MDM: Interacting With Obstacles

Yoav Shkedy, Eviatar Shemesh, Noam Azoulay

June 20, 2024

1 Introduction

Generating natural and expressive human motion is a substantial challenge in the field of computer animation. This complexity arises from the vast diversity of possible motions, the sensitivity of human perception to inaccuracies in motion, and the inherent difficulty in accurately describing human movements. Current generative solutions often fall short, either producing low-quality outputs or lacking in expressiveness.

Motion Diffusion Models (MDMs) offer a promising approach to addressing these challenges. They utilize the advanced generative abilities of diffusion models, previously shown effective in other areas, making them particularly suitable for human motion tasks. However, these models typically require substantial resources and are challenging to manage.

The MDM framework incorporates a transformer-based architecture and draws on existing knowledge in motion generation. A key feature of MDM is its strategy to predict the sample instead of the noise at each step of diffusion, allowing for the application of specific geometric losses on the motion's position and velocities, like the foot contact loss, thus improving both precision and expressiveness.

MDM's general-purpose nature enables it to handle different conditioning modes and various motion generation tasks. Despite being trained on minimal resources, MDM delivers top-tier performance on major benchmarks for text-to-motion and action-to-motion tasks, showcasing its effectiveness and flexibility in creating lifelike and dynamic human motions.

The focus of this paper is on the application of MDMs in environments involving obstacles. Interacting with obstacles adds an additional layer of complexity to motion generation, as it requires the model to account for spatial constraints and ensure collision-free trajectories. By evaluating joint positions, especially the hip joint relative to obstacles, MDM aims to generate realistic and adaptable human motions that navigate through complex environments effectively.

2 Methodology

Joint Position Analysis

To ensure that the generated motions were collision-free we analyzed the position of the joints, particularly the hip joint, relative to obstacles. The model computed whether the joint was within the bounds of an obstacle and evaluated the proximity to enforce a collision-free trajectory. This analysis was critical for maintaining the realism and safety of the generated motions.

Classifier Guidance

Classifier guidance is a technique used in generative models, as detailed in "Classifier-Free Diffusion Guidance" (Dhariwal & Nichol, 2021), which involves using gradients from a classifier to influence the generation process. This approach balances fidelity and diversity in generated outputs by adjusting the strength of the classifier's influence during generation.
In our implementation, we defined a loss function whose gradient is computed to guide the motion generation. During each denoising step, the model adjusts the generated motion in the direction of this gradient. Specifically, we used the guided score function:

$$\mu^{\hat{\theta}}(x_t|y) = \mu_\theta(x_t|y) + s \cdot \Sigma_\theta(x_t|y)\nabla_{x_t} \log p_\phi(y|x_t)$$

Here, $\mu^{\hat{\theta}}(x_t|y)$ is the guided mean function, $\mu_\theta(x_t|y)$ is the original mean function, $\Sigma_\theta(x_t|y)$ is the variance, and $\nabla_{x_t} \log p_\phi(y|x_t)$ is the gradient of the log-probability of the classifier. The parameter s controls the strength of the guidance. By following the gradient of this loss function, our Motion Diffusion Model (MDM) generated high-fidelity motions that adhered to specified conditions, such as avoiding obstacles, while maintaining diversity and realism in the outputs.

Initial Experiment: Trajectory Conditioning

The initial experiment started with a simple motion prompt: "The person is walking". This foundational action provided a baseline to observe and enhance the MDM's response to environmental variables. By starting with a basic motion, the experiment established a controlled scenario to measure the model's initial performance and adaptability. The first step before integrating obstacles was to experiment with trajectory conditioning. In each step of the diffusion process, the trajectory of the input was masked. Mean Squared Error (MSE) loss was applied between the given trajectory and the desired trajectory, and the gradient with respect to this loss was taken. This approach helped in refining the motion trajectories to adhere to the desired paths. By starting with trajectory conditioning, we established a controlled environment to understand the fundamental workings of the MDM and its response to basic conditioning tasks. After establishing the effectiveness of trajectory conditioning, we proceeded to the next phase: evaluating the MDM's ability to interact with obstacles.

