Final Project Proposal

Paper Implementation: [CVPR 2023] MoFusion: A Framework for Denoising-Diffusion-based Motion Synthesis

Group 32
110612117 張仲瑜
110550022 賴柏允

Goal

To implement the a diffusion-based motion synthesis model that can generate a sequence of motion based on the given text prompt or music.

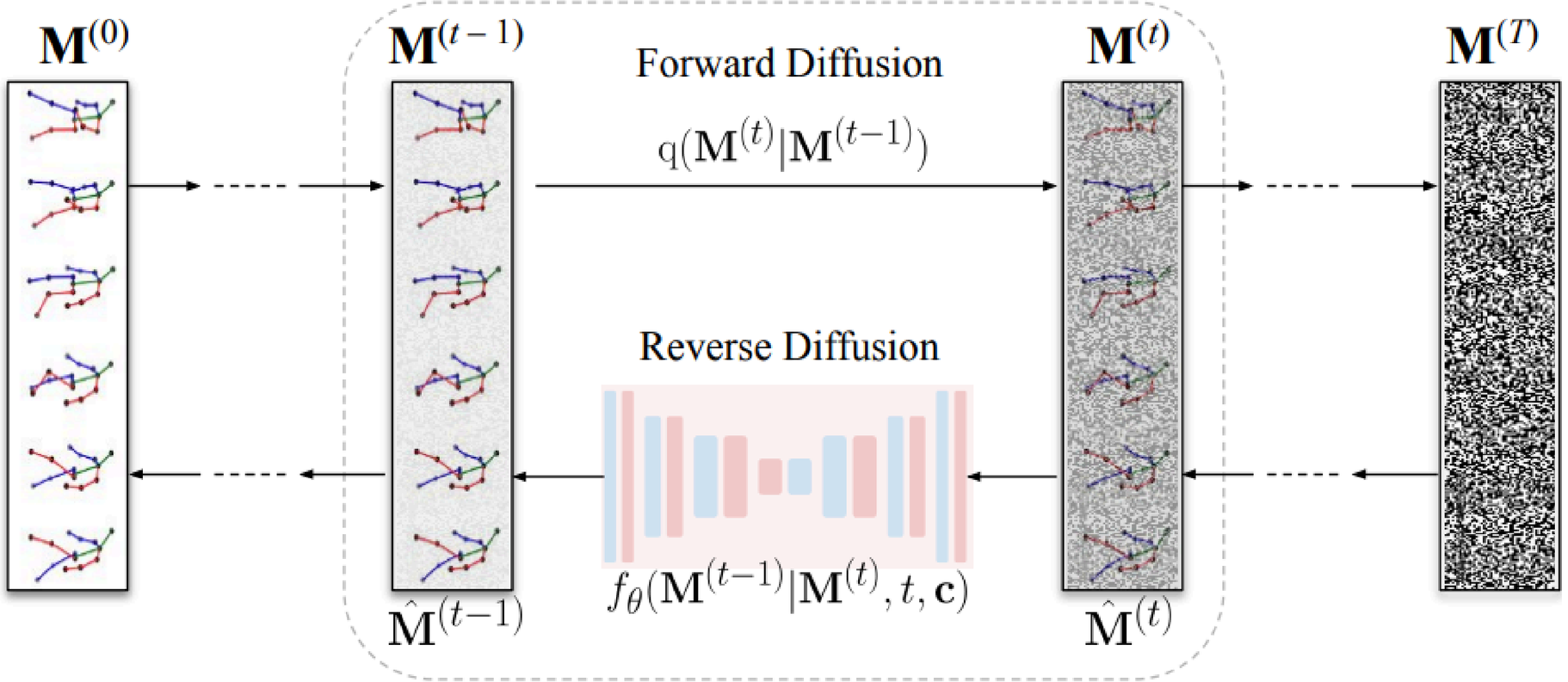
Related work

[ICLR2023]
MDM: Human Motion Diffusion Model

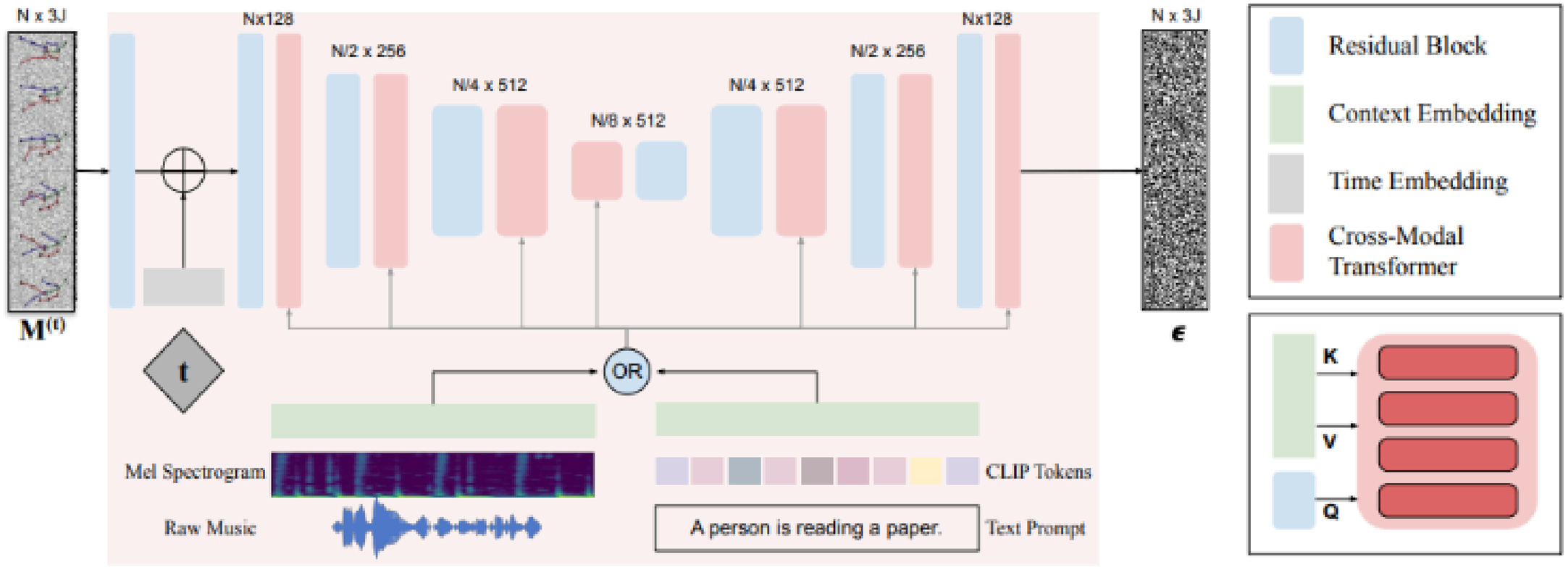
[ICCV 2023]
PhysDiff: Physics-Guided Human Motion Diffusion Model

[SIGGRAPH 2023]
Listen, denoise, action! Audio-driven motion synthesis with diffusion models

Core idea


$\mathbf{M}^{(0)}$
$\mathbf{M}^{(t-1)}$
Forward Diffusion
$\mathrm{q}(\mathbf{M}^{(t)}|\mathbf{M}^{(t-1)})$
$\mathbf{M}^{(t)}$
$\mathbf{M}^{(T)}$
Reverse Diffusion
$f_\theta(\mathbf{M}^{(t-1)}|\mathbf{M}^{(t)}, t, \mathbf{c})$
$\hat{\mathbf{M}}^{(t-1)}$
$\hat{\mathbf{M}}^{(t)}$

Overview


N x 3J
$\mathbf{M}^{(t)}$
Nx128
N/2 x 256
N/4 x 512
N/8 x 512
N/4 x 512
N/2 x 256
Nx128
N x 3J
ϵ
t
OR
Mel Spectrogram
Raw Music
CLIP Tokens
A person is reading a paper.
Text Prompt
Residual Block
Context Embedding
Time Embedding
Cross-Modal Transformer
K
V
Q

Pipeline

Feature Extraction

- Transformer to get the context embeddings

Forward Diffusion

- Adding Gaussian noise to a motion sequence
- Timestamp and context embeddings are included during the training

Backward Diffusion

- Train a model to denoise
- Loss
  - L2 distance between the predicted noise and that added in the forward process
  - Constraints
    - Symmetric Bone length
    - Consistent Bone length
    - Ground Truth supervision

