denoising process. It is worth noting that an Intention adjustment mechanism based on classifier-free guidance is added since we deem goals as pattern indicators rather than destinations, whose confidence can be adjusted.

C. Pattern-balanced Pseudo-label Generation

Before the Independent goal distribution predictor is applied to link the past and future, we need to model the empirical intention distribution first. Drawing on the instance-based anchor generation method [41], we search for similar patterns in previous flight in an observed-future hash structure and induce it as prior knowledge. For each to-be-predicted aircraft in terminal airspace, the observed trajectory X_i is firstly regarded as a key for gathering the flights with previous similar patterns, then we filter out identical samples in each corresponding pattern to form empirical intention distribution and finally use k-means to cluster it as D_e.

While generating the pattern-balanced pseudo-labels, although we follow the same procedures, we make certain changes to make the model easy to learn. First, we only take account of temporal features without social influence, for we are modeling all possible movements for aircraft under all kinds of interactions. Second, we only use the feature similarity of the observed trajectories as a measure of pattern similarity rather than the proximity of the endpoint [41], since finding previous flights with close endpoints hurts covering more modalities. Third, we perform rarity-adaptive data augmentation on rare flight patterns, which are usually violations of FAA guidance.

For an observed trajectory X_i and corresponding empirical intention distribution with N_f cluster center coordinates $D_e = [c_0, c_1, \ldots, c_{N_f}] \in \mathbb{R}^{N_f \times 3}$, where $c_i = [x, y, z]$, we define a rarity r based on the distance to the ground truth endpoint $Y_i^{T_f} = [x, y, z]$, and use $\min(r, N_f/2)$ rare endpoints with small perturbation to replace ordinary endpoints, and then copy these samples for $\min(r, M_{cp})$ times, where M_{cp} is artificially set. Mathematically, rarity is calculated as follows:

$$r = \min_{i \in N_f} \|c_i - Y_i^{T_f}\|_2 \tag{2}$$

Through the above augmentation, the pattern-balanced pseudo-labels representing empirical intention distribution of ith aircraft are established as $G_i = [p_0, p_1, \ldots, p_{N_f}] \in \mathbb{R}^{N_f \times 3}$, where $p_i = [x, y, z]$ is the coordinates of ith pseudo-

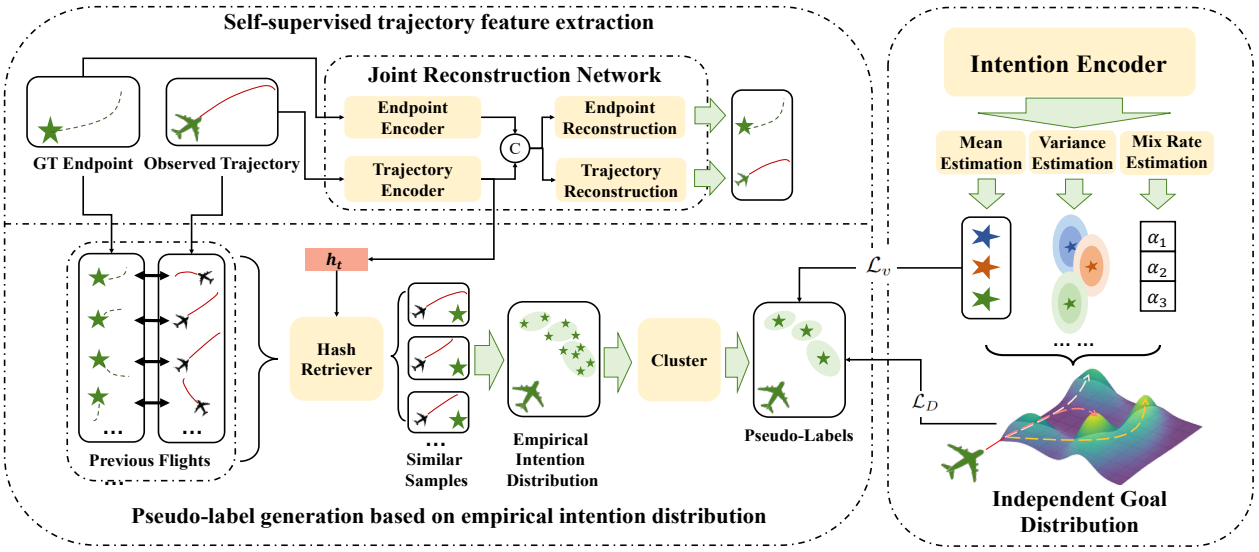

Fig. 3. Pseudo-labelling generation process and the structure of Independent goal distribution predictor. In the Pseudo-label generation process displayed on the left, the observed trajectories are jointly reconstructed with future endpoints to create a motion representation with past and future links, which we use to retrieve similar samples from previous flights and cluster their endpoints as pseudo-labels. The joint reconstruction network is pre-trained for this process and its data stream is shown as a blue dotted line, while the process of constructing the observed trajectory-future endpoint data pairs for previous flights is shown using a yellow solid line, and the data stream for the pseudo-labeling generation process is shown using a black solid line. In the Independent goal distribution predictor displayed on the right, the observed trajectories are taken as input, encoded by an intention encoder, the output is the independent goal distribution under a mixed gaussian prior, the components of which are predicted by different heads.

label and pseudo-labels of all traffic participants can be denoted as $\mathbf{G} = [G_1, G_2, \ldots, G_l] \in \mathbb{R}^{l \times N_f \times 3}$.

D. Independent goal distribution predictor

As shown in Fig. 3, The Independent goal distribution predictor is composed of a temporal encoder and four trainable heads which are responsible for predicting means, standard deviation, mixing rate, and conducting goal sampling respectively.

Building a bridge between the previous experience and the future movement goals, the Independent goal distribution predictor is designed to learn with pseudo label P_i for generating a goal distribution of an aircraft merely considering the temporal feature of its observed trajectory X_i. To fit the multi-cluster distribution in P_i, a gaussian mixture distribution is employed to be the prior of the prediction. Let $f_{Ie}(\cdot)$ be the temporal encoder, $f_{\mu}(\cdot)$, $f_{\sigma}(\cdot)$, $f_{mr}(\cdot)$ to be the three heads predicting the mean, standard deviation and mixing rate of a N_f gaussian components which are denoted as μ_θ, σ_θ, α_θ respectively, and $f_{\hat{\mathbb{S}}}(\cdot)$ to be the sampler for Independent goal distribution $\hat{G}_i = [\hat{g_0}, \hat{g_1}, \ldots, \hat{g_k}]$. The mathematical representation of the process is as follows:

$$\begin{aligned} v_t &= f_{Ie}(X_i) \in \mathbb{R}_h^d, \\ \mu_\theta &= f_{\mu}(v_t) \in \mathbb{R}^{N_f \times 3}, \\ \sigma_\theta &= f_{\Sigma}(v_t) \in \mathbb{R}^{N_f \times 3}, \\ \alpha_\theta &= f_{mr}(v_t) \in \mathbb{R}^{N_f}, \\ \hat{G}_i &= f_{\hat{\mathbf{S}}}(v_t, \mu_\theta, \sigma_\theta, \alpha_\theta) \in \mathbb{R}^{k \times 3}, \end{aligned} \tag{3}$$

To be specific, the temporal encoder $f_{Ie}(\cdot)$ encodes the observed trajectories X_i into a latent representation v_t, whose dimension is d_h, then v_t is delivered into $f_{\mu}(\cdot)$, $f_{\sigma}(\cdot)$ and $f_{mr}(\cdot)$ to generate μ_θ, σ_θ, α_θ for gaussian mixture distribution, in which each component stands for a pseudo label, reflecting the distribution of previous similar flight intentions. Finally, a learnable sampler $f_{\hat{\mathbb{S}}}(\cdot)$ is proposed to differentiably sample k/N_f independent goals from each component. The independent goal predictions are obtained by applying the calculated displacements to the mean position in Eq.3.

To optimize this Independent goal distribution predictor, we assemble the generated gaussian mixture distribution $D_I(\mu_\theta, \mu_\theta, \alpha_\theta)$, and maximize the probability of the pseudo-labels.

$$P(\hat{G}_i; \mu_\theta, \sigma_\theta, \alpha_\theta) = \sum_{i=1}^{k} \sum_{j=1}^{N_f} \alpha_{\theta_j} \mathcal{N}(\hat{g}_i; \mu_{\theta_j}, \sigma_{\theta_j}), \tag{4}$$

while the variance of means is constrained to be close to the variance of the pseudo-label position so that the predicted means are dispersed. This can prevent the model from taking "shortcuts" and make the predictions focus on the mean position of different patterns. This part of the loss can be described as:

$$\begin{aligned} \mathcal{L}_v &= \text{MSE}(\sigma(\mu_\theta), \sigma(G_i)), \\ \mathcal{L}_c &= \text{MSE}(\mu_\theta, G_i) \end{aligned} \tag{5}$$

where $\sigma(\mu_\theta)$ is the variance of the predicted means, and $\sigma(P)$ is the variance of the pseudo-labels.

Furthermore, it is worth mentioning that the method we use to generate pseudo-labels is non-differentiable and generates independent goal distribution predictions. However, compared with the Independent goal distribution predictor we proposed, this vanilla method cannot predict in real-time due to the time-consuming steps such as matching and clustering, and

the non-differentiable generation requires us to train the networks in two stages separately, which will bring in cascade errors, resulting in performance degradation. The performance comparison is discussed experimentally in Section IV-D3.

E. Joint goal distribution cropper

Given the aircraft's estimated complete independent goal distribution, the Joint goal distribution cropper aims to remove redundant and unreachable goals. Drawing on a graph attention network as the basic structure, our proposed model makes two innovations: intention conflict measured Graph generation and graph attention with dynamic weight learning. Fig. 4 shows its basic structure.

1) Intention conflict Graph generation: We model the adjacency between all aircraft in airspace depending on whether their intentions are in conflict. The original graph attention methods construct graph relationships between agents through known observation information, but this is not suitable for high-speed flight trajectory prediction, because although the observation trajectories of the two aircraft are not strongly correlated, this does not mean that there will be no contact between the two aircraft in the short-term future after large-scale movements. Since the flight interaction cannot be effectively constructed only with observation information, we established an adjacency graph based on the intentions output by the Independent goal distribution predictor.

