Truncated Diffusion Process for Temporal Consistent Inference

Qingyuan Jiang[1] and Volkan Isler[1,2]

***Abstract*—Diffusion models have demonstrated strong capabilities for modeling complex multimodal distributions in robotics applications such as human motion prediction, yet their practical deployment is hindered by slow inference speed due to iterative denoising steps. In this paper, we propose a truncated diffusion process that exploits temporal consistency in sequential prediction tasks to significantly accelerate inference while maintaining prediction quality. Our key insight is that in consequential inference problems, samples generated at the previous time step provide valuable priors that are similar to the current prediction target, enabling us to skip a substantial portion of the denoising process. The proposed method operates in two stages: first, we add partial noise to samples from the previous time step to reach an intermediate noise level; second, we denoise these samples with updated conditioning to generate predictions for the current time step. We provide a theoretical analysis showing that the KL divergence between our truncated process and the full diffusion process is upper-bounded by the conditioning change, scaled by a contraction factor that depends on the truncation ratio. We validate our approach on the human trajectory prediction task using both simulated (GTA-IM) and real-world (HPS) datasets. Experimental results demonstrate that our method achieves comparable or better performance than state-of-the-art baselines while using significantly fewer diffusion steps.**

I. INTRODUCTION

Diffusion models [15, 37, 38] are proposed to model complex data distributions for image generation and are now widely used in robotics to generate action sequences [8, 6, 22, 18], or to predict human motions [16, 1]. However, a major challenge of applying diffusion models to robotics in real time is their relatively slow inference speed. A few methods have been proposed to speed up inference [37, 39, 12] by skipping denoising steps in general-purpose applications. However, these methods either sacrifice sample quality or remain computationally expensive.

One property of using diffusion models in robotics, which has yet to be explored, is that the system continually makes predictions over time. In other words, before each inference timestamp, there exists a set of samples from the previous time step that should be 'similar' to the sampling results at the current time. These existing samples can be treated as 'prior' and used to speed up inference. In fact, the idea of using 'prior' is not new in the image inpainting task [38, 42], or in the super-resolution task [37]. In this paper, we borrow these ideas to a consequential human trajectory prediction task, where we naturally have priors, and propose the truncated diffusion process that generates new samples in two stages: (1) we add partial noise to the samples from the previous time step, and (2) we denoise the intermediate noisy samples with the new conditioning.

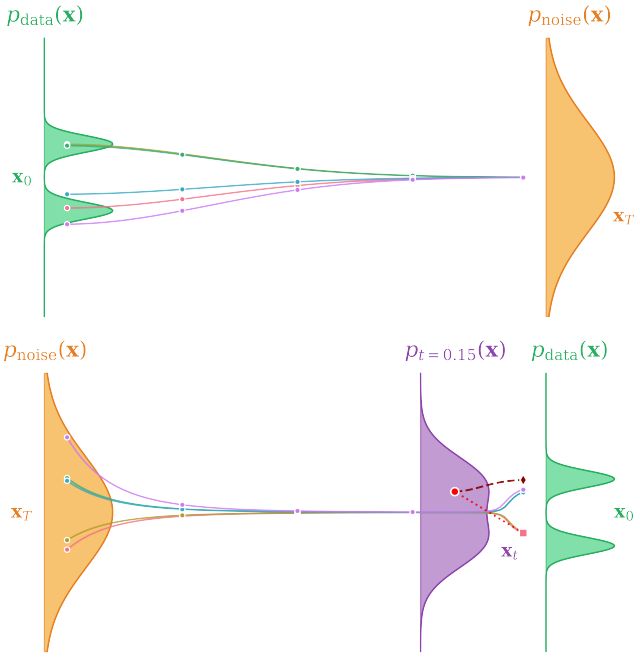


Fig. 1: **1D Mixture of Gaussians Example for Truncated Diffusion Process**. In this toy example, we show the diffusion (**Top**) and denoising process (**Bottom**) of a 1D diffusion model, where the raw data distribution is a mixture of two Gaussians. For each sample, we show the generation process by the corresponding Ordinary Differential Equation (ODE) [38] with a solid line. Additionally, we illustrate the proposed **truncated diffusion process** in the **Bottom**: A sample from the previous time step (from one Gaussian branch) is used as the prior. We add partial (15%) noise to this sample (red dashed line), and denoise (purple dashed line) with proper conditioning to get a new sample that falls from the other Gaussian branch.

The toy example in Fig. 1 explains the intuition behind. In the Fig. 1.bottom, we show a classic diffusion model's sampling process towards a 1D mixture of Gaussian distributions from a zero-mean Gaussian. We visualize that the ODEs from different samples will first converge in the early steps of the denoising process, then diverge in the last few steps. This implies that only a small portion of the last denoising steps is essential for generating samples that fall into different modes of the Gaussian, and the samples at these intermediate noise levels have great values [10]. Therefore, despite the human trajectory being in a higher-dimensional space and a much more complex distribution, our intuition is that, with proper

[1] Department of Computer Science & Engineering, University of Minnesota, 4-192 Keller Hall, 200 Union Street SE, Minneapolis, MN 55455, US

[2] Department of Computer Science, University of Texas at Austin, 2317 Speedway, Austin, Texas 78712, US

conditioning guidance, we can use the existing samples from the previous time step, diffuse them into an intermediate noise level, and only denoise the last few steps to generate samples that fall into a different sub-mode of the distribution.

One challenge for doing so in such long-term, consequential prediction is that the conditioning for the diffusion model keeps changing over time, which may introduce compounding error. Therefore, we theoretically investigate the performance of the truncated diffusion process in approximating the target distribution and reveal its relationship with changes in the conditioning and noise level. More specifically, we show that the approximation error of the truncated diffusion process is upper bounded by a function of the change of the conditioning, multiplied by a factor related to the noise level. We also conduct experiments to evaluate the performance of the truncated diffusion process on a human trajectory prediction task on both simulation and real-world datasets. We show that, with a proper noise level, the truncated diffusion process can achieve performance comparable to or better than the full diffusion process with significantly fewer denoising steps.

Our main contribution is threefold:

1) We propose a truncated diffusion process that exploits temporal consistency in sequential prediction tasks to significantly accelerate inference while maintaining prediction quality.
2) We provide a theoretical analysis of the truncated diffusion process, showing how the approximation error relates to changes in conditioning and the noise level.
3) We conduct experiments to evaluate our method on both simulation and real-world datasets. We show that the proposed method can efficiently predict human trajectories across consecutive frames, achieving comparable or better performance while significantly reducing the number of denoising steps.

II. RELATED WORK

We first summarize the related literature on human motion prediction, then focus on diffusion processes and their acceleration techniques for similar tasks.

A. Human Motion Prediction

Human motion prediction [26] is a well-defined computer vision problem that has many potential applications [31, 21, 27], while remaining challenging in long-term forecasting. Varies of neural network architectures have been proposed to address this challenge, including using Multi-Layer Perceptrons (MLP) [3], Generative Adversarial Networks (GANs) [32], Graph Convolutional Networks (GCN) [28], Transformers [7, 24, 29, 2], and recently diffusion models [40, 16, 1, 4]. In this paper, proposing a new diffusion model architecture is not our focus; rather, we use a state-of-the-art method [16] as our base model and study the diffusion process itself, specifically the possibility of using priors to speed up inference.