Prompts and Obstacle Configurations

To validate the robustness of MDM, a wide array of complex environments and textual prompts were introduced. This rigorous evaluation strategy tested the model against diverse motion challenges and scenarios. The experiments included:

- Various configurations of obstacles to test the model's adaptability.
- Different textual prompts to observe how well the model could generate appropriate motions in response to varying conditions.

While the model achieved notable successes in various settings, it was also found that adjusting the model's hyperparameters was necessary for optimal performance in certain complex environments. This finding emphasized the importance of hyperparameter tuning to enhance the model's adaptability and effectiveness.

Loss Functions

Three specific loss functions were formulated and tested:

- Zero-Infinity Loss: We initiated our experiments with a loss function that assigned zero loss for motion occurring outside of an obstacle's boundaries and an infinite loss for any segments of motion intersecting with the obstacle.

- Trapezoid Loss: This loss function provided a different approach to penalizing interactions with obstacles, using a trapezoidal function to quantify the penalties based on proximity.

- Inverse Distance Loss: This loss function measured the distance between the joint positions and the obstacles, penalizing closer distances more heavily.

The results of these loss functions were analyzed to determine their effectiveness in generating collision-free trajectories.

Tradeoff Between Motion Quality and Condition Discipline

A significant aspect of the methodology was managing the tradeoff between maintaining the original quality of the generated motion and the extent to which the motion adhered to the conditions set by the environment. By using a classifier guidance method, the model could balance this tradeoff, ensuring high-quality motion that was also realistic and adaptable to obstacle interactions.

Interacting with Multiple Obstacles

After successfully interacting with a single obstacle, we advanced to more complex scenarios involving multiple obstacles. This phase tested the MDM's ability to adapt to various configurations of obstacles and maintain collision-free trajectories in more dynamic environments. Different prompts and obstacle configurations were introduced to validate the robustness of the model. This rigorous evaluation strategy led to successes in some cases, showcasing the model's capacity to adapt and perform across a range of settings and specifications.

3 Results and Discussion

Trajectory Conditioning

The first phase of our research involved trajectory conditioning to establish a foundational understanding of the Motion Diffusion Model (MDM) and its response to basic conditioning tasks. Here are the key results from this initial experiment:

Trajectory Masking and Loss Application:

- In each step of the diffusion process, the trajectory of the input motion was masked.

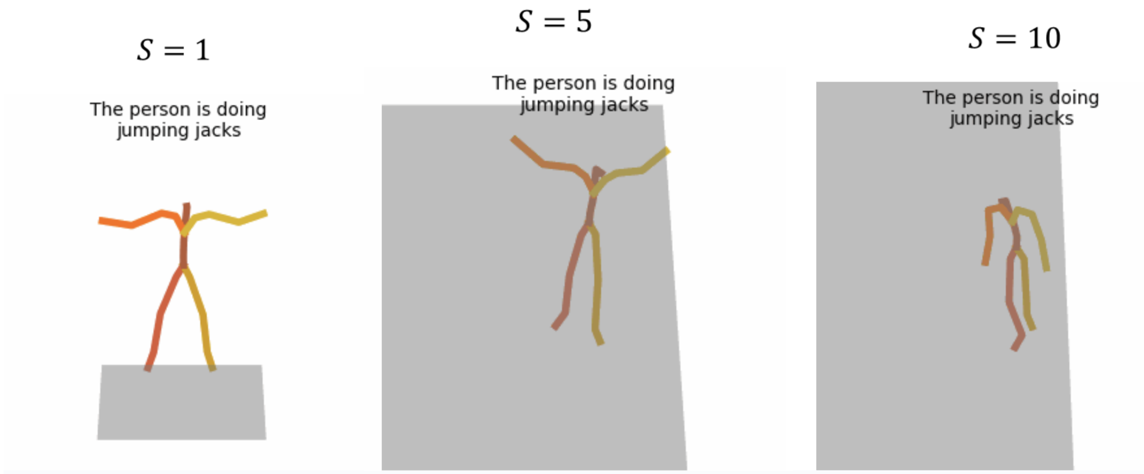
- Mean Squared Error (MSE) loss was applied between the given trajectory and the desired trajectory.