In practice, we model the adjacency relationship between aircraft as the Euclidean distance between independent goal predictions. If this distance is less than the threshold, there is interaction between these aircraft, denoted as 1, and vice versa. The 0-1 adjacency matrix acts as the mask in the attention mechanism to control the scope of interaction modeling. Mathematically, the generation of an adjacency matrix can be expressed as follows:

$$A_{ij} = \begin{cases} 1, & \text{if } \|G_i - G_j\|_2 < \epsilon \\ 0, & \text{if } \|G_i - G_j\|_2 \geq \epsilon \end{cases} \tag{6}$$

where $\|G_i - G_j\|_2$ is the distance between the independent goals of ith and jth aircraft, and A_{ij} is the adjacency relation between them with ϵ as threshold.

2) Graph attention with dynamic weight learning: We implicitly model the influences of aircraft interactions on each other through a graph attention network (GAT). Existing graph attention methods use fixed weights to calculate attention for different interaction scenarios, whose capability is insufficient to represent the rich interactions within the terminal airspace. To better cover the interaction patterns, we leverage the dynamic weight learning [72]–[74] to adapt to different scenarios without changing the network structure.

In practice, the dynamic interaction net takes the concatenated observed trajectories and independent goal distributions as input and maps them to the dynamic weight $\mathbf{W} \in \mathbb{R}^{d_u \times (d_h+3k)}$ and $\mathbf{a} \in \mathbb{R}^{2d_u}$ of graph attention with two linear projections $f_w(\cdot)$ and $f_a(\cdot)$, d_u is the dimension of the embedding in GAT. These weights encoding movements and interactions contain the motion features of different aircraft, which enable the GAT structure to adapt to different flights:

$$\begin{aligned} \mathbf{h}_m &= \text{cat}(\mathbf{X}, \Delta\mathbf{X}, \hat{\mathbf{G}}), \\ \mathbf{V_t} &= f_{Ie}(\mathbf{X}), \\ \mathbf{h_t} &= \text{cat}(\mathbf{V_t}, \hat{\mathbf{G}}), \\ \mathbf{W} &= f_w(\mathbf{h_m}), \\ \mathbf{a} &= f_a(\mathbf{h_m}) \end{aligned} \tag{7}$$

where $\mathbf{h}_m \in \mathbb{R}^{l \times (2T+k) \times 3}$ is the aircraft-specific meta info including the position, velocity and intention information at the observation period, $\mathbf{h}_t \in \mathbb{R}^{l \times (d_h+3k)}$ is the graph node feature includes the temporal encoding $\mathbf{V_t} = [{\mathbf{v_t}}^1, {\mathbf{v_t}}^2, \ldots, {\mathbf{v_t}}^l] \in \mathbb{R}^{l \times d_h}$ in Independent goal distribution predictor and goal estimations $\hat{\mathbf{G}}$.

Inside the graph attention, the interaction features after dynamic weight encoding are modeled by a Multi-Head Attention mechanism restricted by the adjacency matrix, and the assigned probability is obtained through a linear decoder for each node representing aircraft. Mathematically, this process of interaction modeling between node i and its neighbor $j \in \mathcal{N}_i$ defined by A_{ij} can be expressed as:

$$\begin{aligned} u_{ij} &= \mathbf{a}^\top \text{cat}(\mathbf{W}{\mathbf{h_t}}^i, \mathbf{W}{\mathbf{h_t}}^j), \\ \alpha_{ij} &= \frac{\exp(\text{LeakyReLU}(u_{ij}))}{\sum_{k \in N_i} \exp(\text{LeakyReLU}(u_{ik}))}, \\ {\mathbf{h_s}}^i(M) &= \prod_{k=1}^{M} \sigma(\sum_{j \in \mathcal{N}_i} \alpha_{ij}^k \mathbf{W}^k {\mathbf{h_t}}^j), \\ p_g &= f_p({\mathbf{h_s}}^i(M)) \end{aligned} \tag{8}$$

where α_{ij} is the attention coefficient between nodes, after softmax$([u_{i1}, u_{i2}, \ldots, u_{i\mathcal{N}_i}])$, ${\mathbf{h_s}}^i(M) \in \mathbb{R}^{d_u}$ is the aggregated interaction feature of ith node according to α_{ij} , M is the number of heads in multi-head attention, f_p is an MLP to assign probability. Ultimately, we prune out low-probability goal estimates, and the remaining ones will be used as the final goal predictions.

Based on the above designs, the Joint goal distribution cropper can capture interactions during high-speed movements for adapting to different patterns of flights and different interactions in shared airspace. It is optimized along with the entire network, serving for the performance of the overall trajectory prediction task.

To further explain the process of generating this "one-then-all" goal estimation intuitively, we visualized the independent intention distributions and joint intention distributions of single aircraft and multiple aircraft in Fig. 5. We can observe that the independent intention is a distribution that covers all patterns but is full of noise, while the cropped joint intention distribution is more accurate and in line with the scene interaction.

F. Trajectory generator

As shown in Fig. 2, our trajectory generator consists of a trajectory encoder and a noise estimation module. The trajectory generator is a diffusion model that generates actions $\hat{\mathbf{a}}_{\gamma+1} \in \mathbb{R}^{l \times T \times 3}$ conditioned on the goal predictions $(\mathbf{u}_{social}, \hat{\mathbf{G}})$.

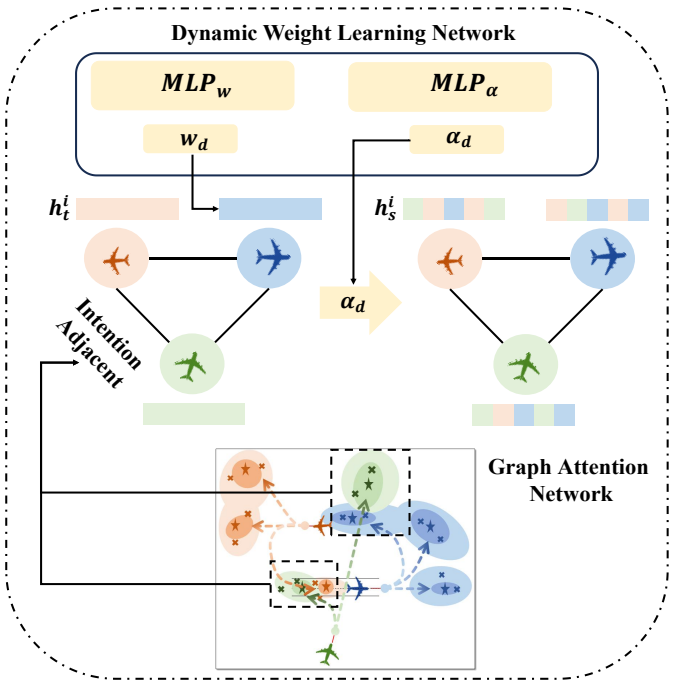

Fig. 4. The detailed structure of Graph attention with dynamic weight learning. The adjacency matrix is defined by the distance of estimated intentions, and the structure of the dynamic weight learning mechanism is shown in pink and light yellow.

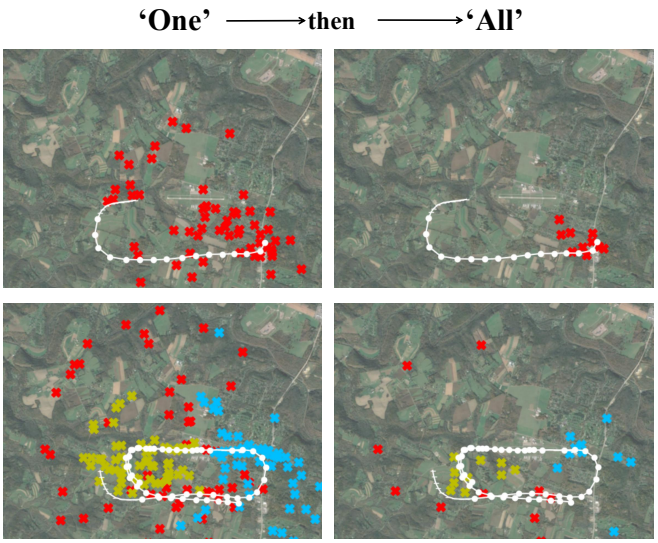

Fig. 5. The Visualization for Generating the Joint Goal Distribution from an Independent Goal Distribution by Using the "One-then-all" process.

The trajectory encoder $f_{Te}(\cdot)$ is composed of alternately superimposed temporal self-attention layers and social attention layers [72]. By rearranging the size of the observed trajectory tensor, it can be adapted to different attention operations, calculation within a pair of encoding layers is implemented as follows:

$$\begin{aligned}
\mathbf{X} &\xrightarrow{\text{reshape}} \mathbf{X} \in \mathbb{R}^{(B\times N)\times T\times D}, \\
\mathbf{u}_{\text{temporal}} &= \text{softmax}(\frac{f_{qt}(\mathbf{X})f_{kt}(\mathbf{X})^T}{\sqrt{d}})f_{vt}(\mathbf{X}), \\
\mathbf{u}_{\text{temporal}} &\xrightarrow{\text{reshape}} \mathbf{h} \in \mathbb{R}^{B\times N\times (T\times D)}, \\
\mathbf{u}_{\text{social}} &= \text{softmax}(\frac{f_{qs}(\mathbf{h})f_{ks}(\mathbf{h})^T}{\sqrt{d}})f_{vs}(\mathbf{h})
\end{aligned} \tag{9}$$

where $\mathbf{u}_{\text{temporal}} \in \mathbb{R}^{(B\cdot l)\times T\times d_h}$ is temporal embeddings and $\mathbf{u}_{\text{social}} \in \mathbb{R}^{B\times l\times (T\cdot d_h)}$ is temporal-social feature.

The reason why the two encoders are designed independently for the goal estimation stage and the trajectory generation stage is that the trajectory generation is an accuracy-oriented task. However, goal estimation is a diversity-oriented task. The two tasks are contradictory to each other, and sharing one backbone network as a multi-task learning framework will cause performance degradation. Experimental verification is in Section IV-D4.