B. Truncated Diffusion Process

Since the success of the diffusion model in image generation [37, 11], various works have been proposed to accelerate the inference process by reducing the number of denoising steps [37, 39, 12, 23]. Among them, some methods [42, 30, 25, 34] use 'prior' to skip denoising steps by starting from a partially noisy image rather than pure Gaussian noise. The 'prior' either comes from an image with missing regions [30], or a similar image from the same category [25], and the problem is formulated as a warm-start [34] strategy or an evolutionary strategy [41] for the diffusion models. In this work, we extend these ideas to the domain of human motion prediction, where we can naturally use predictions from the previous time step as the 'prior'. Further, we provide a theoretical analysis of bounding the distributional difference between a truncated diffusion process and the original method, and demonstrate its effectiveness through experiments.

III. PROBLEM FORMULATION

In this paper, we study this truncated diffusion process in the context of human motion prediction task. We will first introduce the formulation on the human motion prediction task itself, then formulate it in the context of truncated diffusion.

A. Human Motion Prediction

The human motion prediction task defines the following problem: Given the past T-frame human motion $\mathbf{x}_{-(T+1):-1}$ and the surrounding environment $\mathbf{S}_0$, predict the future human motion $\mathbf{x}_{0:T}$. Here the human motion, $\mathbf{x}$, is defined as a sequence of human poses over time. Depending on the application scenario, human motion can be represented in different fidelity. For example, it can be defined as a sequence of 2D positions, and hence $\mathbf{x} \in \mathbb{R}^{T\times 2}$. Alternatively, it can be a sequence of 3D full-body skeleton, where $\mathbf{x} \in \mathbb{R}^{T\times(J\times 3)}$ with J joints in the human pose at each time step. In this paper, for simplicity, we will use the 2D position representation for human motion, and the surrounding environment is represented as a 2D occupancy grid.

Further, from the 'inference as a sampling' perspective [22], the human motion prediction task can be formulated as generating multiple possible future human motion samples $\mathbf{x}_{0:T}$ from the conditional distribution $p_\theta(\mathbf{x}_{0:T}|\mathbf{x}_{-(T+1):-1}, \mathbf{S}_0)$, where θ is the parameter of the model. . Simplifying the notation by using $\mathbf{x}_0$ to denote $\mathbf{x}_{0:T}$, and define the conditioning $\mathbf{c_0} := \{\mathbf{x}_{-(T+1):-1}, \mathbf{S}_0\}$, we have the following formulation:

$$\mathbf{x}_0 \sim p_\theta(\mathbf{x}_0|\mathbf{c_0}) \tag{1}$$

B. Utilizing Temporal Consistency

Suppose we have already generated a set of human motion samples at current time, denoted as $\mathcal{X}_0 = \{\mathbf{x}_0\}$, following the distribution above. At the next time step, we would like to generate another set of human motion samples, $\mathcal{X}_1 = \{\mathbf{x}_1\}$, under the new condition $\mathbf{c_1} = \{\mathbf{x}_{-T:0}, \mathbf{S}_1\}$, i.e. $\mathbf{x}_1 \sim p_\theta(\mathbf{x}_1|\mathbf{c_1})$. In this paper, we study how to make use of the previously generated samples $\mathcal{X}_0$ to speed up the generation of $\mathcal{X}_1$, by leveraging the temporal consistency of the human motion.

IV. APPROACH

In this work, we present a truncated diffusion process that generates human motion prediction samples based on the previously generated samples. We will first introduce the proposed truncated diffusion process. Then we will provide a theoretical analysis of the validity of such method, by providing a bound on the distributional difference between the truncated diffusion process and the original diffusion process.

A. Truncated Diffusion Process

Following the formulated problem in Sec. III-B, suppose we have already generated a set of human motion samples $\mathcal{X}_0 = \{\mathbf{x}_0^0\}$. At the next time step, in order to generate the new samples $\mathcal{X}_1$, we propose the following truncated diffusion process: **(Step 1): Diffusion**: Add noise to the previous samples $\mathbf{x}_0^0$ with k diffusion steps, so that we get intermediate noisy samples $\mathbf{x}_0^k$. Note that here we use superscript to denote the diffusion time steps out of total M-steps, where 0 means the final generated samples, $\mathbf{x}^M$ means the initial pure Gaussian noise. This is different from the real time step index in the human motion sequence in the subscript. **(Step 2): Denoising**: Denoise the noisy samples $\mathbf{x}_0^k$ with the new condition $\mathbf{c}_1$, to get the new samples $\hat{\mathbf{x}}_1^0$.

This process is presented in Algo. 1. Note that since the diffusion process can be derived with an analytical method [35], this truncated process uses only k/M of the total diffusion steps, and therefore significantly speeds up the inference process.

Algorithm 1 Truncated Diffusion Process

Input: Previous samples $\mathbf{x}_0^0$, new condition $\mathbf{c}_1$, truncation steps k

Output: New samples $\hat{\mathbf{x}}_1^0$ approximately from $p_\theta(\mathbf{x}_1|\mathbf{c}_1)$

1: **Diffusion step:** Add noise to $\mathbf{x}_0^0$ with k steps
2: $\epsilon \sim \mathcal{N}(0, I)$
3: $\mathbf{x}_0^k \leftarrow \sqrt{\bar{\alpha}_k}\,\mathbf{x}_0^0 + \sqrt{1-\bar{\alpha}_k}\epsilon$
4: **Denoising step:** Denoise noisy samples with new condition
5: $\hat{\mathbf{x}}_1^0 \leftarrow \text{Denoise}(\mathbf{x}_0^k, \mathbf{c}_1)$
6: **return** $\hat{\mathbf{x}}_1^0$

B. Analysis from Probability View

Here we analysis the truncated diffusion process from a probability view. Define the output distribution from the truncated diffusion process, originally conditioned on $\mathbf{c}_0$ and later on $\mathbf{c}_1$, as $Q(\mathbf{x}_1|\mathbf{c}_1, \mathbf{c}_0)$. In the following section, we will show that the distribution difference between $Q(\mathbf{x}_1|\mathbf{c}_1, \mathbf{c}_0)$ and the original diffusion model distribution $p(\mathbf{x}_1|\mathbf{c}_1)$ is bounded by the following inequality:

$$\begin{aligned} &D_{\text{KL}}\left(Q(\mathbf{x}_1|\mathbf{c}_1, \mathbf{c}_0) \,||\, p(\mathbf{x}_1|\mathbf{c}_1)\right) \\ &\quad \leq \bar{\alpha}_k\, D_{\text{KL}}\left(p(\mathbf{x}_1|\mathbf{c}_0) \,||\, p(\mathbf{x}_1|\mathbf{c}_1)\right) \end{aligned} \tag{2}$$

Here $\bar{\alpha}_k$ is the noise scheduler parameter after k diffusion steps [35, 38], which is a value between 0 and 1, and increases with k. In the following section, we will first formulate the proposed algorithm as a Markov kernel, and define the corresponding stochasitic transition. Then we will show bounding analysis on the KL divergence on Eq. 2.

Definition 1: Truncated Diffusion Process as a Markov Kernel. Given a existing sample from an arbitrary distribution $\mathbf{x} \sim P(\mathbf{x})$, and a new condition $\mathbf{c}$, the probability of a k-step truncated diffusion process on a new sample $\mathbf{x}'$ can be defined by a Markov kernel [1]:

$$K_{\mathbf{c}}^{(k)} := p(\mathbf{x}'|\mathbf{x}, \mathbf{c}) = \int_{\mathbf{x}^k} p(\mathbf{x}'|\mathbf{x}^k, \mathbf{c})\, q(\mathbf{x}^k|\mathbf{x})\, \mathrm{d}\mathbf{x}^k \tag{3}$$

Note that we follow the convention of diffusion model [36], where we define the forward diffusion process as $q(\mathbf{x}^k|\mathbf{x})$, and the reverse denoising process as $p(\mathbf{x}'|\mathbf{x}^k, \mathbf{c})$.