- The gradient with respect to this loss was taken to refine the motion trajectories.

Outcomes of Trajectory Conditioning:

- The application of MSE loss and trajectory masking effectively improved the model's ability to generate motions that closely followed the desired paths.

- This process allowed us to understand how the model reacts to conditioning and provided insights into optimizing trajectory adherence.



Challenges and Observations:

- One of the main observations was the necessity to balance the strength of the trajectory conditioning. Overly strict conditioning could lead to unnatural motions, while too lenient conditioning might result in deviations from the desired path. Soft conditioning ($s = 1$) produced motion based on the provided prompt. Moderately balanced conditioning ($s = 5$) yielded a mix of motions from both the conditioned and prompted motions. Strict conditioning ($s = 10$) caused somewhat unnatural motion, predominantly influenced by the conditioned motion.

- This experiment highlighted the importance of fine-tuning the loss functions and conditioning parameters to achieve optimal results.

These initial results set the stage for the subsequent experiments involving interactions with obstacles. By establishing a controlled environment for trajectory conditioning, we gained valuable insights into the model's capabilities and limitations, which informed our approach to more complex scenarios.

Interacting With a Single Obstacle

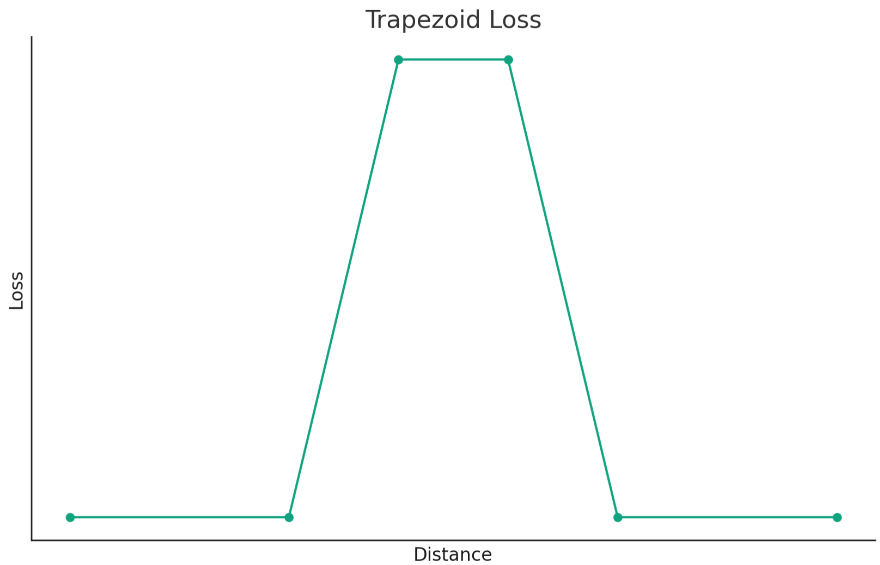
After establishing the effectiveness of trajectory conditioning, we proceeded to the next phase: evaluating the MDM's ability to interact with obstacles.