The noise estimation module $f_\epsilon(\cdot)$ is a transformer-based network designed to estimate the noise to reduce. In addition to the basic denoising process, we added an intention adjustment mechanism based on classifier-free guidance, as shown in Fig. 6. During the training process, the estimation of noise of the γ-th step is defined as follows:

$$\begin{aligned}
\tau &= \begin{cases} \hat{G}_{min} & \text{if } p > 0.8 \\ \emptyset & \text{else} \end{cases}, \\
\epsilon^\gamma_{\theta,c} &= f_\epsilon(\hat{\mathbf{a}}_{\gamma+1}, \mathbf{u}_{social}, \tau)
\end{aligned} \tag{10}$$

where $\hat{G}_{min}$, is the goal estimation closest to the ground truth. And τ is the condition in the training process. According to the estimation of noise $\epsilon^\gamma_{\theta,c} \in \mathbb{R}^{l\times T\times 3}$, the loss function of the trajectory generation stage can be formed as:

$$\mathcal{L}_d = \mathbb{E}_{\epsilon\sim\mathcal{N}(\mathbf{0},\mathbf{I})}\|\epsilon^\gamma_\theta - \epsilon\|_2^2 \tag{11}$$

During the inference phase, for each predicted goal, we generate the corresponding action sequence following the backward process of the diffusion model with the intention adjustment mechanism. The estimation of action series $\hat{\mathbf{a}}_\gamma$ at γth denoising step is defined as follows:

$$\begin{aligned}
\epsilon^\gamma_{\theta,c} &= f_\epsilon(\hat{\mathbf{a}}_{\gamma+1}, \mathbf{u}_{social}, \hat{\mathbf{G}}), \\
\epsilon^\gamma_{\theta,f} &= f_\epsilon(\hat{\mathbf{a}}_{\gamma+1}, \mathbf{u}_{social}), \\
\epsilon^\gamma_\theta &= (\omega+1)\epsilon^\gamma_{\theta,c} - \omega\cdot\epsilon^\gamma_{\theta,f}, \\
\hat{\mathbf{a}}_\gamma &= \frac{1}{\sqrt{\alpha_\gamma}}(\hat{\mathbf{a}}_{\gamma+1} - \frac{1-\alpha_\gamma}{\sqrt{1-\bar{\alpha}_\gamma}}\epsilon^\gamma_\theta) + \sqrt{1-\alpha_\gamma}\cdot\mathbf{z}
\end{aligned} \tag{12}$$

where ω is the intention adjustment scale, positively related to the influence of intention in trajectory generation. α_γ and $\bar{\alpha}_\gamma = \prod_{i=1}^{\gamma}\alpha_i$ are parameters in the diffusion process. $\mathbf{z} \sim \mathcal{N}(\mathbf{0},\mathbf{I})$ are gaussian noises. The generated action sequences $\hat{\mathbf{a}}_0$ are integrated from the current position $\mathbf{X}^0$ to get the final predicted trajectory $\hat{\mathbf{Y}}$:

$$\hat{\mathbf{Y}}^t = \sum_{i=0}^{t-1}\hat{\mathbf{a}}_0^i + \mathbf{X}^0, t \in [1,\ldots,T_f] \tag{13}$$

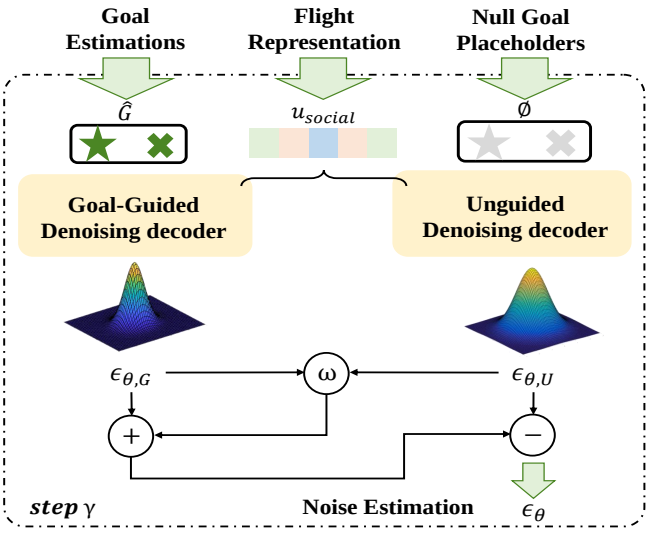

Fig. 6. The structure of the Intention adjustment mechanism, a CFG strategy is employed to adjust the strength of intention guidance.

In the above design, our motivation behind the intention adjustment lies in a different way of thinking about the intentions lying behind goal coordinates. Traditional anchor-based methods often use the ground truth or closest endpoint prediction to train the trajectory generation module alone, which undermines its modeling of movement patterns and degrades it only for linking the observation to each destination, making it vulnerable to unreachable destination predictions and also leading to difficulty in internalizing the operational knowledge of flight fully. We believe that goal prediction should only indicate flight intention, rather than the specific endpoint. Therefore, we used a simple classifier-free guidance mechanism for the goal prediction conditions, which allows a continuous adjustment of the goal prediction conditions.

IV. Experiments

A. Datasets

We evaluate our method on TrajAir [14], a large-scale real-world flight trajectory dataset collected within the terminal of the Pittsburgh-Butler Regional Airport (KBTP). KBTP is a non-towered general airport at the county seat of Butler County, Pennsylvania, United States. The dataset includes 111 days of flight data from 01:00 AM local time to 11:00 PM local time. TrajAir is notable for both trajectory complexity and pattern imbalance compared with datasets in en-route airspace and autonomous driving. The dataset is published with the flight navigation task on the ICRA 2022. The dataset is open source at https://theairlab.org/trajair.

B. Implementation Details and Experimental Design

In our experiment, we used the processed complete dataset 111days for the main training and testing. Meanwhile, we also conducted verification on four processed subset 7days1-4. We set the prediction horizon to $T = 40$ and $T_f = 120$, using the last 40 seconds of observed trajectories to predict the next 120 seconds of flight. To ensure the integrity and uniformity of the trajectories within the scene, we filtered out the incomplete trajectories with a length < 160. Different sampling intervals are used to meet the input and output form of baseline methods. The number of aircraft in the scene is set from 1 to 7. In our method, all scenes are padded to 7 aircraft. Since the take-off and landing behaviors of aircraft are related to absolute positions rather than relative positions, we continued to use the normalized representation relative to the absolute coordinates of the airport instead of the normalized representation relative to the starting position, although this may hinders our model to generalize in the airports with other layouts.

We train our models with a batch size of 256 for 500 epochs, using Adam optimizer with a learning rate of 1×10^{-5}. For key hyperparameters, we set the number of components of the mixed gaussian distribution to 10 and the number of samples to 30, then through the joint goal distribution clipper, we retain the top 20 goals with high confidence. Finally, we generate a trajectory for each goal prediction, composing 20 candidate trajectories for an observed flight. The entire framework is trained on one GTX-4090 GPU, implemented with PyTorch 2.1.0.

Since the existing methods for flight trajectory prediction are very limited, we migrate some pedestrian and vehicle trajectory prediction methods such as STG-CNN [75] and MID [67] as baseline methods for comparison.

C. Evaluation metrics

In our experiment, the Average Displacement Error (ADE) and the Final Displacement Error (FDE), which are commonly used in trajectory prediction, are used as our evaluation metrics. In addition, we propose a Global-local endpoint variance (GLeV) indicator for the evaluation of the predicted diversity on the premise of social acceptance.

ADE and FDE: The Average displacement error (ADE) is defined as the mean L_2 error between all predicted trajectory coordinates and ground truth, and the Final displacement error (FDE) is defined as the L_2 error of the last time step. In stochastic trajectory prediction, only the candidate trajectory closest to the ground truth is used in calculations:

$$\begin{aligned} ADE &= \frac{1}{l \cdot T_f} \sum_{i=0}^{l} \min_{j \in k} \|\hat{Y}_i^j - Y_i\|_2 \\ FDE &= \frac{1}{l} \sum_{i=0}^{l} \min_{j \in k} \|\hat{Y}_i^{j,T_f} - Y_i^{T_f}\|_2 \end{aligned} \tag{14}$$

where $\hat{Y}_i^j$ is the jth trajectory prediction of ith aircraft, $\hat{Y}_i^{j,T_f}$ is its last trajectory point, l is the number of aircraft in the current scene, k is the number of the candidate predictions.

As the most commonly used evaluation metrics in trajectory prediction, the ADE and FDE are evaluated using the candidate trajectory closest to the ground truth, which can reflect the upper bound of a model's performance. However, they are insufficient to reflect the diversity of candidate trajectories because of information leaks and sensitivity to randomisation. For example, the constant velocity model (CVM) with a wider

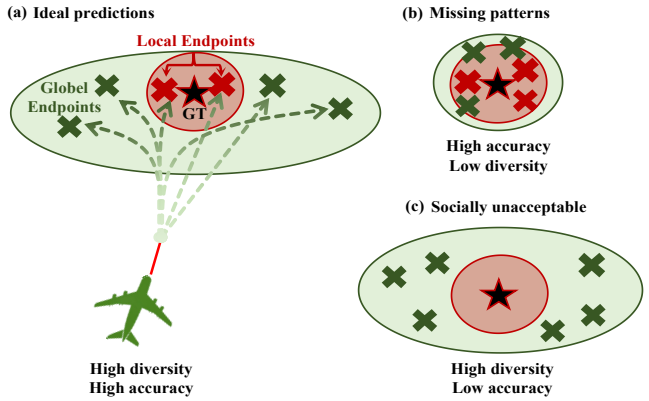

Fig. 7. The diagram illustrates the three types of trajectory prediction results measured by the GLeV evaluation metric: a) shows the ideal trajectory prediction result, where the prediction accurately fits the given ground truth while covering other potential diverse motion patterns. b) shows the trajectory prediction result of missing patterns, where the prediction is overly concentrated around the given ground truth, ignoring other possible patterns. c) shows the socially unacceptable trajectory prediction result, where the prediction overemphasizes diversity and fails to fit the given ground truth, generating unreal prediction outcomes.

angle spread [76] can even outperform deep learning-based models.

Global-local endpoint Variance (GLeV): To fill the blank in diversity evaluation, we propose a simple yet effective metric named Global-local endpoint Variance (GLeV), which calculates the expectation, across aircraft in the current scene, of the ratio of each aircraft between the endpoint variance of its candidate trajectories and the variance of its top$n$ endpoints ($n < k$) closest to ground truth:

$$GLeV = \frac{1}{l}\sum_{i=0}^{l}\frac{\sigma(\hat{\mathbb{Y}}_i^{T_f})}{\sigma(\hat{Y}_i^{T_f})} \tag{15}$$

where l is the number of aircraft in the current scene, T_f is the last time step of the future trajectory, $\sigma(\hat{Y}_i^{T_f})$ is the variance of all candidate endpoint predictions for ith aircraft denoted as $\hat{Y}_i^{T_f} \in \mathbb{R}^{l\times k\times 3}$, $\sigma(\hat{\mathbb{Y}}_i^{T_f})$ is the variance of the closest top$n$ endpoints measured by L_2 distance to the ground truth, which are denoted as $\hat{\mathbb{Y}}_i^{T_f} \in \mathbb{R}^{l\times n\times 3}$ ($\hat{\mathbb{Y}}_i^{T_f} \in \hat{Y}_i^{T_f}$).