Lemma 1: Transition to a Distribution. Apply this kernel to the existing a sample distribution $P(\mathbf{x})$. The output distribution $Q(\mathbf{x}'|\mathbf{c})$ from the truncated diffusion process can be written as:

$$Q(\mathbf{x}'|\mathbf{c}) = \int_{\mathbf{x}} p(\mathbf{x}'|\mathbf{x}, \mathbf{c})\, P(\mathbf{x})\, \mathrm{d}\mathbf{x} \tag{4}$$

Proof has been attached in Appendix VII-A2. In the meantime, we simplify the notation by $Q(\mathbf{x}'|\mathbf{c}) := K_{\mathbf{c}}^{(k)} P(\mathbf{x})$. This way, we interprete the truncated diffusion proess as a transition from an arbitrary existing sample distribution $P(\mathbf{x})$ to a new distribution $Q(\mathbf{x}'|\mathbf{c})$, through the defined Markov kernel $K_{\mathbf{c}}^{(k)}$.

Analysis Goal. Simplify the notation by $p_0 := p(\mathbf{x}_0|\mathbf{c}_0)$, and $p_1 := p(\mathbf{x}_1|\mathbf{c}_1)$. Applying the defined transition in Eq. 4 to p_0, we get the target output distribution as $Q_1 := Q(\mathbf{x}_1|\mathbf{c}_1, \mathbf{c}_0) = K_{\mathbf{c}_1}^{(k)} \circ p_0$. Here $\circ$ defines the transition operation on the distribution. In the meantime, note that if we apply the kernel to p_1, the transition is identity since they are under the same condition $\mathbf{c_1}$, i.e. $K_{\mathbf{c_1}}^{(k)} p_1 = p_1$. [2] The target distribution difference in Eq. 2 can be written as:

Proposition. The distribution difference between the truncated diffusion process and the original diffusion model distribution is bounded by

$$\begin{aligned} D_{\text{KL}}\left(Q_1 \,||\, p_1\right) &= D_{\text{KL}}\left(K_{\mathbf{c}_1}^{(k)} \circ p_0 \,||\, K_{\mathbf{c}_1}^{(k)} \circ p_1\right) \\ &\leq \bar{\alpha}_k\, D_{\text{KL}}\left(p_0 \,||\, p_1\right) \end{aligned} \tag{5}$$

Proof. Revisiting Eq. 3, our Markov kernel can be decomposed into two steps: (a) the noise injection kernel, and (b) the truncated denoising kernel. Each can be formulated as a separate Markov kernel as in Eq. 6. We will investigate the properties for each kernal separately.

$$\begin{aligned} &K_{\mathbf{c_1}}^{(k)} = K_{\text{denoise},\mathbf{c_1}}^{(k)} \circ K_{\text{noise}}^{(k)} \\ &K_{\text{noise}}^{(k)} := p(\mathbf{x}^k|\mathbf{x}_0^0) = \mathcal{N}(\mathbf{x}^k; \sqrt{\bar{\alpha}_k}\,\mathbf{x}_0^0, (1-\bar{\alpha}_k)I) \\ &K_{\text{denoise},\mathbf{c_1}}^{(k)} := p(\mathbf{x}_1^0|\mathbf{x}^k, \mathbf{c}_1) = \Pi_{i=1}^k \mathcal{N}(\mathbf{x}^{i-1}; \mu_\theta(\mathbf{x}^i, \mathbf{c}), \Sigma_\theta^i) \end{aligned} \tag{6}$$

[1] Refer to Appendix VII-A1 for details.
[2] Refer to Appendix VII-A3 for details.

Lemma 2: KL Contractivity of Noise Injection. Define the intermediate distributions after noise injection as $p^k_{\text{noise},0} = K^{(k)}_{\text{noise}} \circ p_0$ and $p^k_{\text{noise},1} = K^{(k)}_{\text{noise}} \circ p_1$. Since the noise injection process, $\mathbf{x}^0 \to \mathbf{x}^k = \sqrt{\bar{\alpha}_k}\,\mathbf{x}^0 + \sqrt{1-\bar{\alpha}_k}\epsilon$, $\epsilon \sim \mathcal{N}(0, I)$, is exactly the form of Gaussian channel [33]. We can apply the Gaussian Strong Data Processing Inequality (SDPI) [33, 20], where we have:

$$D_{\text{KL}}\left(p^k_{\text{noise},0} \,||\, p^k_{\text{noise},1}\right) \leq \bar{\alpha}_k \, D_{\text{KL}}\left(p_0 \,||\, p_1\right) \tag{7}$$

Lemma 3: KL Contractivity of Denoising Step. In this step, we use the standard DDPM sampling procedure [15] but simply starting from an intermediate noised samples. As shown in Eq. 6, decompose the denoising step by $K^{(k)}_{\text{denoise}} = \prod_{i=1}^{k} K^{(i)}_{\text{denoise}}$, where $K^{(i)}_{\text{denoise}} = \mathcal{N}(\mathbf{x}^{i-1}; \mu_\theta(\mathbf{x}^i, \mathbf{c}), \Sigma^i)$. For each denoising step, by Data Processing Inequality (DPI) of KL divergence [9], we have: $D_{\text{KL}}\left(K^{(i)}_{\text{denoise}} \circ p^i_0 \,||\, K^{(i)}_{\text{denoise}} \circ p^i_1\right) \leq D_{\text{KL}}\left(p^i_0 \,||\, p^i_1\right)$. Therefore, starting from the intermediate noised distribution p^k_0 and p^k_1, by applying this inequality recursively from $i = k, k-1, \ldots, 1$, we have:

$$D_{\text{KL}}\left(K^{(k)}_{\text{denoise}} \circ p^k_0 \,||\, K^{(k)}_{\text{denoise}} \circ p^k_1\right) \leq D_{\text{KL}}\left(p^k_0 \,||\, p^k_1\right) \tag{8}$$

Combining the two steps. Combining Lemmas 2 (Eq. 7) and Lemmas 3 (Eq. 8), we prove the proposition as in Eq. 2 and Eq. 5.

$$\begin{aligned}
&D_{\text{KL}}\left(K^{(k)}_{\mathbf{c}_1} p_0 \,||\, K^{(k)}_{\mathbf{c}_1} p_1\right) \\
&= D_{\text{KL}}\left(K^{(k)}_{\text{denoise}} \circ K_{\text{noise}}\, p_0 \,||\, K^{(k)}_{\text{denoise}} \circ K_{\text{noise}}\, p_1\right) \\
&\leq D_{\text{KL}}\left(K_{\text{noise}}\, p_0 \,||\, K_{\text{noise}}\, p_1\right) \quad \text{(by DPI of denoising step)} \\
&\leq \bar{\alpha}_k \, D_{\text{KL}}\left(p_0 \,||\, p_1\right) \quad \text{(by SDPI of noise injection)} \quad \blacksquare
\end{aligned} \tag{9}$$

Analysis. As shown in Eq. 9, the output distribution error becomes smaller with more truncation steps k, with $\bar{\alpha}_k$ decrease with k. This is consistent with the intuition that if the conditioning change is small, we may use fewer diffusion steps, whereas if the conditioning change is large, for example, in our scenario, if the human moves fast between two time steps, we should use more diffusion steps to get better results.