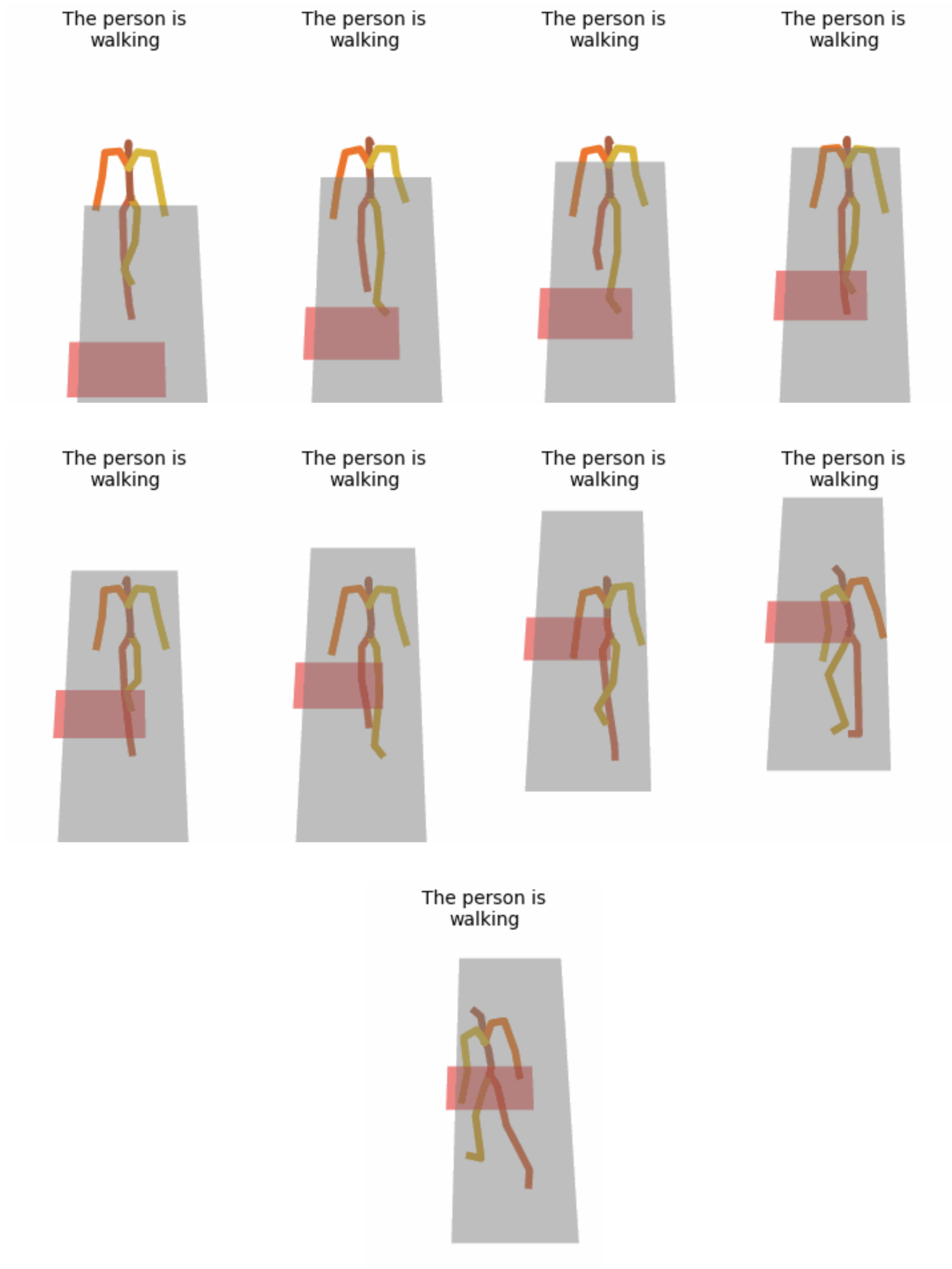
Loss Function Formulation:

We tried to develop a loss function designed specifically to quantify the interaction between the motion model and a single obstacle:

- Zero-Infinity Loss: This binary approach, although conceptually straightforward, presents a critical flaw for learning-based models: it lacks differentiability. The absence of a gradient within the obstacle's space rendered the loss function impractical for gradient-descent-based optimization methods used in neural networks. Consequently, we disregarded this approach and sought to develop a more effective function.