As shown in Fig. 7, GLeV needs to comprehensively reflect both prediction diversity and accuracy, with acceptable prediction accuracy being a prerequisite for generating as diverse trajectory predictions as possible. We penalize predictions that are highly accurate but have low diversity, as well as those that have high diversity but low accuracy. Mathematically, this is reflected in the ratio form of the GLeV formula 15. For each aircraft, prediction diversity is reflected as the global variance of the predicted endpoint coordinates, while accuracy is reflected as the variance of the predicted endpoint coordinates that fall around the ground truth, when local variance decreases and global variance increases, the GLeV value increases to represent this more diverse and plausible trajectory prediction. The effectiveness of this evaluation metric will be demonstrated in Section IV-E2.

TABLE II
QUANTITATIVE PREDICTION RESULTS FOR THE TRAJAIR DATASET

Methods	ADE(km)↓	FDE(km)↓
Const. Vel [77].	1.85	4.16
Nearest Neigh [14].	1.97	2.25
STG-CNN [75]	1.37	2.91
TransformerTF [78]	1.67	3.94
A-VRNN [79]	0.62	1.49
DAG-Net [79]	0.77	1.61
S-PEC [80]	1.04	2.18
Expert-Traj [40]	0.55	<u>0.72</u>
PECNet [36]	0.67	1.14
TrajAirNet [14]	0.79	1.58
Social-PatteRNN [16]	0.69	1.44
Social-PatteRNN-ATT [16]	0.67	1.40
MID [67]	<u>0.55</u>	0.87
GooDFlight(Ours)	**0.29**(48.21%↓)	**0.39**(45.83%↓)

D. Quantitative Results and Analysis

In this section, we quantitatively compared the performance of GooDFlight with baselines (IV-D1) and analyzed the result in prediction diversity (IV-E2), end-to-end training (IV-D3), separate encoder structure (IV-D4), and intention adjustment mechanism (IV-E3) to support our technical improvements.

1) Quantitative Results: In Table II, we compare with other baseline trajectory prediction methods in non-tower airport terminal airspace. The sufficient coverage of flight pattern diversity allows the GooDFlight to significantly outperform the baseline models on both ADE and FDE. Compared with the state-of-the-art methods on this task, our GooDFlight method improved by +48.21% on ADE and +45.83% on FDE.

Specifically, due to the application of generative methods and the consideration of social relationships, existing social-aware trajectory prediction methods significantly exceed the single aircraft prediction frameworks such as STG-CNN and TransformerTF. However, limited by inadequate coverage of flight patterns, these methods are overfitting on hovering. The recent diffusion-based model MID outperforms CVAE and VRNN based models with its superior expressiveness, but it still has difficulties in fully covering the flight patterns and results in suboptimal forecasting. Born to tackle this major obstacle, GooDFlight takes a great leap in prediction diversity, reducing the ADE from about 0.6km to less than 0.3km, and the FDE from about 0.7km to less than 0.4km. It is worth noting that in the goal-based method, Expert-Traj, which also uses historical information as ours, has excellent performance with a small difference between ADE and FDE on this task, while PECNet based solely on learning struggles to achieve comparable performance.

We also carried out experiments on four subsets: 7days1-4 in Table III. The experimental results show high consistency across different methods and datasets, our method has stably achieved the optimal performance on each subset.

TABLE III
THE ADE/FDE (KM) COMPARISON ACROSS THE 7DAYS1-4 SUBSETS

Methods	7days1	7days2	7days3	7days4
TrajAirNet [14]	0.72/1.45	0.80/1.59	0.88/1.67	0.70/1.44
Social-PatteRNN-ATT [16]	0.61/1.42	0.76/1.67	0.75/1.65	0.67/1.51
GooDFlight(Ours)	**0.27/0.41**	**0.32/0.40**	**0.36/0.48**	**0.30/0.40**

TABLE IV
QUANTITATIVE COMPARISON BETWEEN OUR DIVERSITY FOCUSING MODELS AND EXISTING ADVANCED METHODS

Methods	ADE(km)↓	FDE(km)↓	GLeV↓
Minimalist Network	0.75	1.61	0.57
TrajAirNet	0.79	1.58	0.42
Social-PatteRNN-ATT	0.67	1.40	0.49
Social-PatteRNN-ATT(k=15)	0.69	1.43	0.50
Social-PatteRNN-ATT(k=10)	0.74	1.52	0.61
MN with goal estimation	0.56	0.71	0.17
GooDFlight(Ours)	**0.29**	**0.39**	**0.08**

2) Quantitative Analysis of Prediction Diversity and GLeV: This section quantitatively analyzes the importance of diversity in GA trajectory prediction task. We first applied a minimalist network consisting of a 6-layer LSTM, a 2-layer transformer, and a MLP trajectory decoder on TrajAir. The ADE/FDE of this minimalist model is 0.75km/1.61km, which is comparable to TrajAirNet. A minimal model is comparable to a carefully designed model, indicating that the key to this task is not interactive modeling and generative prediction. We further add the goal estimation stage to the minimalist network, resulting in 0.56km/0.71km on ADE/FDE. This +50% performance improvement demonstrates the importance of capturing the diversity of flight patterns in trajectory prediction tasks.

As shown in Table IV. In the first group, GLeV remains high for all methods. In the second group, there is a rapid drop in GLeV, decreasing by 80%. Theoretically, for the methods that ignore the trajectory diversity, despite the network complexity and the number of candidate trajectories, the prediction diversity will remain low due to the high consistency in candidate trajectories. For methods enhancing diversity, the prediction diversity should be fundamentally improved. GLeV, as an indicator of diversity, accurately reflects the phenomenon. At the same time, GLeV also cooperates well with ADE and FDE. In the experiment, when ADE and FDE increase, GLeV decreases, diversity and accuracy are positively correlated, which is in line with common sense.

3) Quantitative Analysis of End-to-end Training: This section quantitatively analyzes the benefits of end-to-end training. Anchor-based trajectory prediction is often divided into two separate stages, introducing inevitable cascading errors. This problem is tricky for Instance-Based methods based on retrieving similar previous samples since their retrieval process is non-differentiable. To analyze how this cascade error affects the prediction performance, in Table V, we compare the performance between original GooDFlight, GooDFlight with separately trained stages, GooDFlight with a non-differentiable retrieval process and other existing methods.

TABLE V
QUANTITATIVE COMPARISON BETWEEN STAGE CONNECTING PARADIGMS AND TRAINING FRAMEWORKS

Methods	ADE(km)↓	FDE(km)↓	GLeV↓	Time(s)↓
TrajAirNet	0.79	1.58	0.424	**0.25**
Social-PatteRNN-ATT	0.67	1.40	0.493	2.45
Pseudo-labels replacement	0.32	0.45	0.121	>10
Two-step training	0.32	0.45	0.095	0.57
End-to-End (GooDFlight)	**0.29**	**0.39**	**0.083**	0.57

From Table V, the end-to-end trained GooDFlight achieves the most accurate and diverse performance, trump the two-stage training and even the direct use of pseudo-label. The Instance-based method obtains the suboptimal result with unbearable time-consuming. TrajAirNet has the best timeliness with the worst accuracy and diversity. The prediction diversity of GooDFlight surpasses directly using pseudo-labels because it considers goal estimates as pattern indicators rather than destinations and filters out unreachable goal predictions, making the estimated goals more in line with the overall accuracy demand, which is also the reason why its accuracy is better than directly using pseudo-labels.

Furthermore, it is worth noting that theoretically, clustered empirical goals, as the pseudo-label in the goal estimation stage, should be the performance upper bound. However, the end-to-end training method enables the two-stage model to jointly optimize towards the final goal of trajectory prediction. This comprehensive optimization process allows the goal estimation stage to exceed the theoretical performance limit.

4) Quantitative Analysis of Separate Encoders: This section quantitatively analyzes the structural relationship between the goal estimation stage and the trajectory generation stage in end-to-end training. In Table VI, we quantitatively compare the performance between the model with a shared backbone, separated encoders, and the two-stage training mentioned before. The separated encoders achieve the best performance and the shared backbone obtains the suboptimal prediction where the decline in diversity is particularly severe.

We further analyze this phenomenon from the nature of this two-stage task: The goal estimation stage needs to cover all patterns and emphasizes diversity, while the trajectory generation stage constrains the authenticity of the trajectory and emphasizes accuracy, this fundamental disagreement makes them drag each other down in the shared backbone, resulting in suboptimal performance. Based on this counterintuitive insight, we design two separate encoders for the goal estimation stage and the trajectory generation stage. In practice, the encoder for the goal estimation stage is very lightweight, having little impact on the overall parameter amount of the model.

5) Quantitative Analysis of Intention adjustment mechanism: This section quantitatively analyzes the effectiveness of the Intention adjustment mechanism we proposed. Based on our standpoint that the goal estimations should be the

TABLE VI
QUANTITATIVE ANALYSIS ON DIFFERENT ENCODER DESIGNS

Methods	ADE(km)↓	FDE(km)↓	GLeV↓
Two-step training	0.32	0.45	<u>0.095</u>
Shared Backbone	<u>0.30</u>	<u>0.42</u>	0.106
Separate Encoders	**0.29**	**0.39**	**0.083**

TABLE VII
QUANTITATIVE ANALYSIS OF INTENTION ADJUSTMENT WEIGHTS

Intention adjustment weight	ADE(km)↓	FDE(km)↓	GLeV↓
$w = 0$	0.53	0.71	0.253
$w = 0.3$	0.42	0.67	0.209
$w = 0.6$	0.36	0.56	0.142
$w = 1$	0.34	0.50	0.116
$w = 2$	<u>0.30</u>	**0.39**	0.088
$w = 3$	**0.29**	**0.39**	**0.083**
$w = 4$	<u>0.30</u>	<u>0.40</u>	<u>0.088</u>
$w = 5$	0.31	0.42	0.090
$w = 6$	0.32	0.43	0.092
$w = 7$	0.32	0.44	0.095

indicators of flight patterns rather than the destinations in physical space, we compared the performance under different continuous intention strengths whose ω = 1-7 in Table VII. It shows that between no intention guidance and full intention guidance, when the intention adjustment weight $\omega = 3$, the performance reaches the optimum and diminishes to both sides. This phenomenon not only reveals that intention has a decisive role in trajectory generation but also reveals that the intention estimations cannot be completely used as the endpoint position coordinate.