C. Diffusion Model Setup

The diffusion model's architecture design is not the focus of this paper. We choose the transformer-based diffusion model (**Dif**) as [16] for the base diffusion model. The noise predictor ϵ_θ is a transformer network that takes the history trajectory $\mathbf{x}$, the map context S as the conditioning input $\mathbf{c}$, and predicts the noise. The denoising process follows the standard DDPM sampling procedure [15]. In the meantime, note that the noise scheduler will significantly affects the value of $\bar{\alpha}_k$, and hence the results of truncated diffusion. In this paper, we use the linear noise scheduler [15].

V. EXPERIMENTS

We conduct a series of experiments to evaluate the performance of the proposed truncated diffusion process on the human motion prediction task. We mainly investigate two questions: (1) How well does the truncated diffusion process perform on the human motion prediction task, especially compared to the full diffusion process that starts from a Gaussian noise? (2) How does the truncation ratio affect the sampling quality? Does the theoretical analysis in Sec. IV-B hold in practice? In the following section, we first introduce our experiment setup (dataset, metric, baselines), then provide our analysis for both questions in Sec. V-B.

A. Experiment setup

Datasets. In this paper, we select two datasets for the human trajectory prediction task: **GTA-IM** [5] (simulation) and **HPS** [13] (real-world). Both datasets record human motion in a much longer horizon (a few minutes), and in a relatively large area (the whole floor of a building), compared to PROX [14] or Human3.6M [17]. We use 15 frames (5 fps, 3 seconds) of the human 2D trajectory and the local occupancy map as conditioning inputs, and predict the next 15 frames (3 seconds) of the human 2D trajectory. We sample $N = 10$ future trajectories to cover the distribution of future outcomes.

Metric. To evaluate the prediction results, we use Average Displacement Error **(ADE)** and Final Displacement Error **(FDE)** as our two metrics. Additionally, to measure the coverage of multiple predictions on the whole distribution, we use **1-minADE** and **k-minADE** to evaluate the average of 1- and k-minimum average ADE, and also **1-minFDE** and **k-minFDE** to evaluate 1- and k- minimum average FDE, similar to [40]. In our experiment, we set $k = 5$. For all these metrics, lower error means better performance.

Baselines. We select PathNet [19] and STPOTR [27] as two representatives of state-of-the-art one-prediction baselines for human motion prediction. For both baselines, since only one prediction result is predicted, 1-minADE and k-minADE are the same as ADE. So as the FDEs. In the meantime, we implement the full diffusion method (**Dif**) that starts from a Gaussian noise as the third baseline. To compare with another speed-up methods for diffusion models, we select DDIM [37] as our last baseline, which is a widely used method for fast sampling from diffusion models. To fairly compare, they use the same conditioning architecture as our method (**Dif-TR**).

B. Results and Analysis

We show the quantitative results in the table. I, and qualitative results in Fig. 2. Both diffusion methods use $K = 100$ as the full diffusion steps. The truncated ratio is provided by **Dif-TR**(k) where $k = 0.05$ indicates that only 5-steps are used. From Table. I, we show that the full-process diffusion method (**Dif**) can already outperform single-prediction baselines [19, 27] consistently on 1-minADE, k-minADE, 1-minFDE, and k-minFDE on both datasets, indicating that the

TABLE I: **Human Trajectory Prediction**

GTA-IM (traj)										
Metric \ Method	PathNet [19]	STPOTR [27]	Dif [16]	Dif-TR (0.75)	Dif-TR (0.50)	Dif-TR (0.20)	Dif-TR (0.05)	DDIM (0.50)	DDIM (0.20)	DDIM (0.05)
1-ADE	0.052	0.185	**0.031**	0.032	0.034	0.041	0.058	0.042	0.039	0.036
k-ADE	0.052	0.185	**0.042**	0.043	0.045	0.054	0.072	0.060	0.056	0.049
ADE	0.052	0.185	**0.059**	0.060	0.063	0.072	0.091	0.088	0.082	0.072
1-FDE	0.087	0.384	**0.031**	0.032	0.036	0.051	0.085	0.058	0.054	0.044
k-FDE	0.087	0.384	**0.058**	0.061	0.065	0.082	0.119	0.097	0.091	0.076
FDE	0.087	0.384	**0.099**	0.101	0.106	0.125	0.162	0.161	0.151	0.133
HPS										
Metric \ Method	PathNet [19]	STPOTR [27]	Dif [16]	Dif-TR (0.75)	Dif-TR (0.50)	Dif-TR (0.20)	Dif-TR (0.05)	DDIM (0.50)	DDIM (0.20)	DDIM (0.05)
1-ADE	0.051	0.238	**0.037**	0.039	0.041	0.048	0.070	0.066	0.063	0.057
k-ADE	**0.051**	0.238	**0.051**	0.053	0.056	0.064	0.088	0.090	0.087	0.078
ADE	**0.051**	0.238	0.071	0.075	0.080	0.089	0.113	0.129	0.124	0.114
1-FDE	0.105	0.446	**0.039**	0.043	0.047	0.061	0.104	0.104	0.101	0.092
k-FDE	0.105	0.446	**0.072**	0.078	0.083	0.099	0.145	0.156	0.151	0.138
FDE	**0.105**	0.446	0.119	0.128	0.137	0.156	0.200	0.236	0.228	0.212

diffusion model can better cover the whole distribution of human future motion. Furthermore, as we decrease the truncation ratio from 1.00 to 0.05, the sampling quality decreases, which is consistent with our theoretical analysis in Sec. IV-B.

The comparison between the truncated diffusion method (**Dif-TR**) and the DDIM baseline is interesting since we find DDIM performs the best with the smallest number of sampling steps (5 steps). From the table, **Dif-TR** outperforms the best of DDIM with the truncated ratio of 0.50 for GTA-IM and 0.20 for HPS. Therefore, a trade-off can be selected upon application requirement between the sampling quality and the number of sampling steps, i.e. between the truncated diffusion method and DDIM.

To further analyze the effect of different truncation ratios on the sampling quality, we vary the truncation ratio from 0.05 to 1.0 (full diffusion) with step size 0.05, and report the sampling quality (1-minADE) in Fig. 3. Additionally, we plot the value of $\bar{\alpha}_k$ and the corresponding Signal-to-Noise Ratio SNR $= \frac{\bar{\alpha}_k}{1-\bar{\alpha}_k}$ during the denoising process in Fig. 3. As we increase the denoising ratio k from 1.0 to 0.05, the 1-minADE decreases with the SNR, indicating a strong correlation between sampling quality and SNR.

VI. CONCLUSIONS

In this paper, we presented a Truncated Diffusion Process that can be used for consequential inference problems such as human motion prediction. We formulated the consequential human motion prediction problem as a sampling-as-inference method, and proposed the Truncated Diffusion Process with mainly two stages: (1) a forward noise injection process to inject noise into existing samples from the previous time step to an intermediate noise level; (2) a reverse denoising process to obtain the samples under the new conditioning. We provided a theoretical analysis of the proposed method's performance. We showed that the KL divergence between the truncated diffusion process and the full diffusion process is upper bounded by the KL divergence between the sampling distribution under different conditionings from pure Gaussian noise, multiplied by a factor related to the truncation ratio. We conducted experiments on both simulation and real-world datasets to evaluate the sampling quality of the proposed method. We quantitatively demonstrated that the truncated diffusion process can save significant sampling steps while still maintaining good sampling quality. Through experiments, we validated our theoretical analysis and demonstrated that sampling quality is strongly correlated with the Signal-to-Noise Ratio (SNR) during denoising.