- Trapezoid Loss: To overcome the differentiability challenge, we engineered the trapezoid loss function. This approach embodies a continuous, piecewise-linear loss landscape, allowing for gradient-based optimization. We established a "safe zone" where the model incurs zero loss, promoting trajectories that maintain a specified minimum distance from the obstacle. As the subject approaches the obstacle, the loss increases linearly — mirroring the sides of a trapezoid. Entering the obstacle triggers a high, flat penalty akin to the trapezoid's plateau. Through meticulous experimentation with the trapezoid loss function parameters, we achieved a good configuration that significantly enhanced the model's obstacle avoidance capabilities:

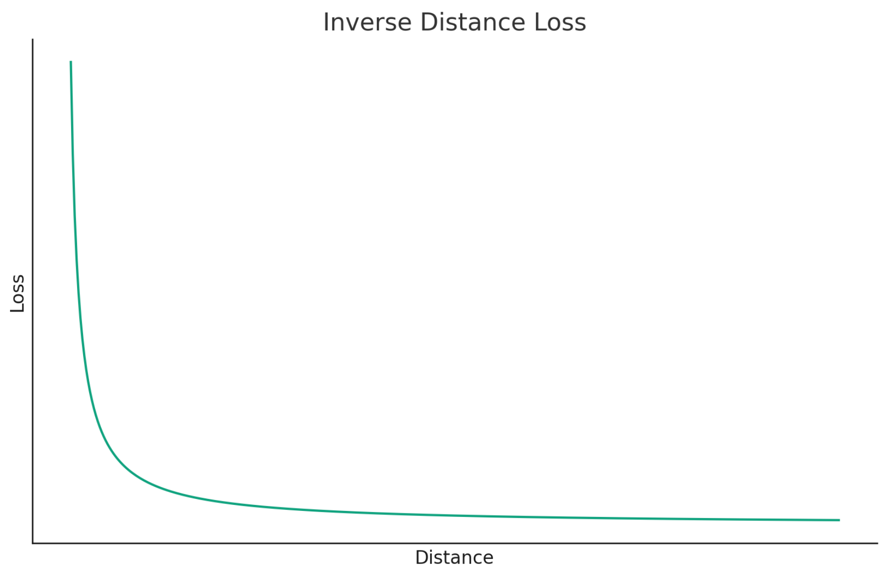