From a technical perspective, during this growing and cascading variation, our Intention adjustment mechanism demonstrates its unique ability to continuously and implicitly adjust between anchor-free and anchor-based paradigms. When ω = 0, the model is equivalent to the anchor-free paradigm, and as ω increases, the model constantly converges to the anchor-based paradigm. Our work achieves this flexible and superior performance for the first time, employing only the most commonly used classifier-free guidance strategy, which hopefully can be an inspiration for subsequent research.

E. Visualization and Qualitative Analysis

To support the quantitative comparison and analysis, in this section, we visualize the trajectory prediction and qualitatively compare the proposed GooDFlight framework with existing baselines. Then we further analyze our contributions in terms of prediction diversity (IV-E2), intention adjustment mechanism (IV-E3), and prediction stability (IV-E4).

1) Visualization Results: In this section, we qualitatively analyze the proposed GooDFlight framework and existing baselines dedicated to aircraft trajectory prediction under the same conditions, by visualizing the flight trajectory prediction results in real-time flight scenarios in non-towered terminal airspace, including entering and leaving terminal airspace, flying on different legs, switching legs, and some unconventional flights that violate the FAA guidance. These scenarios can be considered to fully represent the high maneuverability and complex interactions of flights in the complex terminal airspace of non-towered airports, and accurate prediction of these typical and even difficult flight samples also validates the effectiveness and robustness of the proposed method in real-world applications.

As shown in Fig. 8, taking the Pittsburgh-Butler Regional Airport remote sensing map as the background, all trajectories are depicted in aerial view, and the wind direction is from left to right. The observed trajectory is represented by white lines with short horizontal marks. the ground truth of the future trajectory is represented by white dotted lines, and the best result in 20 predicted candidates of different aircraft is represented by thick solid lines in different colors. It should be noted that the current method does not have a unified observation-prediction horizon, we have to use separate drawings for each method in visualization.

It is clear that the GooDFlight achieves the best prediction performance based on the visualization results. As shown in the bottom row of Fig. 8, our model accurately predicts the trajectory of one or multiple aircraft landing, taking off, hovering, or even violating regulations in the shared airspace simultaneously. At the same time, the TrajAirNet and the Social-PateRNN-ATT as the most advanced methods on this task, seem to hardly handle the complex flight scenarios other than most common hovering flight maneuvers. In the top row of Fig. 8, the results of TrajAirNet show its flaws in inaccuracy and instability, which are especially significant when aircraft perform large-scale maneuvers such as 180-degree U-turns. At the same time, for uncommon flight patterns, such as the red trajectory in (c) and the blue trajectory in (d) indicating an illegal right turn departure, the model first predicts it as a common left turn departure, the right turn pattern can be predicted only when there is an observed tendency, but even so, the generated trajectory is unrealistic. Furthermore, for similar common patterns, such as the yellow trajectory in (f) indicating a 45-degree departure, the model can not recognize it from the hovering flight pattern. In the second row of Fig. 8, Social-PateRNN-ATT utilized the short-term intentions, making its accuracy and stability far exceed TrajAirNet in common flight patterns such as hovering and 180-degree U-turn, but it is still severely limited by insufficient flight pattern coverage, the left turn landing in (a) and the runway departure maneuver in (b) were incorrectly predicted as a straight departure and a left turn, respectively.

2) Qualitative Analysis of Prediction Diversity and GLeV: The core idea of GooDFlight proposed in this article is to focus on the diversity of predictions to cover all possible flight patterns. In this section, we elaborate to visually discuss the diversity of predictions and our corresponding evaluation metric GLeV in Fig. 9. We sequentially visualized all 20 candidates of the TrajAirNet, Social-PateRNN-ATT, and GooDFlight, where we also visualized its goal estimations. In addition, we also show the previous similar flights and corresponding pseudo-

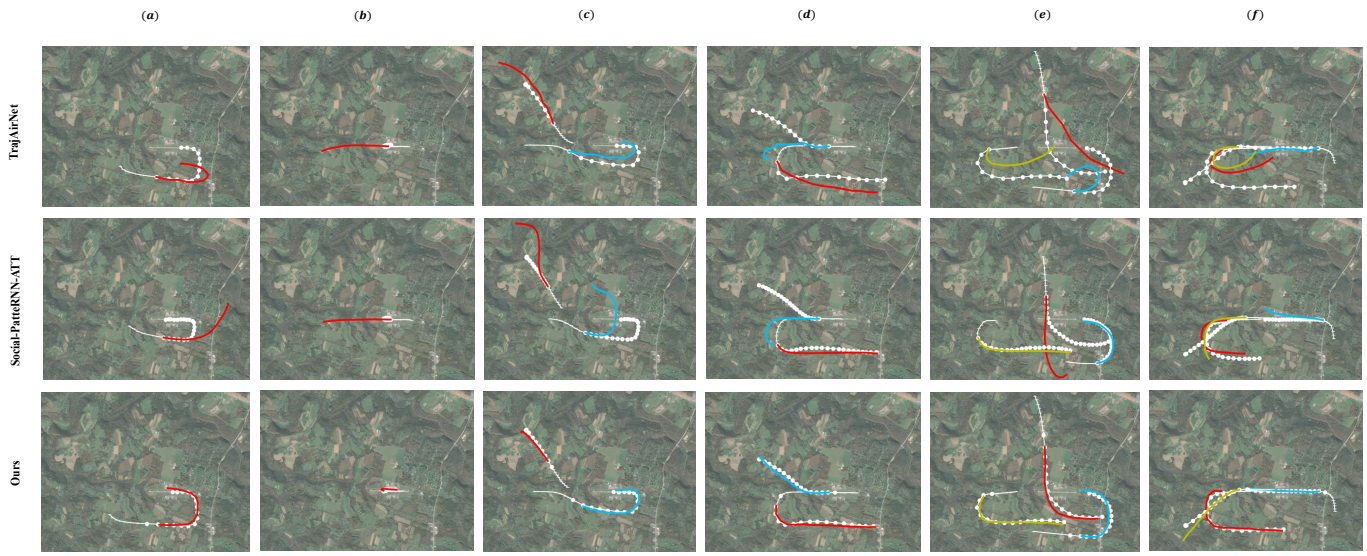

Fig. 8. The overall visualized comparison between the best in 20 predictions of GooDFlight, TrajAir, and Social-PateRNN-ATT under different flight patterns and different numbers of aircraft. The observed trajectory is represented by white lines with short horizontal marks. the ground truth of the future trajectory is represented by white dotted lines, and the best result in 20 predicted candidates of different aircraft are represented by thick solid lines in different colors

labels on the right.

As shown in the first and second columns of the figure, The predicted trajectories of TrajAirNet and Social-PateRNN-ATT both show a highly centralized distribution, regardless of the number of candidate trajectories. This distribution characteristic results in extremely limited coverage of flight patterns. At the same time, because the imbalance between flight modes is not considered, this concentrated distribution often falls in the most common hovering maneuver flight, which can explain why there is a high tendency to predict hovering flight and they only have accurate prediction results for hovering scenarios.

In contrast, the GooDFlight can fully cover all possible flight patterns in the future even if they are very different from each other. As shown in the third column, when faced with an observed trajectory hovering within the downwind leg, in addition to continuing to hover to the base leg, the possible right turn departures were also considered in different trajectories. Analysis from the perspective of the macroscopic distribution of predicted trajectories, the goal estimation stage comprehensively considers the empirical goal distribution of a single aircraft and the interactive relationship of all aircraft in the scenario, which covers all plausible future flight patterns. Analyzing from a micro perspective, the trajectory generation stage uses a powerful diffusion model with Intention adjustment mechanism to set different confidence levels for goal conditions, prompting diversity when generating trajectories under different flight operations for similar goal estimations.

From the perspective of the GLeV metric, the visualization results show that when the distribution of the endpoints of the predicted trajectories close to the ground truth endpoint are concentrated, and the endpoints of all predicted trajectories are dispersed, the GLeV indicator is lower, indicating that the diversity of model predictions is enhanced. In the comparison of the first, second, and last columns, we can find out that the prediction diversity of GooDFlight is significantly higher than other methods, and its GLeV shows a lower order of magnitude, which preliminarily verifies the GLeV effectiveness. In addition, we can see in the second row that even though the global diversity of TrajAirNet is slightly larger than Social-PateRNN-ATT, the latter's accurate endpoint predictions are more closely distributed, which leads to a lower GLeV. Besides, even though the accurate endpoint predictions of GooDFlight are sparse, its rich global diversity makes up for it.

Through the above visualization, we not only verified the effectiveness of GLeV in evaluating prediction diversity but also explained how the predicted endpoint distribution near the ground truth and the global predicted endpoint distribution will affect GLeV.

At the same time, this also shows a series of intermediate results generated by explicit modeling in GooDFlight framework. By querying these results in reverse, we can easily trace the predicted trajectory to previous similar samples which is in line with the physiological law of humans predicting the development of future events by recalling the past, thus greatly improving the interpretability of the model.

3) Qualitative Analysis of Intention adjustment mechanism: In addition to quantitative analysis, visual qualitative analysis is more helpful for understanding the Intention adjustment mechanism, in this section, we visualized and compared the trajectory generation results under different intention adjustment weights in Fig. 10. From left to right $\omega = 0, 1, 3, 5$, showing the transformation from anchor-free to anchor-based trajectory prediction.

