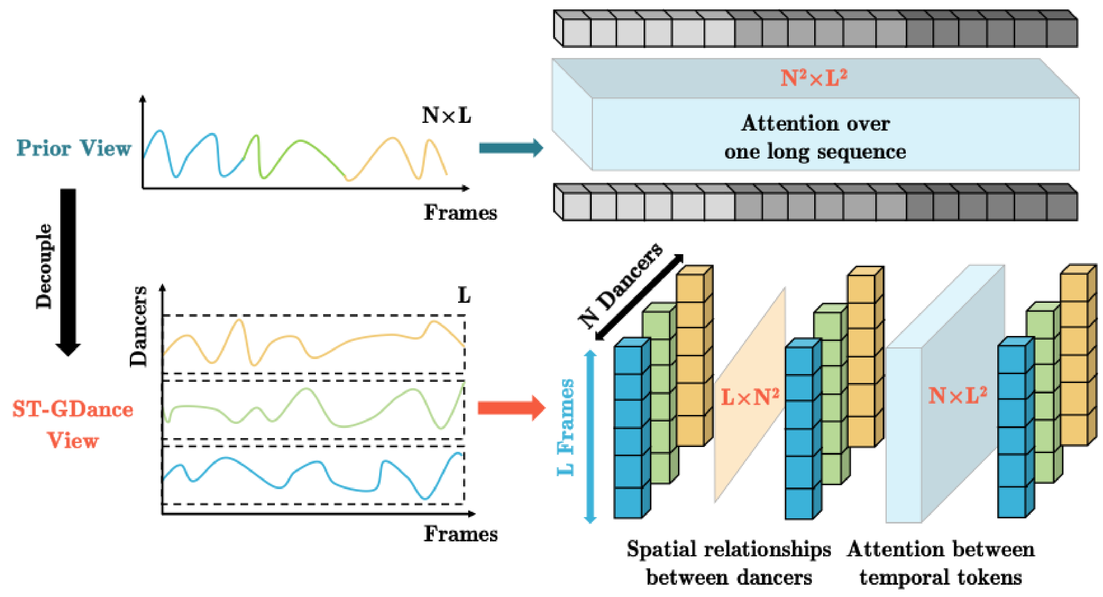
# ST-GDance: Long-Term and Collision-Free Group Choreography from Music

Jing Xu[1] Weiqiang Wang[1,†] Cunjian Chen[1,†] Jun Liu[2] Qiuhong Ke[1]
[1] Department of Data Science and AI, Monash University, Melbourne, Australia
[2] School of Computing and Communications, Lancaster University, Lancaster, UK

## Motivation

We identify that the efficiency challenges in long-term group dance generation mainly stem from **two factors**: the quadratic complexity growth in transformers with respect to **(1)** temporal sequence length ($L^2$) and **(2)** the number of dancers ($N^2$).