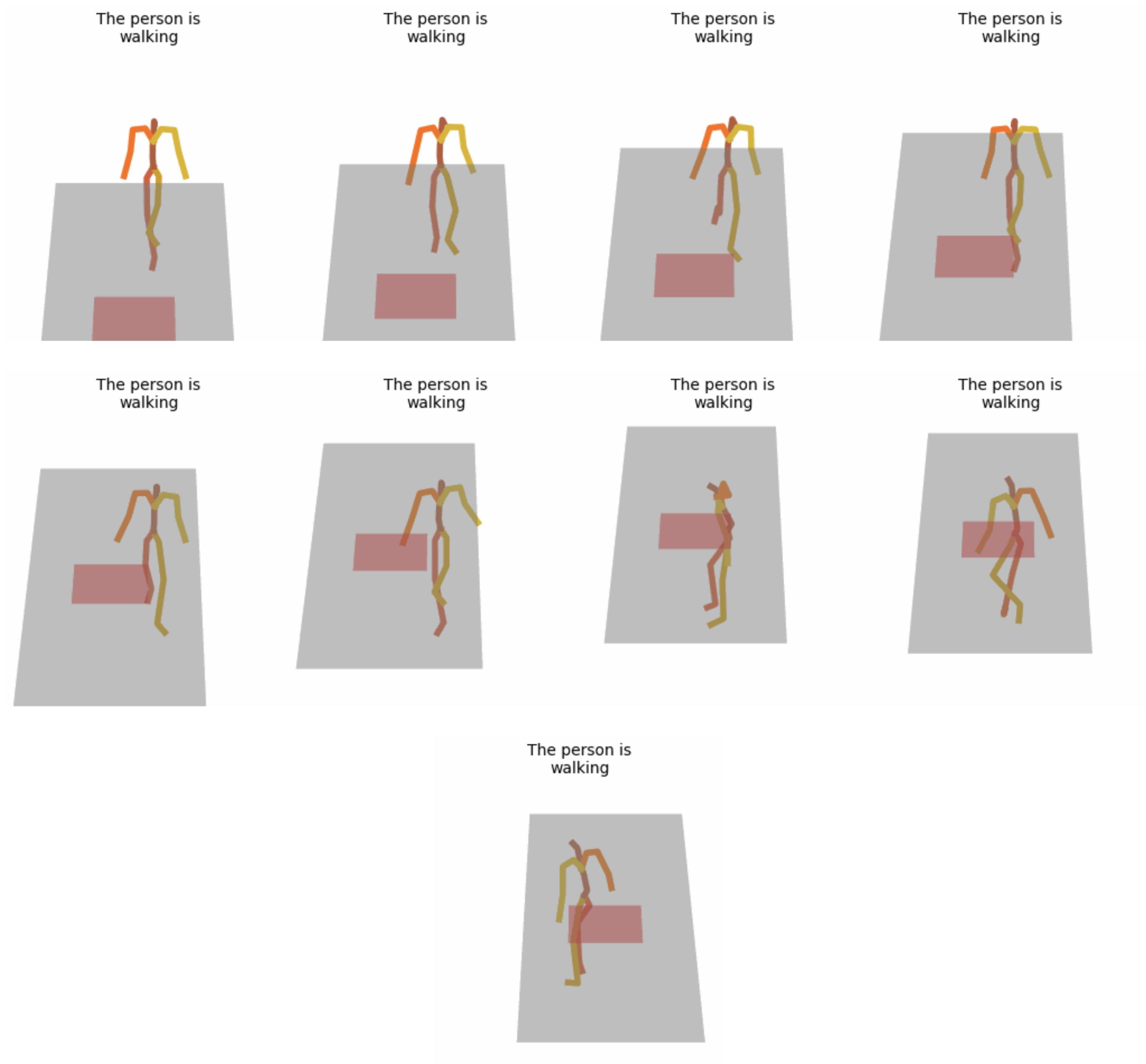


Although the subject kept a distance from the center of the obstacle and attempted to avoid it, the results were still not satisfactory. Therefore, we proceeded to develop a more effective loss function that would ensure total avoidance of the obstacle without the subject stepping in it.

- Inverse Distance Loss: Adopting an inverse distance strategy, our loss function assigned penalties that increased as the model's predicted motion approached the obstacle. To decisively discourage intersection with the obstacle, the function imposes a substantial constant penalty whenever the model's trajectory enters the obstacle's boundaries:

$$\sum_{x \notin \text{square}} \frac{1}{d(x)+\varepsilon} + \sum_{x \in \text{square}} C$$



