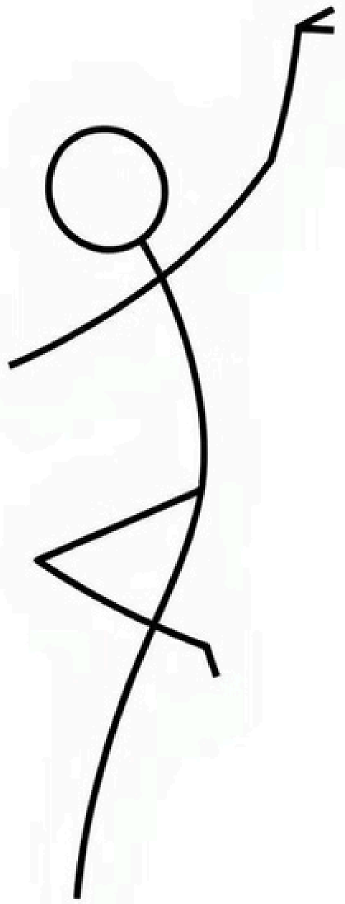


Rhythm-Aware Human Motion Diffusion Model for Dance Generation

Innovating Expressive & Realistic Dance Synthesis

Team: AdityaChaturvedi, Aditya Bajaj, Lovish Kumar

Supervisor: Dr. Alka Singhal

Institute: JIIT

Semester: 5th

Abstract

Our project tackles the challenge of generating realistic and rhythm-aware human dance motions. We propose a novel diffusion-based model that leverages **rhythmic conditioning** to synthesize diverse, high-quality dance sequences. This approach overcomes limitations of traditional generative models, offering significant applications in animation, gaming, and virtual reality.

Introduction: The Dance Generation Challenge

Generatingconvincing humandance motionis complexdue to its expressiveness, temporal dynamics, and critical connection to rhythm. Existing methods often fall short:

- **Recurrent Neural Networks (RNNs):** Struggle with long-term temporal consistency and diverse motion generation.
- **Variational Autoencoders (VAEs):** Can lack sharp details and sometimes produce blurry or unrealistic motions.
- **Generative Adversarial Networks (GANs):** Often suffer from training instability and mode collapse, limiting motion diversity.

Diffusion Models offer a powerful alternative, excelling in high-fidelity content generation and overcoming many of these limitations.

Project Objectives

Realistic & Expressive Motion

Generatehuman dance motionsthatare visually indistinguishable from real performances.

Rhythm-Aware Generation

Ensuresynthesized movements areperfectly synchronized and harmonized with input audio rhythms.

Style-Specific Control

Enable generationof dancein distinct styles (e.g., hip-hop, ballet) based on user conditioning.

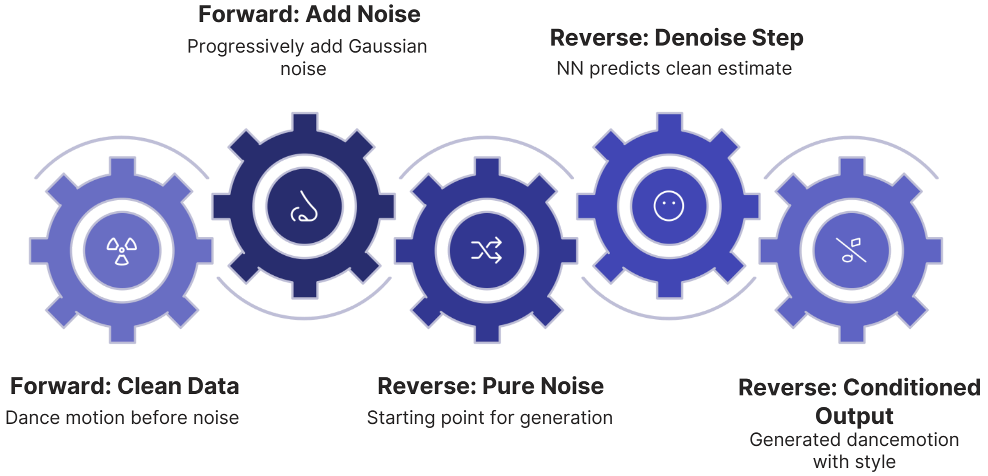
Temporal Consistency

Maintain smoothandcoherentmotion trajectories over extended durations.

Diffusion Models (MDMs) Overview

Diffusion Modelsoperate by progressively adding **Gaussian noise**to data (forward process) and then learning to reverse this process to recover the original data (reverse process).

- **Forward Process:** Gradually corrupts clean dance motion into pure noise.
- **Reverse Process:** A neural network learns to denoise step by step, reconstructing the dance from noise.
- **Conditional Generation:** We introduce **rhythmic and style information** at each denoising step, guiding the generation towards desired dance motions.



Algorithmic Landscape for Motion Generation

Abrief comparisonofgenerativemodel architectures:

RNN / LSTM	Sequential Data	Long-term Dependencies	Basic Sequences	Poor for expressiveness
VAE	Latent Space Encoding	Blurry Outputs	Abstract Motion	Lacks sharp detail
GAN	Sharp, Realistic Samples	Training Instability	High-fidelity, limited diversity	Mode Collapse issues
Transformer	Long-range Context	Computational Cost	Complex Interactions	Can be effective, but heavy
Diffusion	High Quality, Diversity, Stability	Inference Speed	Realistic, Controllable Motion	Excellent for expressive dance

Methodology Pipeline

Problem Statement

Rhythm-aware dance generation.

Dataset Preparation

AIST++, Mixamo, Human3.6M.

Data Preprocessing

Normalizationandrhythmfeature extraction.

Model Training

Losses: denoising,rhythm;Optimizer: AdamW.

Our pipeline integrates diverse datasets and specific loss functions to train a robust diffusion model capable of generating rhythmically accurate dance.

Inference Pipeline: Prompt to Dance

Input Prompt

Textdescription(e.g., "hip-hop dance," "energetic popping")

Audio Input

Musictrack(extracts rhythmic features)

MDM Inference

Denoisingprocess guided by prompt & audio

Output Dance Clip

High-fidelity,rhythm-aware dance motion

Users provide a text prompt or an audio track, and our model synthesizes a unique, synchronized dance performance.

Key Features & Strengths

Diffusion-Based Fidelity

Generateshigh-resolution,perceptually realistic motion.

Multi-Modal Conditioning

Precisecontrolviatext,audio,andstyle inputs.

Enhanced Generalization

Capableofsynthesizingdiversemotions beyond training data.

Temporal Coherence

Ensuressmooth,fluidtransitionsover long sequences.

Expected Output & Applications

Our Rhythm-Aware Human Motion Diffusion Model has the potential to revolutionize several industries.

1. **Animation & Film**

   Rapid generation of complex dance sequences for characters.

2. **Video Games**

   Dynamic and varied in-game character animations.

3. **Virtual & Augmented Reality**

   Creating immersive experiences with realistic avatars.

4. **Social Media Content**

   Personalized dance clips and virtual influencer content.