REFERENCES

[1] Hyemin Ahn, Esteve Valls Mascaro, and Dongheui Lee. Can We Use Diffusion Probabilistic Models for 3D Motion Prediction?, February 2023. URL http://arxiv.org/abs/2302.14503. arXiv:2302.14503 [cs].

[2] Emre Aksan, Manuel Kaufmann, Peng Cao, and Otmar Hilliges. A Spatio-temporal Transformer for 3D Human Motion Prediction. In *2021 International Conference on 3D Vision (3DV)*, pages 565–574, December 2021. doi: 10.1109/3DV53792.2021.00066.

[3] Arij Bouazizi, Adrian Holzbock, Ulrich Kressel, Klaus Dietmayer, and Vasileios Belagiannis. MotionMixer: MLP-based 3D Human Body Pose Forecasting. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence*, pages 791–798, Vienna, Austria, July 2022. International Joint Conferences on Artificial Intelligence Organization. ISBN 978-1-956792-00-3. doi: 10.24963/ijcai.2022/111. URL https://www.ijcai.org/proceedings/2022/111.

[4] Leo Bringer, Joey Wilson, Kira Barton, and Maani Ghaffari. MDMP: Multi-modal Diffusion for supervised

Motion Predictions with uncertainty, June 2025. URL http://arxiv.org/abs/2410.03860. arXiv:2410.03860 [cs].

[5] Zhe Cao, Hang Gao, Karttikeya Mangalam, Qi-Zhi Cai, Minh Vo, and Jitendra Malik. Long-term Human Motion Prediction with Scene Context, July 2020. URL http://arxiv.org/abs/2007.03672. arXiv:2007.03672 [cs].

[6] Joao Carvalho, An T. Le, Mark Baierl, Dorothea Koert, and Jan Peters. Motion Planning Diffusion: Learning and Planning of Robot Motions with Diffusion Models, August 2023. URL http://arxiv.org/abs/2308.01557. arXiv:2308.01557 [cs].

[7] Lujing Chen, Rui Liu, Xin Yang, Dongsheng Zhou, Qiang Zhang, and Xiaopeng Wei. STTG-net: a Spatio-temporal network for human motion prediction based on transformer and graph convolution network. *Visual Computing for Industry, Biomedicine, and Art*, 5(1): 19, December 2022. ISSN 2524-4442. doi: 10.1186/s42492-022-00112-5. URL https://vciba.springeropen.com/articles/10.1186/s42492-022-00112-5.

[8] Cheng Chi, Siyuan Feng, Yilun Du, Zhenjia Xu, Eric Cousineau, Benjamin Burchfiel, and Shuran Song. Diffusion Policy: Visuomotor Policy Learning via Action Diffusion, June 2023. URL http://arxiv.org/abs/2303.04137. arXiv:2303.04137 [cs].

[9] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, July 2006. ISBN 0-471-24195-4. Published: Hardcover.

[10] Giannis Daras, Yeshwanth Cherapanamjeri, and Constantinos Daskalakis. How much is a noisy image worth? Data Scaling Laws for Ambient Diffusion, November 2024. URL http://arxiv.org/abs/2411.02780. arXiv:2411.02780 [cs].

[11] Prafulla Dhariwal and Alex Nichol. Diffusion Models Beat GANs on Image Synthesis, June 2021. URL http://arxiv.org/abs/2105.05233. arXiv:2105.05233 [cs].

[12] Kevin Frans, Danijar Hafner, Sergey Levine, and Pieter Abbeel. One Step Diffusion via Shortcut Models, October 2024. URL http://arxiv.org/abs/2410.12557. arXiv:2410.12557.

[13] Vladimir Guzov, Aymen Mir, Torsten Sattler, and Gerard Pons-Moll. Human POSEitioning System (HPS): 3D Human Pose Estimation and Self-localization in Large Scenes from Body-Mounted Sensors. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2021.

[14] Mohamed Hassan, Vasileios Choutas, Dimitrios Tzionas, and Michael J. Black. Resolving 3D Human Pose Ambiguities with 3D Scene Constraints, August 2019. URL http://arxiv.org/abs/1908.06963. arXiv:1908.06963 [cs].

[15] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising Diffusion Probabilistic Models. In *Advances in Neural Information Processing Systems*, volume 33, pages 6840–6851. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/4c5bcfec8584af0d967f1ab10179ca4b-Abstract.html.

[16] Siyuan Huang, Zan Wang, Puhao Li, Baoxiong Jia, Tengyu Liu, Yixin Zhu, Wei Liang, and Song-Chun Zhu. Diffusion-based generation, optimization, and planning in 3d scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16750–16761, 2023.

[17] Catalin Ionescu, Dragos Papava, Vlad Olaru, and Cristian Sminchisescu. Human3.6M: Large Scale Datasets and Predictive Methods for 3D Human Sensing in Natural Environments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 36(7):1325–1339, July 2014.

[18] Michael Janner, Yilun Du, Joshua B. Tenenbaum, and Sergey Levine. Planning with Diffusion for Flexible Behavior Synthesis, December 2022. URL http://arxiv.org/abs/2205.09991. arXiv:2205.09991 [cs].

[19] Qingyuan Jiang, Burak Susam, Jun-Jee Chao, and Volkan Isler. Map-Aware Human Pose Prediction for Robot Follow-Ahead, March 2024. URL http://arxiv.org/abs/2403.13294. arXiv:2403.13294 [cs].

[20] Lifu Jin, Amedeo Roberto Esposito, and Michael Gastpar. Properties of the Strong Data Processing Constant for Rényi Divergence, May 2024. URL http://arxiv.org/abs/2403.10656. arXiv:2403.10656 [cs].

[21] Nikhil Karnad and Volkan Isler. Modeling human motion patterns for multi-robot planning. In *2012 IEEE International Conference on Robotics and Automation*, pages 3161–3166. IEEE, 2012.

[22] Sergey Levine. Reinforcement Learning and Control as Probabilistic Inference: Tutorial and Review, May 2018. URL http://arxiv.org/abs/1805.00909. arXiv:1805.00909 [cs, stat].

[23] Jiacheng Liu, Chang Zou, Yuanhuiyi Lyu, Junjie Chen, and Linfeng Zhang. From Reusing to Forecasting: Accelerating Diffusion Models with TaylorSeers, August 2025. URL http://arxiv.org/abs/2503.06923. arXiv:2503.06923 [cs].

[24] Thomas Lucas, Fabien Baradel, Philippe Weinzaepfel, and Grégory Rogez. PoseGPT: Quantization-based 3D Human Motion Generation and Forecasting, October 2022. URL http://arxiv.org/abs/2210.10542. arXiv:2210.10542 [cs].

[25] Andreas Lugmayr, Martin Danelljan, Andres Romero, Fisher Yu, Radu Timofte, and Luc Van Gool. RePaint: Inpainting using Denoising Diffusion Probabilistic Models, August 2022. URL http://arxiv.org/abs/2201.09865. arXiv:2201.09865 [cs].

[26] Kedi Lyu, Haipeng Chen, Zhenguang Liu, Beiqi Zhang, and Ruili Wang. 3D Human Motion Prediction: A Survey, March 2022. URL http://arxiv.org/abs/2203.01593. arXiv:2203.01593 [cs].