Specifically, in yellow dotted boxes, when $\omega = 0$, GooDFlight is equivalent to anchor-free trajectory prediction. The trajectory prediction is only determined by the observation trajectory and is not affected by the goal predictions. When $\omega = 1, 3$, GooDFlight is gradually enhanced by goal predic-

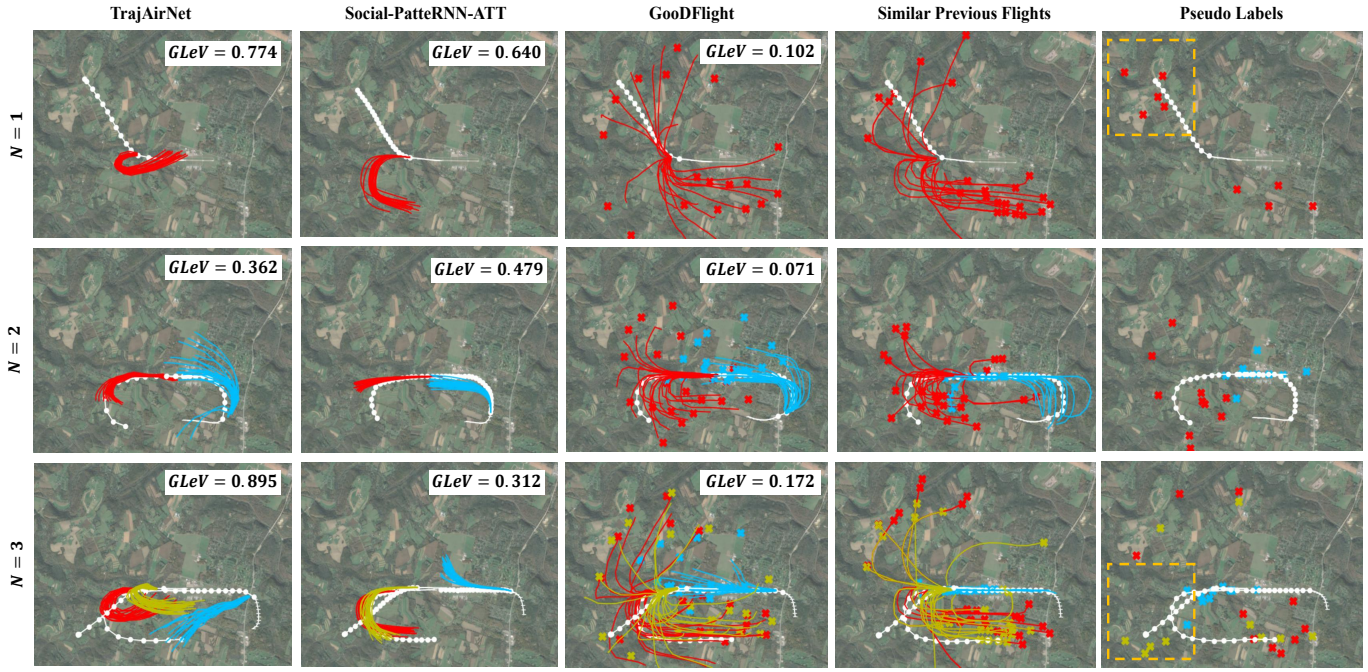

Fig. 9. The visualized comparison between all the predictions of GooDFlight, TrajAir, and Social-PateRNN-ATT under different flight patterns and different numbers of aircraft. The goal predictions are represented by solid crosses in different colors, the GLeV for each scene is recorded in the white block, and the enhanced goals in pseudo-labels are outlined in yellow dotted boxes.

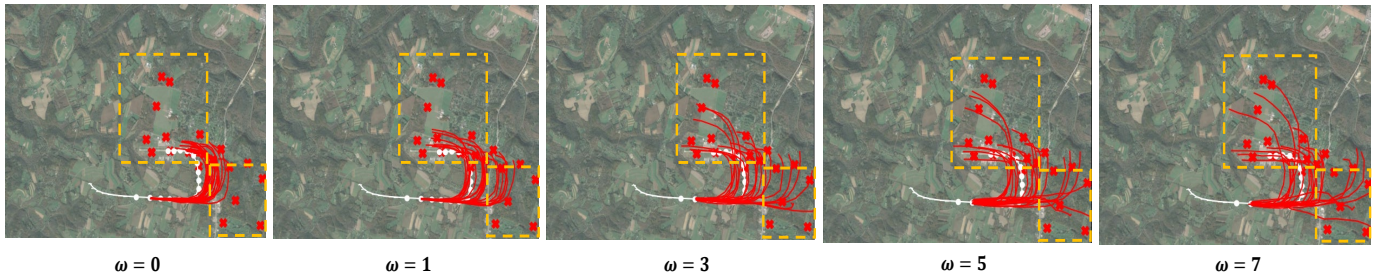

Fig. 10. The visualized trajectory generation results of GooDFlight under different intention adjustment weights from 0 to 7. The major differences are outlined in yellow dotted boxes.

tions, the endpoints of trajectory prediction begin to approach goal predictions. When $\omega = 5, 7$, GooDFlight converges to anchor-based prediction, the trajectories are closer to completely connecting the endpoints.

The value of ω has the final impact on prediction, when ω is small, the diversity of anchor-free predictions is too low to cover different patterns, but when ω is too large, just like the anchor-based method, the generated trajectory will produce excessively biased or unrealistic predictions due to forcibly approaching the goal conditions. This is especially obvious in GooDFlight which emphasizes the diversity of goals. Fortunately, due to the advantage that our ω can be continuously adjusted, we can always find an optimal condition strength to achieve the best prediction performance.

4) Qualitative Analysis of Prediction Stability: In addition to the diversity and accuracy of prediction, we cannot ignore that trajectory prediction is actually a time-continuous generation in practical applications, which has high requirements on the stability of the generation, especially under conditions that emphasize diversity. In Fig. 11, we visualize open-loop best trajectory predictions within an 8-minute hovering flight over all legs and compare the stability of different methods.

Even in the face of the most common hovering flight, existing methods still cannot output stable and accurate predictions. In the first and second row of Fig. 11, the results of TrajAirNet and Social-PateRNN-ATT show huge fluctuations in the predicted trajectory, making it almost impossible to achieve stable prediction. In comparison, in the third row, our proposed GooDFlight method perfectly achieves stable and accurate flight trajectory prediction. This is because our method is able to cover all flight modes based on previous similar flights, which are often similar in consecutive observation trajectories.

F. Ablation Study

In this section, in order to verify the effectiveness of our contributions to the proposed GooDFlight framework, we designed an ablation study to gradually add the different components we proposed to the plain diffusion model. The

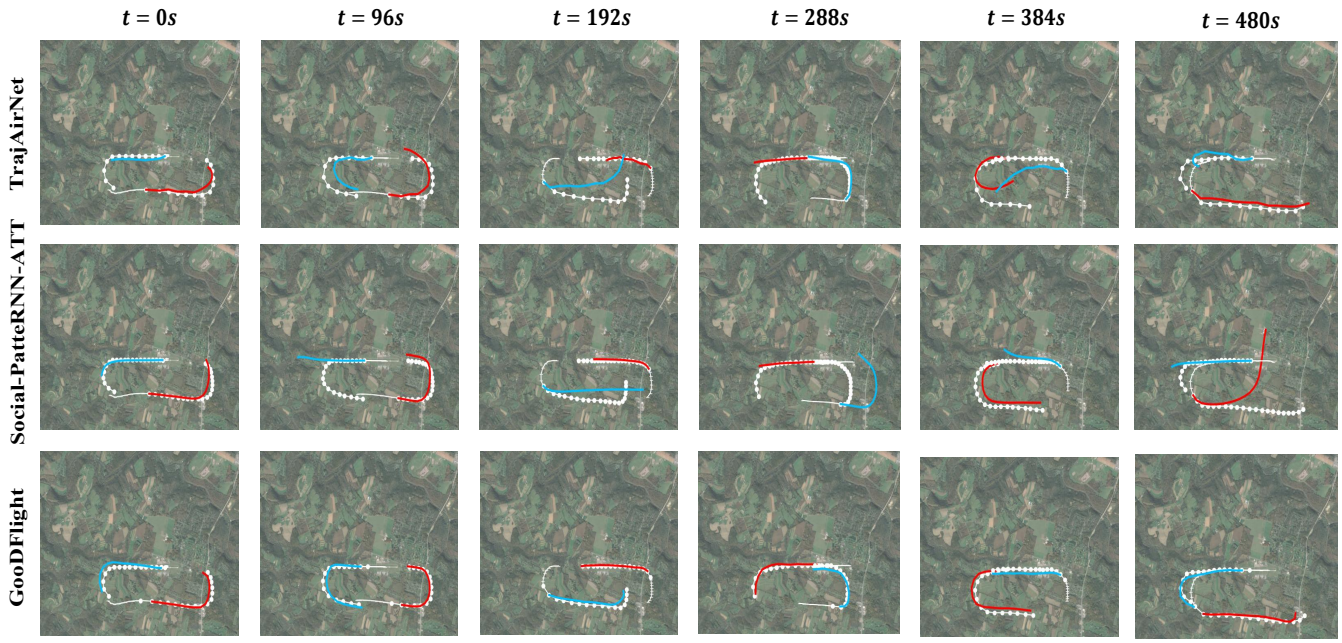

Fig. 11. The visualized best in 20 predictions of GooDFlight, TrajAir, and Social-PateRNN-ATT within an 8-minute hovering flight.

Independent goal distribution predictor, the Joint goal distribution Predictor, the DWGAT, and the Intention adjustment mechanism are added on step by step until the complete GooDFlight framework is composed, and their performance improvements are recorded from the top to the bottom in Table IX.

It should be noted that in the above ablation studies, to ensure a consistent number of predicted goals, we sampled only 20 goal estimates when using the Independent goal distribution predictor alone. At the same time, we consider the goal estimates as the positional coordinates of the endpoints until the Intention adjustment mechanism is added.

From the overall performance changes from top to bottom in Table IX, we can see that our proposed One-then-all goal estimation makes a fundamental leap in the performance of trajectory prediction, while our DWGAT network and Intention adjustment mechanism also lead to a great improvement in the overall performance. In terms of accuracy, every module added to the network brings a performance gain. In terms of accuracy, each addition to the network has resulted in a performance gain, and our manipulation of goal estimation has allowed the network to predict future flight patterns and destinations more accurately, which has resulted in significant increases in ADE, especially in FDE metrics. Meanwhile, in terms of diversity, our framework achieves the improvement of GLeV from two perspectives, with the introduction of the Independent goal distribution predictor, the global endpoint diversity is much higher than that of the Base Model, and even slightly higher than the model with Joint goal distribution Predictor. This undoubtedly results in a much lower GLeV metric than the model that does not take into account predictive diversity. However, contrary to the effect of global diversity, its GLeV metric is higher than that of the model with the Joint goal distribution Predictor.