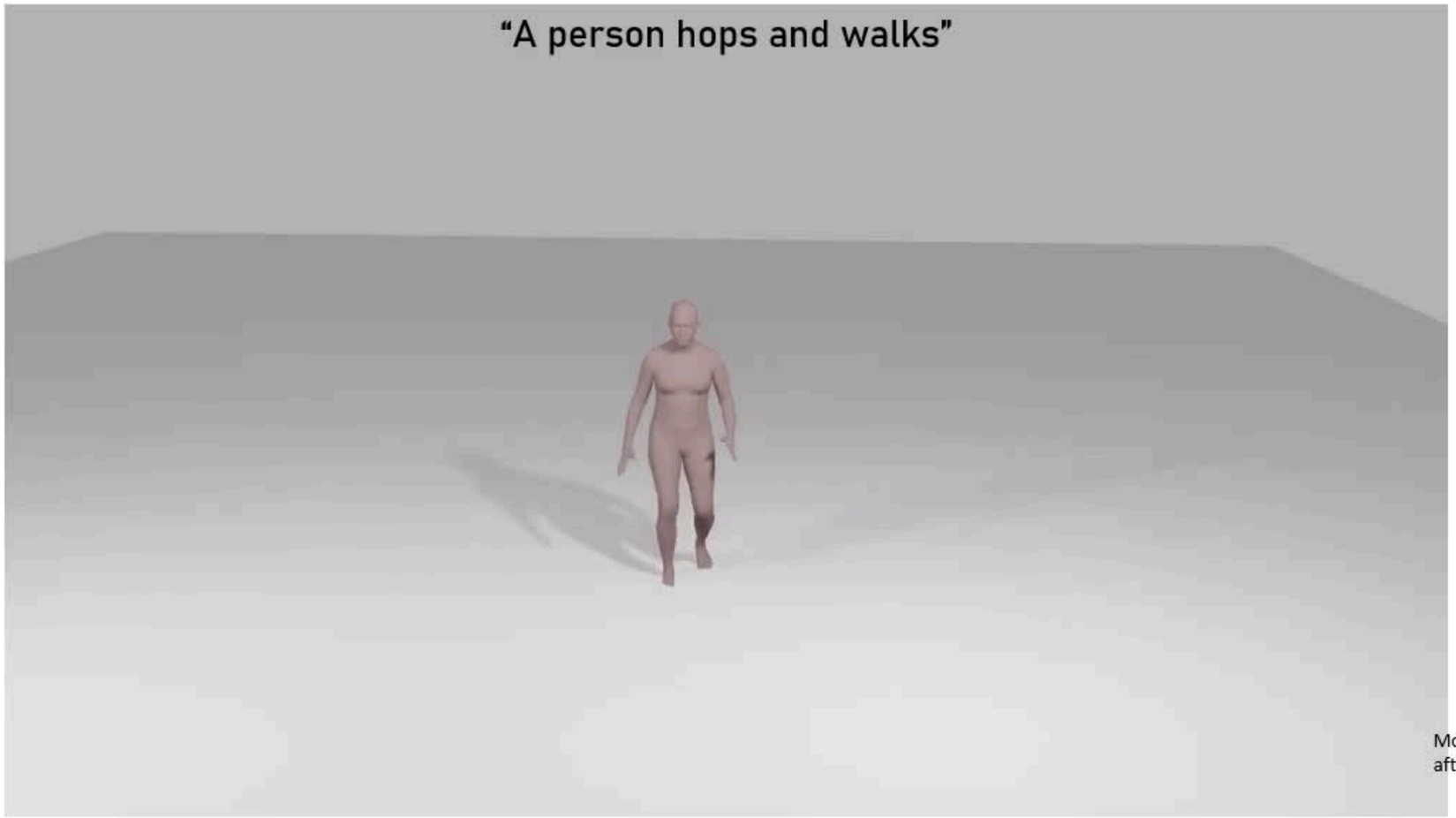
Expected Result

Things that-

- Are guaranteed (We'll try our best!) to work:
  - Code of building the Unconditional generation
  - Code of text to motion synthesis
  - Plan B: We will make a blender animation if things above don't work out the way we hope.

- May (might not) work:
  - Audio to motion synthesis
  - Same performance as the thesis (due to poor hardware, RTX 3060 Ti vs RTX A40)

Example


Text-to-Motion Synthesis
“A person hops and walks”
MoFusion results
after IK