[27] Mohammad Mahdavian, Payam Nikdel, Mahdi Taher-Ahmadi, and Mo Chen. STPOTR: Simultaneous Human Trajectory and Pose Prediction Using a Non-Autoregressive Transformer for Robot Following Ahead, September 2022. URL http://arxiv.org/abs/2209.07600.

arXiv:2209.07600 [cs].

[28] Wei Mao, Miaomiao Liu, Mathieu Salzmann, and Hongdong Li. Learning Trajectory Dependencies for Human Motion Prediction. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 9488–9496, Seoul, Korea (South), October 2019. IEEE. ISBN 978-1-7281-4803-8. doi: 10.1109/ICCV.2019.00958. URL https://ieeexplore.ieee.org/document/9009559/.

[29] Angel Martinez-Gonzalez, Michael Villamizar, and Jean-Marc Odobez. Pose Transformers (POTR): Human Motion Prediction with Non-Autoregressive Transformers. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pages 2276–2284, Montreal, BC, Canada, October 2021. IEEE. ISBN 978-1-6654-0191-3. doi: 10.1109/ICCVW54120.2021.00257. URL https://ieeexplore.ieee.org/document/9607511/.

[30] Chenlin Meng, Yutong He, Yang Song, Jiaming Song, Jiajun Wu, Jun-Yan Zhu, and Stefano Ermon. SDEdit: Guided Image Synthesis and Editing with Stochastic Differential Equations, January 2022. URL http://arxiv.org/abs/2108.01073. arXiv:2108.01073 [cs].

[31] Ronja Möller, Antonino Furnari, Sebastiano Battiato, Aki Härmä, and Giovanni Maria Farinella. A Survey on Human-aware Robot Navigation, June 2021. URL http://arxiv.org/abs/2106.11650. arXiv:2106.11650 [cs] version: 1.

[32] Payam Nikdel, Mohammad Mahdavian, and Mo Chen. DMMGAN: Diverse Multi Motion Prediction of 3D Human Joints using Attention-Based Generative Adversarial Network. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 9938–9944, May 2023. doi: 10.1109/ICRA48891.2023.10160401.

[33] Yury Polyanskiy and Yihong Wu. Strong Data-Processing Inequalities for Channels and Bayesian Networks. pages 211–249. Springer Nature, April 2017.

[34] Jonas Scholz and Richard E. Turner. Warm Starts Accelerate Conditional Diffusion, September 2025. URL http://arxiv.org/abs/2507.09212. arXiv:2507.09212 [cs].

[35] Jascha Sohl-Dickstein, Eric A. Weiss, Niru Maheswaranathan, and Surya Ganguli. Deep Unsupervised Learning using Nonequilibrium Thermodynamics, November 2015. URL http://arxiv.org/abs/1503.03585. arXiv:1503.03585 [cond-mat, q-bio, stat].

[36] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning Structured Output Representation using Deep Conditional Generative Models. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://papers.nips.cc/paper_files/paper/2015/hash/8d55a249e6baa5c06772297520da2051-Abstract.html.

[37] Jiaming Song, Chenlin Meng, and Stefano Ermon. Denoising Diffusion Implicit Models, October 2022. URL http://arxiv.org/abs/2010.02502. arXiv:2010.02502 [cs].

[38] Yang Song, Jascha Sohl-Dickstein, Diederik P. Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-Based Generative Modeling through Stochastic Differential Equations, February 2021. URL http://arxiv.org/abs/2011.13456. arXiv:2011.13456 [cs].

[39] Yang Song, Prafulla Dhariwal, Mark Chen, and Ilya Sutskever. Consistency Models, May 2023. URL http://arxiv.org/abs/2303.01469. arXiv:2303.01469 [cs].

[40] Weizhuo Wang, C. Karen Liu, and Monroe Kennedy III. EgoNav: Egocentric Scene-aware Human Trajectory Prediction, August 2024. URL http://arxiv.org/abs/2403.19026. arXiv:2403.19026.

[41] Brian Yang, Huangyuan Su, Nikolaos Gkanatsios, Tsung-Wei Ke, Ayush Jain, Jeff Schneider, and Katerina Fragkiadaki. Diffusion-ES: Gradient-free Planning with Diffusion for Autonomous Driving and Zero-Shot Instruction Following, July 2024. URL http://arxiv.org/abs/2402.06559. arXiv:2402.06559 [cs].

[42] Huangjie Zheng, Pengcheng He, Weizhu Chen, and Mingyuan Zhou. Truncated Diffusion Probabilistic Models and Diffusion-based Adversarial Auto-Encoders, September 2023. URL http://arxiv.org/abs/2202.09671. arXiv:2202.09671 [stat].

VII. APPENDIX

We provide extra details on the theoretical analysis of the truncated diffusion process in Sec. IV-B, including the definition of the Markov kernel for the truncated diffusion process, and the proof of the KL contractivity.

A. Definition details

1) Markov Kernel Definition: In Sec. IV-B, we define the k-step truncated diffusion process as a Markov kernel $K_{\mathbf{c}}^{(k)}$. Here we provide the detailed math process. Given a sample $\mathbf{x}$ and a new condition $\mathbf{c}$, the probability of generating a new sample $\mathbf{x}'$ can be calculated by:

$$\begin{aligned} K_{\mathbf{c}}^{(k)} &:= p_\theta(\mathbf{x}'|\mathbf{x},\mathbf{c}) \\ &= \int_{\mathbf{x}^k} p_\theta(\mathbf{x}',\mathbf{x}^k|\mathbf{x},\mathbf{c})\,\mathrm{d}\mathbf{x}^k \\ &= \int_{\mathbf{x}^k} p_\theta(\mathbf{x}'|\mathbf{x},\mathbf{x}^k,\mathbf{c})p(\mathbf{x}^k|\mathbf{x},\mathbf{c})\,\mathrm{d}\mathbf{x}^k \\ &= \int_{\mathbf{x}^k} p_\theta(\mathbf{x}'|\mathbf{x}^k,\mathbf{c})q(\mathbf{x}^k|\mathbf{x})\,\mathrm{d}\mathbf{x}^k \end{aligned} \tag{10}$$

In the forth row, we remove the dependency of $\mathbf{x}$ in $p_\theta(\mathbf{x}'|\mathbf{x},\mathbf{x}^k,\mathbf{c})$ because the denoising process only depends on the noised sample $\mathbf{x}^k$ and the condition $\mathbf{c}$. Similarly, we remove the dependency of $\mathbf{c}$ in $p(\mathbf{x}^k|\mathbf{x},\mathbf{c})$ because the forward diffusion process is independent of the condition, and we replace the notation with $q(\mathbf{x}^k|\mathbf{x})$ to follow the convention of diffusion models [36].

2) Output Distribution Definition: Given the Markov kernel defined above, we can calculate the output distribution from the truncated diffusion process by applying this kernel to the existing arbitrary distribution $P(\mathbf{x})$.

$$\begin{aligned} Q_\theta(\mathbf{x}'|\mathbf{c}) &:= \int_{\mathbf{x}} p(\mathbf{x}',\mathbf{x}|\mathbf{c})\,\mathrm{d}\mathbf{x} \quad \text{(marginalization)} \\ &= \int_{\mathbf{x}} p(\mathbf{x}'|\mathbf{x},\mathbf{c})p(\mathbf{x}|\mathbf{c})\,\mathrm{d}\mathbf{x} \quad \text{(chain rule)} \\ &= \int_{\mathbf{x}} p(\mathbf{x}'|\mathbf{x},\mathbf{c})P(\mathbf{x})\,\mathrm{d}\mathbf{x} \quad \text{(Definition of } P(\mathbf{x})\text{)} \end{aligned} \tag{11}$$

Note that in the third row, we replace the notation of $p(\mathbf{x}|\mathbf{c})$ with $P(\mathbf{x})$ because the input distribution is independent of the condition $\mathbf{c}$, and we use $P(\mathbf{x})$ to follow the convention of probability distributions.