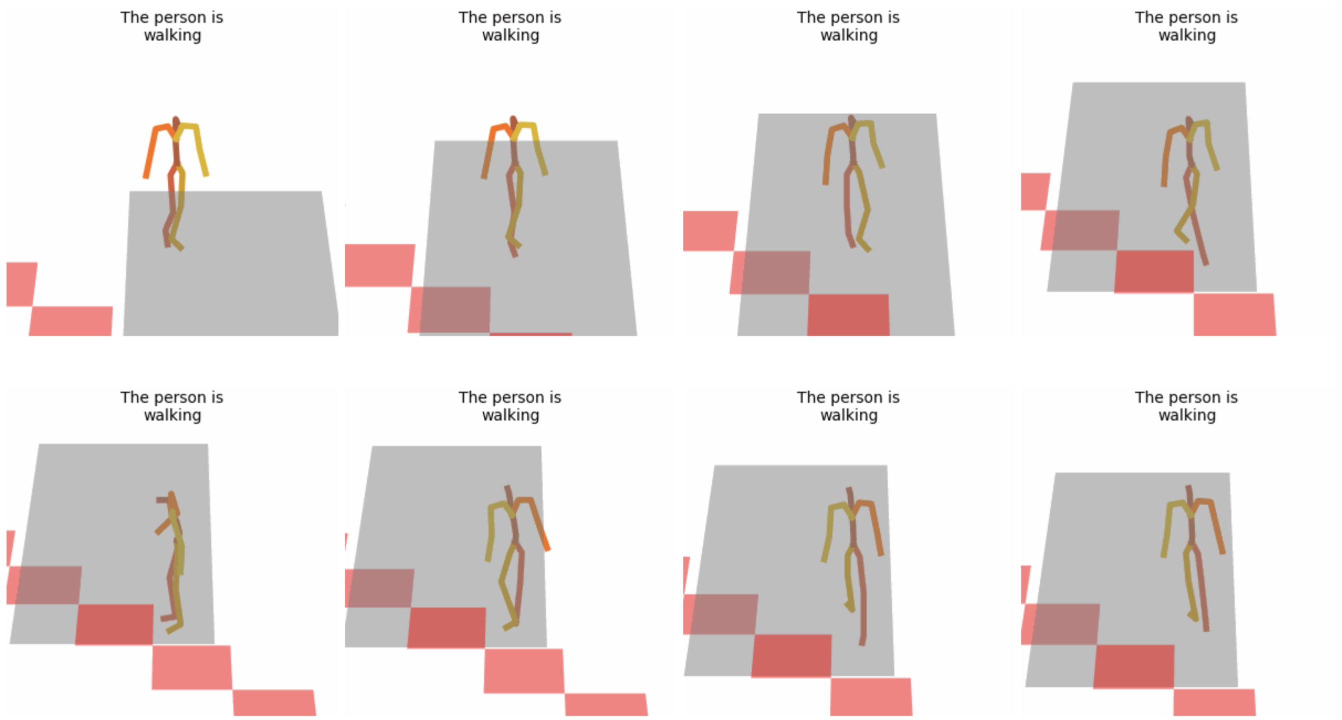


The inverse distance loss formula proved to be highly effective, outperforming previous iterations by delivering more reliable obstacle avoidance while maintaining the fluidity and realism of the motion