## Problem and Contribution

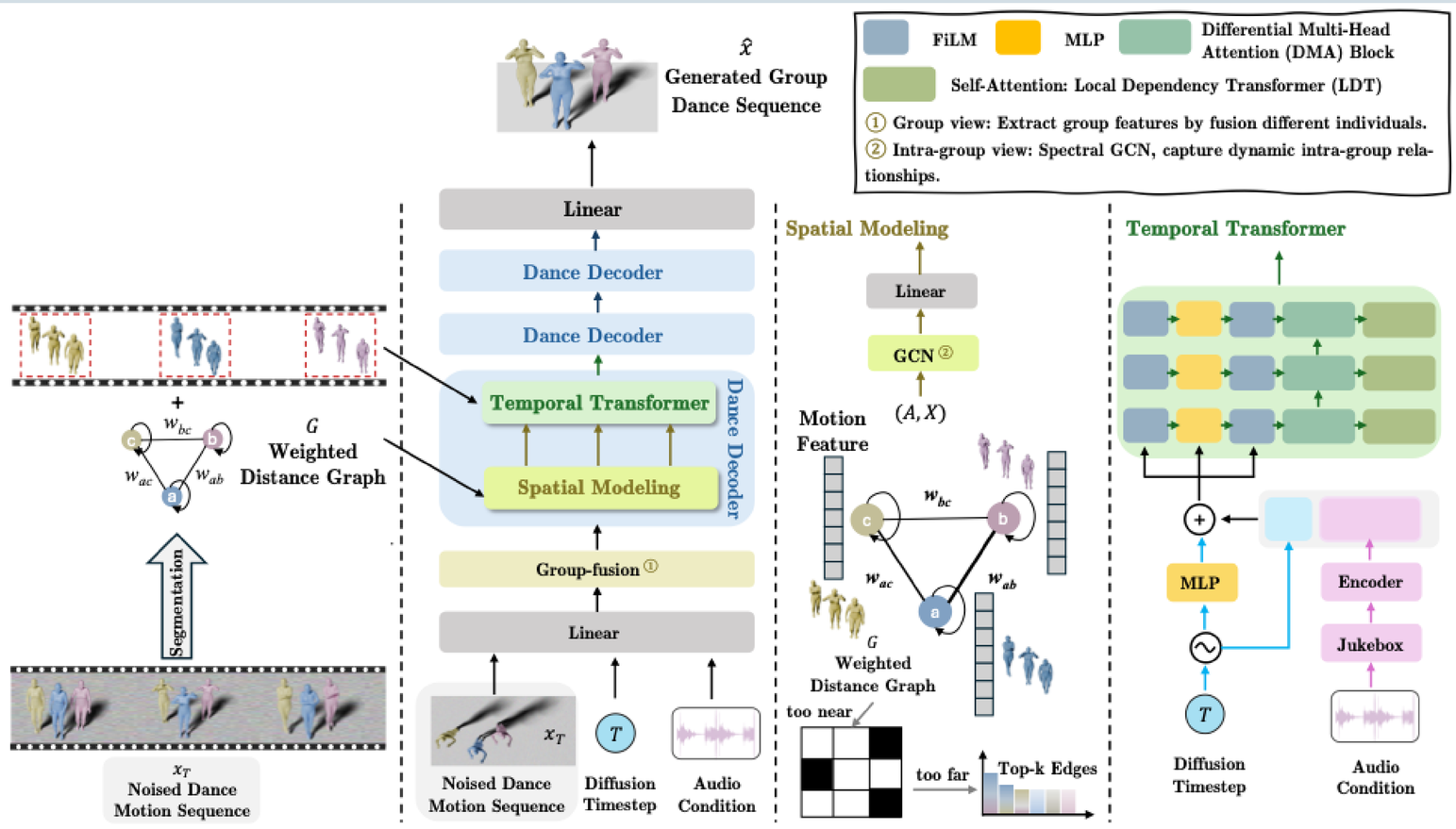
Goal: Generate long-term, collision-free group dances from music efficiently and coherently.

### Main Contributions:

- We propose **ST-GDance**, a novel **spatial-temporal** framework that decouples group dance generation into spatial and temporal components, enabling lightweight and modular processing.
- We propose a spatial modeling method that incorporates **spatial distance-aware constraints**, promoting structured and coordinated group choreography and preventing dance movement collisions.
- Our method avoids the traditional long-sequence paradigm and **reduces the denoising complexity from $O((NL)^2)$ to nearly $O(NL)$**, allowing efficient generation under limited resources, particularly for long frame sequences.

## Method

1. **Group Fusion** – Aggregate multi-dancer motions into a unified group representation to reduce dancer ambiguity.
2. **Spatial Modeling** – Apply a distance-aware Graph Convolution Network (GCN) to capture spatial relations and prevent collisions between dancers.
3. **Temporal Modeling** – Use a sparse Temporal Transformer with Differential Attention and Local Dependency modules to model long-term motion dynamics efficiently.
4. **Audio & Timestep Conditioning** – Inject music and timestep embeddings through FiLM and cross-attention to synchronize rhythm and refine motion coherence.