This effect is a result of higher prediction accuracy. In all previous discussions we have emphasized the GLeV metric's representation of predictive diversity by highlighting the impact of diversity, but as a metric defined by division, its numerator has an equally important impact on the results as a representation of accuracy. It is the subsequent inclusion of modules to take into account more detailed factors such as aircraft interactions that leads to a significant improvement in the accuracy of the model and introduces a significant reduction in the numerator of the GLeV while a small change in the denominator, this explains the overall numerical reduction in the GLeV.

Faced with the decision-making pressure of aircraft in a high-dynamic environment, we conducted ablation experiments on the number of predicted candidate trajectories in addition to the designed modules in Table VIII. We compared the performance of the GooDFlight and TrajAirNet when the number of candidate trajectories k is equal to 15, 10, 5, and 1. This will contribute to the application of our model in the actual environment.

The comparison results show that as the number of candidate trajectories decreases, both the ADE and FDE of the two models decline significantly. The change of the TrajAirNet method on GLeV is slight (0.44-0.62), which indicates that the reduced trajectories do not affect the prediction distribution of TrajAir, or in other words, the predictions are redundant. We can also see overly consistent predicted trajectories in the visualization. GoodFlight shows a relatively large increase on GLeV (0.11-0.47), which means that the reduced candidate trajectories have affected the overall trajectory output distribution, that is to say, most of the candidate trajectories predicted are meaningful. Finally, even in the extreme case where $k = 1$,

Fig. 12. The visualized predictions of GooDFlight when $k = 1$.

TABLE VIII
THE ABLATION STUDY ON ADE/FDE (KM)/GLEV COMPARISON ACROSS $k = 1 - 15$

Methods	k=15	k=10	k=5	k=1
TrajAirNet [14]	0.80/1.61/0.44	0.82/1.68/0.42	0.93/1.80/0.62	1.04/2.01/nan
GooDFlight(Ours)	**0.30/0.45/0.11**	**0.35/0.55/0.18**	**0.45/0.81/0.47**	**0.79/1.53**/nan

TABLE IX
ABLATION STUDY

Incremental Methods↓	IGD Predictor	JGD Predictor	DWGAT	IAM	ADE(km)↓	FDE(km)↓	GLeV↓
Base model (Diffusion)	×	×	×	×	0.53	0.71	0.25
+ IGD Predictor	✓	×	×	×	0.38	0.58	0.17
+ JGD Predictor	✓	✓	×	×	0.34	0.49	0.12
+ DWGAT	✓	✓	✓	×	0.32	0.45	0.10
+ IAM	✓	✓	✓	✓	**0.29**	**0.39**	**0.08**

our model can still output performance comparable to that of TrajAirNet when $k = 15$, demonstrating the powerful pattern recognition ability. As shown in Fig. 12 we visualize the predicted trajectory when k = 1, our model can still predict reasonable trajectories.

V. CONCLUSION AND FUTURE WORK

In this work, we propose a novel framework called GooDFlight for trajectory prediction within terminal airspace at non-towered general aviation airports. For the first time, we explored this task from the perspective of flight pattern coverage and designed a two-stage solution in the end-to-end training process. At the same time, a One-then-all goal estimation pipeline is innovatively proposed to enhance the trajectory prediction diversity for anchor-based methods. Finally, we also designed a new evaluation metric called GLeV to describe the diversity and social acceptability of candidate trajectories. Comprehensive experimental results on large-scale real-world datasets have demonstrated that the proposed model significantly outperforms existing state-of-the-art methods both quantitatively and qualitatively. The effectiveness of all technological innovations has been demonstrated through ablation studies and result-analysis sessions. Furthermore, our evaluation metric also demonstrates its generality in visual comparisons of multiple methods.

Beyond the technological improvement, we fully revealed the decisive role of diversity in this task and the pattern imbalance of flight trajectories for the first time. Hopefully, these insights can better guide future research on trajectory prediction in terminal airspace of non-towered airports.

Finally, we present the possible limitations of our method as well as the future research directions. Although methods based on historical behaviors, such as Expert-Traj and our proposed GoodFlight, can achieve leading results in the task of aircraft trajectory prediction, we have to consider the situation where there is a lack of similar historical behaviors. For example, the flight pattern of a fighter jet is very likely to not match any general aviation historical behaviors. In such a case, the wrong historical behaviors will be matched, this mismatched condition will then lead to the failure of the entire model. Fortunately, our model can adjust the intensity of the historical guidance. If the above situation occurs, our model will ignore the intention estimation and degenerate into a diffusion model, generating reasonable trajectories despite the significant decline in performance.

An interesting future research direction is to explore the multi-modal flight trajectory prediction. Since different flight patterns can be described by a specific set of instructions, jointly predicting future trajectories and simultaneously generating a language description of the prediction can make it easier for pilots to understand in highly dynamic flights. In the future, it may appear as an artificial intelligence virtual tower in a non-towered airport scenario to assist navigation.

REFERENCES

[1] Y. Lin, J.-w. Zhang, and H. Liu, "Deep learning based short-term air traffic flow prediction considering temporal–spatial correlation," *Aerospace Science and Technology*, vol. 93, p. 105113, 2019.

[2] Z. Wang, M. Liang, and D. Delahaye, "A hybrid machine learning model for short-term estimated time of arrival prediction in terminal manoeuvring area," *Transportation Research Part C: Emerging Technologies*, vol. 95, pp. 280–294, 2018.

[3] W. Liu and I. Hwang, "Probabilistic trajectory prediction and conflict detection for air traffic control," *Journal of Guidance, Control, and Dynamics*, vol. 34, no. 6, pp. 1779–1789, 2011.

[4] Z. Chen, D. Guo, and Y. Lin, "A deep gaussian process-based flight trajectory prediction approach and its application on conflict detection," *Algorithms*, vol. 13, no. 11, p. 293, 2020.

[5] F. A. Administration. (2018) General aviation safety. [Online]. Available: https://www.faa.gov/newsroom/general-aviation-safety

[6] N. T. S. Board. (2019) Annual summary of us civil aviation accidents. [Online]. Available: https://www.ntsb.gov/safety/data/Pages/AviationDataStats2019.aspx

[7] M. E. Johnson and Y. Gu, "Estimating airport operations at general aviation airports using the faa npias airport categories," *International Journal of Aviation, Aeronautics, and Aerospace*, vol. 4, no. 1, p. 3, 2017.

[8] J. Zhang, J. Liu, R. Hu, and H. Zhu, "Online four dimensional trajectory prediction method based on aircraft intent updating," *Aerospace Science and Technology*, vol. 77, pp. 774–787, 2018.

[9] R. Dalmau, M. Pérez-Batlle, and X. Prats, "Real-time identification of guidance modes in aircraft descents using surveillace data," in *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*. IEEE, 2018, pp. 1–10.

[10] R. Dalmau, X. Prats, R. Verhoeven, F. Bussink, and B. Heesbeen, "Comparison of various guidance strategies to achieve time constraints in optimal descents," *Journal of Guidance, Control, and Dynamics*, vol. 42, no. 7, pp. 1612–1621, 2019.

[11] D. Guo, E. Q. Wu, Y. Wu, J. Zhang, R. Law, and Y. Lin, "Flightbert: binary encoding representation for flight trajectory prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 24, no. 2, pp. 1828–1842, 2022.

[12] Z. Shi, M. Xu, and Q. Pan, "4-d flight trajectory prediction with constrained lstm network," *IEEE transactions on intelligent transportation systems*, vol. 22, no. 11, pp. 7242–7255, 2020.

[13] A. Nuic, "User manual for the base of aircraft data (bada) revision 3.10," *Atmosphere*, vol. 2010, p. 001, 2010.

[14] J. Patrikar, B. Moon, J. Oh, and S. Scherer, "Predicting like a pilot: Dataset and method to predict socially-aware aircraft trajectories in non-towered terminal airspace," in *2022 international conference on robotics and automation (icra)*. IEEE, 2022, pp. 2525–2531.

[15] J. Patrikar, J. Dantas, B. Moon, M. Hamidi, S. Ghosh, N. Keetha, I. Higgins, A. Chandak, T. Yoneyama, and S. Scherer, "Tartanaviation: Image, speech, and ads-b trajectory datasets for terminal airspace operations," 2024. [Online]. Available: https://arxiv.org/pdf/2403.03372.pdf

[16] I. Navarro and J. Oh, "Social-patternn: Socially-aware trajectory prediction guided by motion patterns," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2022, pp. 9859–9864.

[17] C.-F. Lin, A. G. Ulsoy, and D. J. LeBlanc, "Vehicle dynamics and external disturbance estimation for vehicle path prediction," *IEEE Transactions on Control Systems Technology*, vol. 8, no. 3, pp. 508–518, 2000.

[18] N. Kaempchen, B. Schiele, and K. Dietmayer, "Situation assessment of an autonomous emergency brake for arbitrary vehicle-to-vehicle collision scenarios," *IEEE Transactions on Intelligent Transportation Systems*, vol. 10, no. 4, pp. 678–687, 2009.

[19] B. Jin, B. Jiu, T. Su, H. Liu, and G. Liu, "Switched kalman filter-interacting multiple model algorithm based on optimal autoregressive model for manoeuvring target tracking," *IET Radar, Sonar & Navigation*, vol. 9, no. 2, pp. 199–209, 2015.

[20] R. Zhang, L. Cao, S. Bao, and J. Tan, "A method for connected vehicle trajectory prediction and collision warning algorithm based on v2v communication," *International Journal of Crashworthiness*, vol. 22, no. 1, pp. 15–25, 2017.

[21] D. J. Phillips, T. A. Wheeler, and M. J. Kochenderfer, "Generalizable intention prediction of human drivers at intersections," in *2017 IEEE intelligent vehicles symposium (IV)*. IEEE, 2017, pp. 1665–1670.

[22] M.-F. Chang, J. Lambert, P. Sangkloy, J. Singh, S. Bak, A. Hartnett, D. Wang, P. Carr, S. Lucey, D. Ramanan *et al.*, "Argoverse: 3d tracking and forecasting with rich maps," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 8748–8757.

[23] Y. Xing, C. Lv, and D. Cao, "Personalized vehicle trajectory prediction based on joint time-series modeling for connected vehicles," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 2, pp. 1341–1352, 2019.

[24] S. Capobianco, N. Forti, L. M. Millefiori, P. Braca, and P. Willett, "Recurrent encoder–decoder networks for vessel trajectory prediction with uncertainty estimation," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 3, pp. 2554–2565, 2023.

[25] J. Zhang, Z. Shi, A. Zhang, Q. Yang, G. Shi, and Y. Wu, "Uav trajectory prediction based on flight state recognition," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 60, no. 3, pp. 2629–2641, 2024.

