

硕 士 研 究 生 读 书 报 告



题目Real-time Controllable Motion Transition for Characters

作者姓名 李蕊伶

作者学号 222151154

指导教师 李启雷

学科专业 电子信息

所在学院 软件学院

提交日期 2022年 12月29日

Real-time Controllable Motion Transition for Characters

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Electronic Information

Advisor: Qilei Li

By

Ruiling Li

Zhejiang University, P.R. China

2022

# 摘要

文章提出了一种新的实时转换方法来解决实时补间动作生成问题。问题的核心挑战在于需要同时满足三个关键条件:质量、可控性和速度，这使得任何需要离线计算或不能合并用户控制的方法都是不可取的。文章提出的方法包括运动流形和条件过渡。前者学习重要的低层运动特征及其动力学; 而后者综合过渡条件下的目标帧和所需的过渡持续时间。运动流形通过多模态映射机制明确地模拟人体运动中的内在过渡随机性。在生成过程中，文章设计了一个过渡模型，它本质上是一种基于目标帧和目标过渡持续时间从学习的流形中采样的抽样策略。在不允许后处理或离线计算的任务中，文章的方法在不同的数据集上得到了验证。通过详尽的评估和比较，该方法能够生成在多个指标下测量的高质量运动。

**关键词**：运动捕捉 神经网络 运动流形

# Abstract

This paper proposes a new real-time transition method to address the real-time in-between motion generation challenges. Its core challenge lies in the need to satisfy three critical conditions simultaneously: quality, controllability and speed, which renders any methods that need offline computation or cannot incorporate user control undesirable. The new real-time transition method consists of motion manifold and conditional transitioning. The former learns the important low-level motion features and their dynamics; while the latter synthesizes transitions conditioned on a target frame and the desired transition duration. The method first learn a motion manifold that explicitly models the intrinsic transition stochasticity in human motions via a multi-modal mapping mechanism. During generation, it designs a transition model which is essentially a sampling strategy to sample from the learned manifold, based on the target frame and the aimed transition duration. The method is validated on different datasets in tasks where no post-processing or offline computation is allowed. Through exhaustive evaluation and comparison, the method is able to generate high-quality motions measured under multiple metrics.

**Keywords**: Motion capture; Neural networks; Motion manifold.

# 引言

补间动作生成一直是计算机图形/动画中长期存在的问题。在深度学习的背景下。游戏中的离线动画管线和在线动作合成都非常依赖这一技术。前者不需要后处理或离线计算就能快速生成高质量的运动，后者往往是需要的。

在实时补间动作生成中存在两个主要挑战。其中最重要的是运动质量。因为动作需要快速生成，所以后期处理是非常不可取的。因此，离线计算和任何人为干预都将被严格排除在外。一个可行的解决方案是生成一个运动模型，它可以捕捉不同动作的精细动态，并作为运动生成的来源。然而设计这样的模型需要考虑人体运动固有的过渡模糊性，即多个帧或动作可能跟随一个给定的帧或动作。这就引出了第二个挑战:可控性。

早期的方法将补间运动作为运动规划问题，这需要解决复杂的优化问题，并且对于实时应用程序来说速度非常慢。为了处理运动和目标帧之间的任意数据，内存中所需数据的大小会呈指数级增长。在深度学习时代，中间运动可以被解释为运动流形学习问题或控制问题。

文章提出了一种新的实时转换方法来解决实时补间动作生成问题。该方法包括一个自然运动流形模型和一个用于运动生成的采样器。首先学习低水平运动动力学的自然运动流形。然后，引入采样策略，从满足目标帧和目标过渡持续时间的学习流形中生成运动。