3) Identity Transition: Here we prove the identity transition property of the Markov kernel is an identity mapping when sampling under the same condition, i.e., $K_{\mathbf{c}}^{(k)} \circ p(\mathbf{x}|\mathbf{c}) = p(\mathbf{x}|\mathbf{c})$.

$$\begin{aligned} K_{\mathbf{c}}^{(k)} \circ p(\mathbf{x}|\mathbf{c}) &:= \int_{\mathbf{x}} p(\mathbf{x}'|\mathbf{x},\mathbf{c})p(\mathbf{x}|\mathbf{c})\,\mathrm{d}\mathbf{x} \\ &= \int_{\mathbf{x}} p(\mathbf{x}',\mathbf{x}|\mathbf{c})\,\mathrm{d}\mathbf{x} \quad \text{(chain rule)} \\ &= p(\mathbf{x}'|\mathbf{c}) \quad \text{(marginalization)} \end{aligned} \tag{12}$$

B. KL Contractivity Proof

In this section, we provide a fully rigourous and unconditional proof of the KL contractivity.

Lemma 1 (Data Processing Inequality for KL Divergence [9]): Let $K(x_t|x_0)$ be any Markov kernel, then for any distributions $p(x)$ and $q(x)$, we have:

$$D_{\mathrm{KL}}\left(Kp \,||\, Kq\right) \le D_{\mathrm{KL}}\left(p \,||\, q\right) \tag{13}$$

Proof:

Define the push-forward distribution through the kernel on x_t as, $\tilde{p}(x_t) = \int K(x_t|x_0)p(x_0)\,\mathrm{d}x_0$, and similarly $\tilde{q}(x_t) = \int K(x_t|x_0)q(x_0)\,\mathrm{d}x_0$. we introduce the joint distributions $P(x_0,x_t) = p(x_0)K(x_t|x_0)$ and $Q(x_0,x_t) = q(x_0)K(x_t|x_0)$.

$$\begin{aligned} P(x_0,x_t) &= p(x_0)K(x_t|x_0) = \tilde{p}(x_t)p(x_0|x_t) \\ Q(x_0,x_t) &= q(x_0)K(x_t|x_0) = \tilde{q}(x_t)q(x_0|x_t) \end{aligned} \tag{14}$$

The KL divergence between the joint distributions is:

$$\begin{aligned} &D_{\mathrm{KL}}\left(P(x_0,x_t) \,||\, Q(x_0,x_t)\right) \\ &= \int_{x_0}\int_{x_t} P(x_0,x_t)\log\frac{P(x_0,x_t)}{Q(x_0,x_t)}\,\mathrm{d}x_t\,\mathrm{d}x_0 \\ &= \int_{x_0}\int_{x_t} P(x_0,x_t)\log\frac{\tilde{p}(x_t)p(x_0|x_t)}{\tilde{q}(x_t)q(x_0|x_t)}\,\mathrm{d}x_t\,\mathrm{d}x_0 \\ &= \int_{x_0}\int_{x_t} P(x_0,x_t)\log\frac{\tilde{p}(x_t)}{\tilde{q}(x_t)}\,\mathrm{d}x_t\,\mathrm{d}x_0 \quad + \\ &\quad \int_{x_0}\int_{x_t} P(x_0,x_t)\log\frac{p(x_0|x_t)}{q(x_0|x_t)}\,\mathrm{d}x_t\,\mathrm{d}x_0 \\ &= T_1 + T_2 \end{aligned} \tag{15}$$

$$\begin{aligned} T_1 &= \int_{x_0}\int_{x_t} P(x_0,x_t)\log\frac{\tilde{p}(x_t)}{\tilde{q}(x_t)}\,\mathrm{d}x_t\,\mathrm{d}x_0 \\ &= \int_{x_t}\left(\int_{x_0} P(x_0,x_t)\,\mathrm{d}x_0\right)\log\frac{\tilde{p}(x_t)}{\tilde{q}(x_t)}\,\mathrm{d}x_t \\ &= \int_{x_t}\tilde{p}(x_t)\log\frac{\tilde{p}(x_t)}{\tilde{q}(x_t)}\,\mathrm{d}x_t \\ &= D_{\mathrm{KL}}\left(\tilde{p} \,||\, \tilde{q}\right) \end{aligned} \tag{16}$$

$$\begin{aligned} T_2 &= \int_{x_0}\int_{x_t} P(x_0,x_t)\log\frac{p(x_0|x_t)}{q(x_0|x_t)}\,\mathrm{d}x_t\,\mathrm{d}x_0 \\ &= \int_{x_t}\left(\int_{x_0} P(x_0,x_t)\log\frac{p(x_0|x_t)}{q(x_0|x_t)}\,\mathrm{d}x_0\right)\,\mathrm{d}x_t \\ &= \int_{x_t}\left(\int_{x_0} \tilde{p}(x_t)p(x_0|x_t)\log\frac{p(x_0|x_t)}{q(x_0|x_t)}\,\mathrm{d}x_0\right)\,\mathrm{d}x_t \\ &= \int_{x_t}\tilde{p}(x_t)\left(\int_{x_0} p(x_0|x_t)\log\frac{p(x_0|x_t)}{q(x_0|x_t)}\,\mathrm{d}x_0\right)\,\mathrm{d}x_t \\ &= \int_{x_t}\tilde{p}(x_t)D_{\mathrm{KL}}\left(p(x_0|x_t) \,||\, q(x_0|x_t)\right)\,\mathrm{d}x_t \\ &= \mathbb{E}_{x_t\sim\tilde{p}}\left[D_{\mathrm{KL}}\left(p(x_0|x_t) \,||\, q(x_0|x_t)\right)\right] \end{aligned} \tag{17}$$

Combining the two terms, we have:

$$
\begin{aligned}
&D_{\mathrm{KL}}\left(P(x_0, x_t) \,||\, Q(x_0, x_t)\right) = \\
&\quad D_{\mathrm{KL}}\left(\tilde{p} \,||\, \tilde{q}\right) + \mathbb{E}_{x_t \sim \tilde{p}}\left[D_{\mathrm{KL}}\left(p(x_0|x_t) \,||\, q(x_0|x_t)\right)\right]
\end{aligned} \tag{18}
$$

Since the KL divergence term $D_{\mathrm{KL}}\left(p(x_0|x_t) \,||\, q(x_0|x_t)\right)$ is non-negative, its expectation is also non-negative, i.e. $\mathbb{E}_{x_t \sim \tilde{p}}\left[D_{\mathrm{KL}}\left(p(x_0|x_t) \,||\, q(x_0|x_t)\right)\right] \geq 0$, we have:

$$
D_{\mathrm{KL}}\left(\tilde{p} \,||\, \tilde{q}\right) \leq D_{\mathrm{KL}}\left(P(x_0, x_t) \,||\, Q(x_0, x_t)\right) \tag{19}
$$