[26] S. Capobianco, L. M. Millefiori, N. Forti, P. Braca, and P. Willett, "Deep learning methods for vessel trajectory prediction based on recurrent neural networks," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 6, pp. 4329–4346, 2021.

[27] 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.

[28] V. Kosaraju, A. Sadeghian, R. Martín-Martín, I. Reid, H. Rezatofighi, and S. Savarese, "Social-bigat: Multimodal trajectory forecasting using bicycle-gan and graph attention networks," *Advances in neural information processing systems*, vol. 32, 2019.

[29] J. Amirian, J.-B. Hayet, and J. Pettré, "Social ways: Learning multi-modal distributions of pedestrian trajectories with gans," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[30] P. Dendorfer, S. Elflein, and L. Leal-Taixé, "Mg-gan: A multi-generator model preventing out-of-distribution samples in pedestrian trajectory prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 13 158–13 167.

[31] 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.

[32] G. Chen, J. Li, N. Zhou, L. Ren, and J. Lu, "Personalized trajectory prediction via distribution discrimination," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 15 580–15 589.

[33] M. Halawa, O. Hellwich, and P. Bideau, "Action-based contrastive learning for trajectory prediction," in *European Conference on Computer Vision*. Springer, 2022, pp. 143–159.

[34] P. Xu, J.-B. Hayet, and I. Karamouzas, "Socialvae: Human trajectory prediction using timewise latents," in *European Conference on Computer Vision*. Springer, 2022, pp. 511–528.

[35] E. Rehder and H. Kloeden, "Goal-directed pedestrian prediction," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2015, pp. 50–58.

[36] K. Mangalam, H. Girase, S. Agarwal, K.-H. Lee, E. Adeli, J. Malik, and A. Gaidon, "It is not the journey but the destination: Endpoint conditioned trajectory prediction," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part II 16*. Springer, 2020, pp. 759–776.

[37] 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.

[38] P. Dendorfer, A. Osep, and L. Leal-Taixé, "Goal-gan: Multimodal trajectory prediction based on goal position estimation," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[39] R. Rezaie and X. R. Li, "Destination-directed trajectory modeling, filtering, and prediction using conditionally markov sequences," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 57, no. 2, pp. 820–833, 2021.

[40] H. Zhao and R. P. Wildes, "Where are you heading? dynamic trajectory prediction with expert goal examples," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 7629–7638.

[41] C. Xu, W. Mao, W. Zhang, and S. Chen, "Remember intentions: Retrospective-memory-based trajectory prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 6488–6497.

[42] F. Marchetti, F. Becattini, L. Seidenari, and A. D. Bimbo, "Mantra: Memory augmented networks for multiple trajectory prediction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 7143–7152.

[43] M. Meng, Z. Wu, T. Chen, X. Cai, X. Zhou, F. Yang, and D. Shen, "Forecasting human trajectory from scene history," *Advances in Neural Information Processing Systems*, vol. 35, pp. 24 920–24 933, 2022.

[44] 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.

[45] M. G. Hamed, D. Gianazza, M. Serrurier, and N. Durand, "Statistical prediction of aircraft trajectory: regression methods vs point-

mass model," in *ATM 2013, 10th USA/Europe Air Traffic Management Research and Development Seminar*, 2013, pp. pp–xxxx.
[46] Y. Fukuda, M. Shirakawa, and A. Senoguchi, "Development and evaluation of trajectory prediction model," in *Proceedings of the 27th international congress of the aeronautical sciences*, 2010, pp. 1–8.
[47] I. Lymperopoulos and J. Lygeros, "Sequential monte carlo methods for multi-aircraft trajectory prediction in air traffic management," *International Journal of Adaptive Control and Signal Processing*, vol. 24, no. 10, pp. 830–849, 2010.
[48] R. Rezaie and X. R. Li, "Trajectory modeling and prediction with waypoint information using a conditionally markov sequence," in *2018 56th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2018, pp. 486–493.
[49] C. E. Seah and I. Hwang, "A hybrid estimation algorithm for terminal-area aircraft tracking," in *AIAA guidance, navigation and control conference and exhibit*, 2007, p. 6691.
[50] Z. Wang, M. Liang, and D. Delahaye, "Short-term 4d trajectory prediction using machine learning methods," in *SID 2017, 7th SESAR Innovation Days*, 2017.
[51] A. De Leege, M. van Paassen, and M. Mulder, "A machine learning approach to trajectory prediction," in *AIAA Guidance, Navigation, and Control (GNC) Conference*, 2013, p. 4782.
[52] Z. Shi, M. Xu, Q. Pan, B. Yan, and H. Zhang, "Lstm-based flight trajectory prediction," in *2018 International joint conference on neural networks (IJCNN)*. IEEE, 2018, pp. 1–8.
[53] J. Ren, X. Wu, Y. Liu, F. Ni, Y. Bo, and C. Jiang, "Long-term trajectory prediction of hypersonic glide vehicle based on physics-informed transformer," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 59, no. 6, pp. 9551–9561, 2023.
[54] Y. Fan, Y. Tan, L. Wu, H. Ye, and Z. Lyu, "Global and local interattribute relationships-based graph convolutional network for flight trajectory prediction," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 60, no. 3, pp. 2642–2657, 2024.
[55] F. A. Administration, *Airplane flying handbook (FAA-H-8083-3A)*. Skyhorse Publishing Inc., 2011.
[56] Y. Yin, S. Zhang, Y. Zhang, Y. Zhang, and S. Xiang, "Context-aware aircraft trajectory prediction with diffusion models," in *2023 IEEE 26th International Conference on Intelligent Transportation Systems (ITSC)*. IEEE, 2023, pp. 5312–5317.
[57] ——, "Aircraft trajectory prediction in terminal airspace with intentions derived from local history," *Neurocomputing*, vol. 615, p. 128843, 2025.
[58] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in neural information processing systems*, vol. 33, pp. 6840–6851, 2020.
[59] Z. Yuan, C. Hao, R. Zhou, J. Chen, M. Yu, W. Zhang, H. Wang, and X. Sun, "Efficient and controllable remote sensing fake sample generation based on diffusion model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–12, 2023.
[60] L. Liu, B. Chen, H. Chen, Z. Zou, and Z. Shi, "Diverse hyperspectral remote sensing image synthesis with diffusion models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.
[61] Z. Zhao, X. Dong, Y. Wang, and C. Hu, "Advancing realistic precipitation nowcasting with a spatiotemporal transformer-based denoising diffusion model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024.
[62] X. Zou, K. Li, J. Xing, Y. Zhang, S. Wang, L. Jin, and P. Tao, "Diffcr: A fast conditional diffusion framework for cloud removal from optical satellite images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.
[63] Y. Wen, X. Ma, X. Zhang, and M.-O. Pun, "Gcd-ddpm: A generative change detection model based on difference-feature-guided ddpm," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–16, 2024.
[64] Y. Xiao, Q. Yuan, K. Jiang, J. He, X. Jin, and L. Zhang, "Ediffsr: An efficient diffusion probabilistic model for remote sensing image super-resolution," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.
[65] N. Chen, J. Yue, L. Fang, and S. Xia, "Spectraldiff: A generative framework for hyperspectral image classification with diffusion models," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 61, pp. 1–16, 2023.
[66] X. Wang, Z. Wang, Z. Xiong, Y. Yang, C. Zhu, and J. Gao, "Reconstructing regularly missing seismic traces with a classifier-guided diffusion model," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–14, 2024.
[67] T. Gu, G. Chen, J. Li, C. Lin, Y. Rao, J. Zhou, and J. Lu, "Stochastic trajectory prediction via motion indeterminacy diffusion," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17 113–17 122.
[68] W. Mao, C. Xu, Q. Zhu, S. Chen, and Y. Wang, "Leapfrog diffusion model for stochastic trajectory prediction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 5517–5526.
[69] K. Wu, Y. Zhou, H. Shi, X. Li, and B. Ran, "Graph-based interaction-aware multimodal 2d vehicle trajectory prediction using diffusion graph convolutional networks," *IEEE Transactions on Intelligent Vehicles*, 2023.
[70] C. Liu, S. He, H. Liu, and J. Chen, "Intention-aware denoising diffusion model for trajectory prediction," *arXiv preprint arXiv:2403.09190*, 2024.
[71] J. Song, C. Meng, and S. Ermon, "Denoising diffusion implicit models," *arXiv preprint arXiv:2010.02502*, 2020.
[72] G. Aydemir, A. K. Akan, and F. Güney, "Adapt: Efficient multi-agent trajectory prediction with adaptation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 8295–8305.
[73] Q. Zhang, X. Liao, Q. Liu, J. Xu, and B. Zheng, "Leaving no one behind: A multi-scenario multi-task meta learning approach for advertiser modeling," in *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, 2022, pp. 1368–1376.
[74] Z. L. Li, G. W. Zhang, J. Yu, and L. Y. Xu, "Dynamic graph structure learning for multivariate time series forecasting," *Pattern Recognition*, vol. 138, p. 109423, 2023.
[75] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel, "Social-stgcnn: A social spatio-temporal graph convolutional neural network for human trajectory prediction," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 14 424–14 432.
[76] C. Schöller, V. Aravantinos, F. Lay, and A. Knoll, "What the constant velocity model can teach us about pedestrian motion prediction," *IEEE Robotics and Automation Letters*, vol. 5, no. 2, pp. 1696–1703, 2020.
[77] T. Salzmann, B. Ivanovic, P. Chakravarty, and M. Pavone, "Trajectron++: Dynamically-feasible trajectory forecasting with heterogeneous data," in *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVIII 16*. Springer, 2020, pp. 683–700.
[78] F. Giuliari, I. Hasan, M. Cristani, and F. Galasso, "Transformer networks for trajectory forecasting," in *2020 25th international conference on pattern recognition (ICPR)*. IEEE, 2021, pp. 10 335–10 342.
[79] A. Monti, A. Bertugli, S. Calderara, and R. Cucchiara, "Dag-net: Double attentive graph neural network for trajectory forecasting," in *2020 25th International Conference on Pattern Recognition (ICPR)*. IEEE, 2021, pp. 2551–2558.
[80] D. Zhao and J. Oh, "Noticing motion patterns: A temporal cnn with a novel convolution operator for human trajectory prediction," *IEEE Robotics and Automation Letters*, vol. 6, no. 2, pp. 628–634, 2020.