Navigating Multiple Obstacles

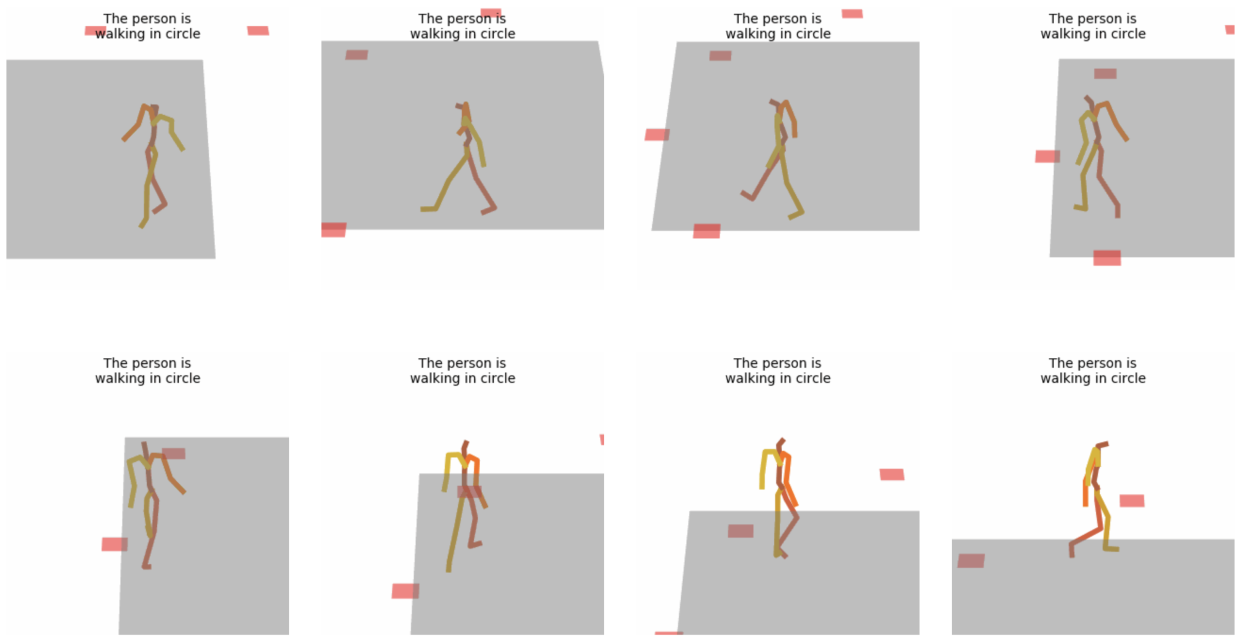
Overcoming Challenges:

The initial transition presented a significant hurdle. The aggregation of high penalties from multiple obstacles resulted in an overwhelming total loss, which complicated the learning process and motion optimization. To tackle the heightened loss levels, we delved into strategic adjustments. Our aim was to balance the penalties to prevent the loss sum from overshadowing the gradient signals necessary for effective learning:



Different Prompts and Obstacle Configurations:

In our quest to validate the robustness of the Motion Diffusion Model, we introduced a wide array of complex environments and textual prompts, aimed at testing the model against diverse motion challenges and scenarios. This rigorous evaluation strategy led to successes in some cases, showcasing the model's capacity to adapt and perform across a range of settings and specifications. While the model achieved notable successes in various settings, our experiments also uncovered a critical limitation: the necessity to adjust the model's hyperparameters for optimal performance in certain complex environments and prompts. This finding highlights a key area for further development, emphasizing the importance of hyperparameter tuning to enhance the model's adaptability and effectiveness in navigating the complexities of dynamic and varied scenarios.



4 Future Work

The findings from our research on Motion Diffusion Models interacting with obstacles have opened several promising avenues for future exploration. To further enhance the capabilities and applications of MDMs, the following areas can be the focus of subsequent studies:

Complex Choreography and Obstacles

One of the primary goals for future work is to simulate more intricate environments. Simple, straightforward motions should progress to modeling sophisticated choreographies that include complex sequences of movements. This will involve introducing a richer variety of obstacle configurations, such as dynamic obstacles that move or change shape over time. By doing so, we will be able to evaluate the model's robustness and adaptability in highly dynamic and unpredictable environments, providing a more comprehensive assessment of its capabilities.

Beyond Flat Surfaces

Current experiments have primarily focused on flat, square obstacles. Future work will venture beyond these limitations to encompass obstacles with diverse shapes, sizes, and dimensions. This includes modeling obstacles that are three-dimensional and have varying heights and textures. The inclusion of such obstacles will simulate more realistic and challenging scenarios, thereby pushing the boundaries of the MDM's ability to generate adaptive and collision-free trajectories.

Data-Driven Obstacle Encounters

To enhance the realism and accuracy of the MDM, leveraging obstacle datasets that include empirical data on obstacle interactions should be considered. Incorporating real-world data into the training regimen will enable the model to learn from actual obstacle encounters, improving its predictive accuracy and performance. This data-driven approach will ensure that the generated motions are not only theoretically sound but also practically applicable in real-world scenarios.