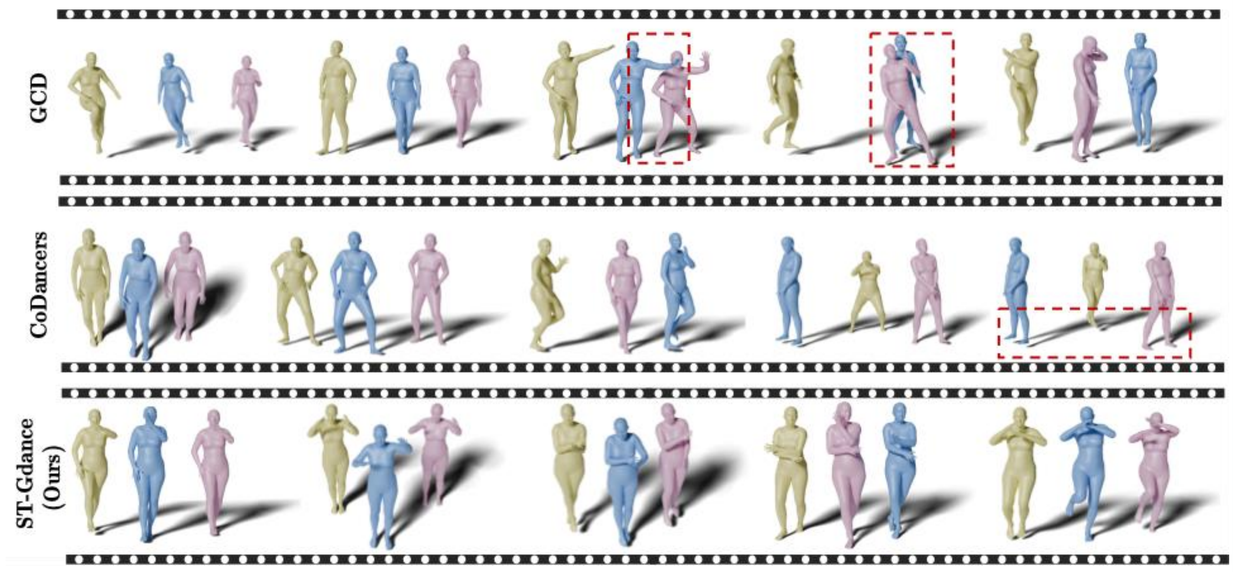

(a) Group Data Segmentation (b) Group Dance Generation Architecture (c) Spatial intra-group modeling (d) Speed-up temporal block

## Quantitative and Qualitive Results

### Experiments on GDance dataset.

| Setting | Method | Group Dance | | | Single Dance | | | Efficiency | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | GMR ↓ | GMC ↑ | TIF ↓ | FID ↓ | Div ↑ | PFC ↓ | FLOPs (GFLOPs) | Params (M) | Train Time / epoch (min) | Inf Time (s) |
| **Short-term** | EDGE | 70.26 | 62.24 | 0.42 | 33.24 | 10.11 | 3.07 | 12.50 | 52.13 | 0:24 | 8.59 |
| | GCD | 30.22 | 80.22 | 0.19 | 39.24 | 9.64 | 2.53 | 27.69 | 62.16 | 1:04 | 4.75 |
| | CoDancers | 26.34 | 74.22 | **0.10** | **23.98** | 9.48 | 3.26 | – | – | – | – |
| | TCDiff | 15.36 | **82.77** | 0.11 | 37.31 | **14.01** | **0.51** | – | – | – | – |
| | **ST-GDance** | **14.76** | 81.24 | 0.11 | 28.87 | 12.83 | 0.97 | **6.75** | **50.21** | 0:37 | **3.47** |
| **Long-term** | EDGE | 76.24 | 59.86 | 0.57 | 42.79 | 6.32 | 4.23 | 23.20 | 53.83 | 1:02 | 35.59 |
| | Lodge | 67.32 | 62.31 | 0.41 | **30.19** | 9.31 | 3.21 | 13.23 | 38.37 | **0:46** | 25.93 |
| | GCD | 42.64 | 67.85 | 0.20 | 32.42 | 8.94 | 3.95 | 31.60 | 62.29 | 2:51 | 31.42 |
| | CoDancers | 40.73 | 68.22 | 0.14 | 31.92 | 8.45 | 4.03 | – | – | – | |
| | **ST-GDance** | **31.76** | **75.18** | **0.13** | 30.21 | **11.35** | **3.11** | **8.12** | **50.35** | 1:25 | **17.84** |

### Qualitative comparison across different models.




## Acknowledgements

This research was supported by the Australian Government through the Australian Research Council's DECRA funding scheme (Grant No.: DE250100030).