Note that by definition of the joint distributions, we have:

$$
\begin{aligned}
&D_{\mathrm{KL}}\left(P(x_0, x_t) \,||\, Q(x_0, x_t)\right) \\
&= \int_{x_0} \int_{x_t} p(x_0) K(x_t|x_0) \log \frac{p(x_0) K(x_t|x_0)}{q(x_0) K(x_t|x_0)} \,\mathrm{d}x_t \,\mathrm{d}x_0 \\
&= \int_{x_0} \int_{x_t} p(x_0) K(x_t|x_0) \log \frac{p(x_0)}{q(x_0)} \,\mathrm{d}x_t \,\mathrm{d}x_0 \\
&= \int_{x_0} \left(\int_{x_t} K(x_t|x_0) \,\mathrm{d}x_t \right) p(x_0) \log \frac{p(x_0)}{q(x_0)} \,\mathrm{d}x_0 \\
&= \int_{x_0} p(x_0) \log \frac{p(x_0)}{q(x_0)} \,\mathrm{d}x_0 \\
&= D_{\mathrm{KL}}\left(p \,||\, q\right)
\end{aligned} \tag{20}
$$

Therefore, we have proved the lemma:

$$
D_{\mathrm{KL}}\left(Kp \,||\, Kq\right) \leq D_{\mathrm{KL}}\left(p \,||\, q\right) \tag{21}
$$

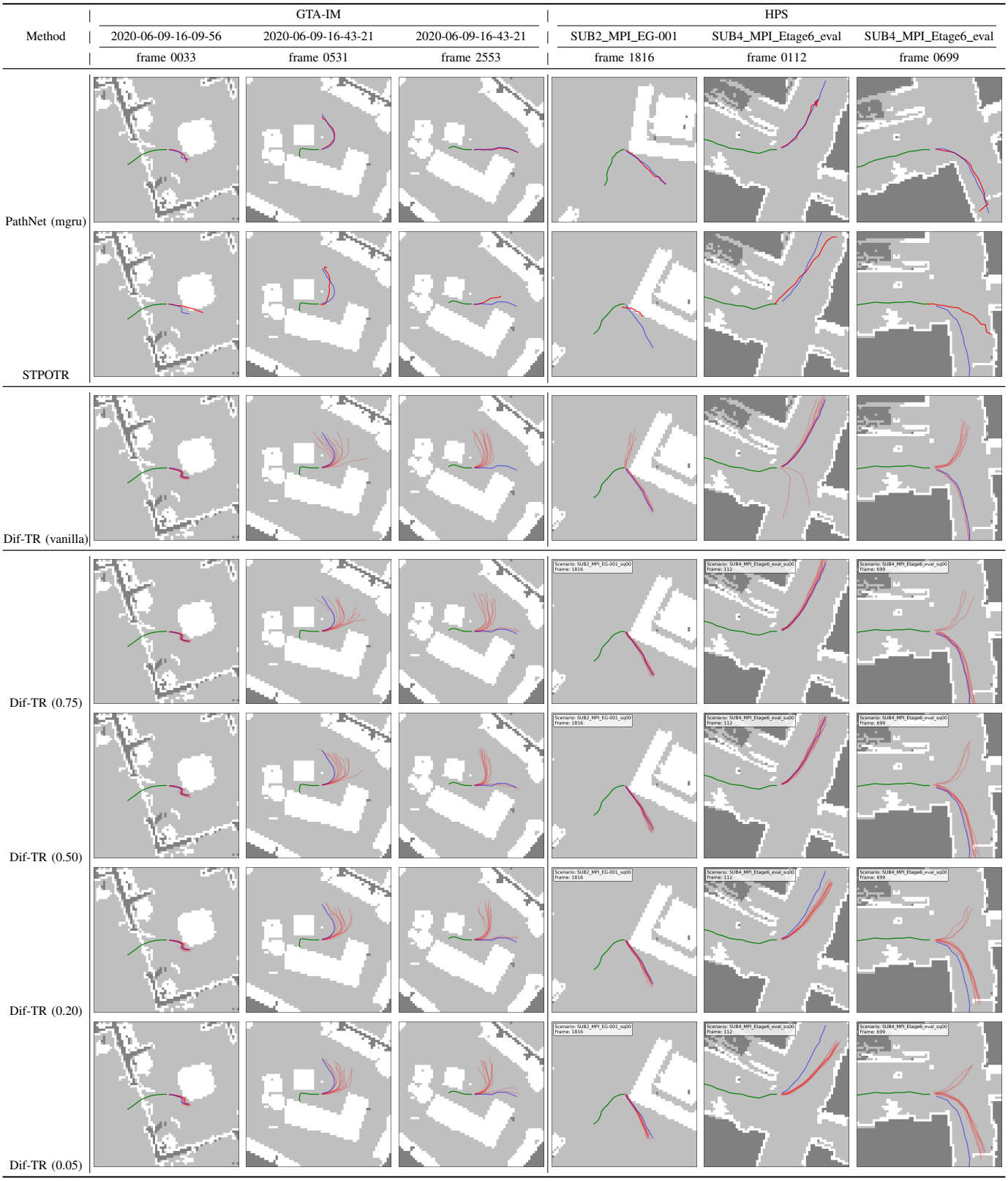
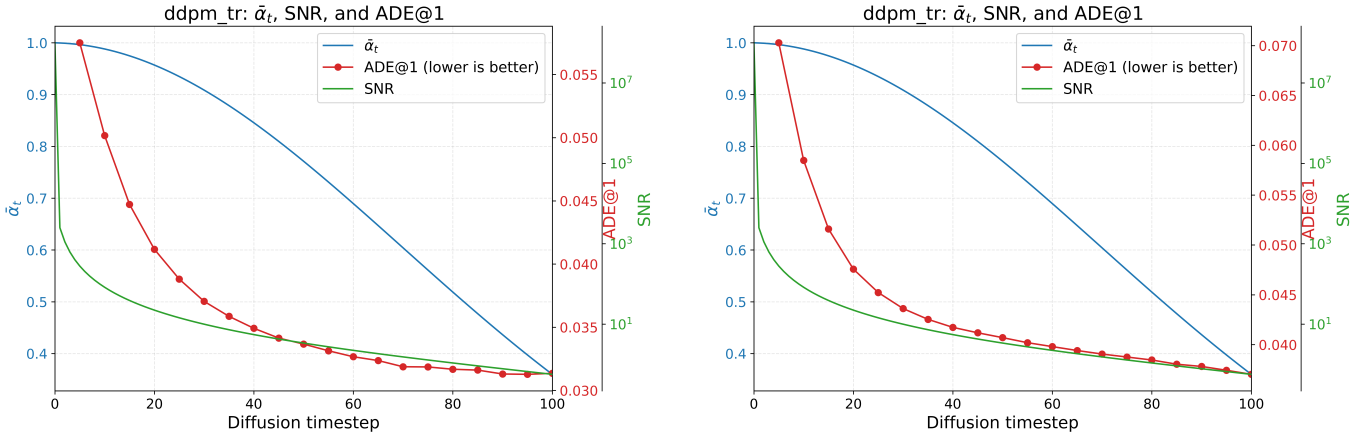


Fig. 2: **Qualitative results for human motion prediction**. We plot the human motion history in green, the predicted human motion in red, and the ground truth in blue. The local occupancy map is visualized by: white - obstacles, grey - free space, dark - unknown. Each row is a method, and each column is a selected sample.



(a) GTA-IM

(b) HPS

Fig. 3: **Sampling quality change along the truncated diffusion steps**. We plot the changes in $\bar{\alpha}_t$, 1-minADE, and SNR with respect to the number of truncated diffusion steps and show their correlation.