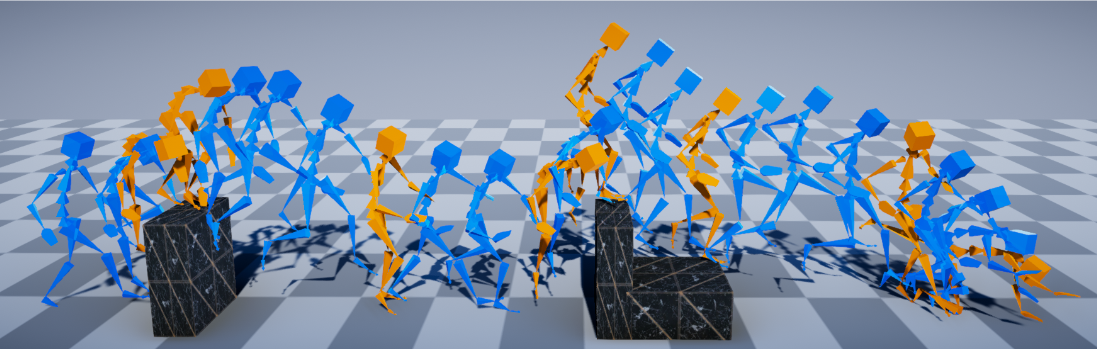


图2-1 实验效果

# 2. 内容介绍

## 2.1 数据格式

文章使用了LaFAN1数据集和Human3.6M数据集。从Human3.6M数据集中删除手腕和拇指关节，剩下21个关节。LaFAN1数据集的字符有22个关节。如图2-1所示，对8个下部关节采用基于位置的表示，对上部关节采用基于旋转的表示。所有较低的关节连接少于两个其他关节，以方便确定它们的方向。

关节位置j包含一个三维矢量表示关节的速度，一个三维矢量表示向上方向。为了均匀性，文章将3D向上方向替换为2轴旋转矩阵。

两个数据集都包含多个主题。测试集中的受试者与实验中的训练集不同，这保证了运动模型在训练后可以泛化到不同的受试者。文章以25HZ的采样率工作。这两个数据集的训练集都被分成多个50帧的窗口，两个连续的窗口有25个重叠的帧。

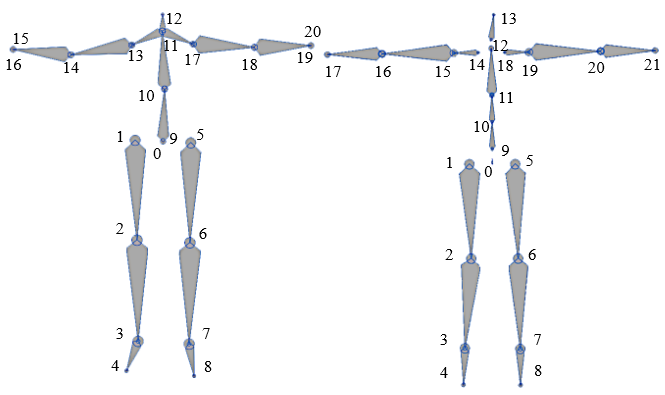


图2-1 数据格式

## 2.2 运动流形

**2.2.1 理论推导**

文章将所有关节分为三组:上半身、髋关节和下半身。根据经验发现，它们动作生成中具有不同的重要性。髋关节速度可以很好地指示下一帧。由于运动会存在潜在的脚部滑动，下体关节的运动会显著影响运动的视觉质量。相对而言，上肢关节约束较小。因此，本方法重点学习下半身和髋关节的运动。

从前n-1帧中产生一个运动M = {}，每帧用表示，其中p、r、v分别为关节位置、旋转和速度，下标L、h、U分别表示下体、髋关节和上体关节。给定起始帧，目标帧，目标过渡时间, M的联合概率可以表示为:

假设 和之间是独立的然后省略它们。在马尔可夫假设下，P(M)可分解为:

其中P可以通过多种方式学习，在这里使用了共嵌入策略，因为它很容易允许引入条件变量。引入一个潜变量z来编码两个连续帧的共嵌入，并使用下一帧的髋关节速度作为条件变量。z编码两个连续帧的转换概率，可以帮助消除下一帧的歧义，对此我们将在引入转换采样时给出详细信息。将z和引入得到:

重点学习下半身和髋关节的运动，对公式进行修改：

其中由当前框架的下半身和髋关节组成。假设在预测期间给出，因此它的先验分布可以被移除。是角速度。由于z编码两个连续的帧，所以它可以通过或进行独立学习。如果假设z ~ N(0, I)，其中N是正态分布，那么P(z)可以被认为是条件VAE或CVAE的编码器，潜空间分布被约束为正态分布，如图2-2所示。

其中解码器的条件变量是专门设计来捕捉人体运动固有的模糊过渡，旨在通过多模态映射学习判别转换。具体来说，给定一个非判别嵌入z，解码以当前帧的和未来髋关节速度()为条件。在解码器中采用了CMoEs模型，如图2所示。在学习过程中，CMoEs可以学习判别映射，其中每个专门的网络倾向于专注于学习运动的一个阶段。

**2.2.2 训练细节**

文章在训练中首先使用重构损失和kl散度损失，粗略的训练网络。然后增加了模型的滑步损失和骨长损失。同时文章将潜变量z和髋关节运动速度传递给专家网络的每一层，以避免后验塌陷。

CVAE的编码器有两个隐藏层，包含256个单元，ELU激活。门控网络有两个隐藏的前馈层，ELU激活。门控网络的输出层采用Softmax激活。每个专家网络是一个三层前馈网络，隐含层中有256个单元，ELU激活。实验中根据经验将专家网络数设置为6。第一次采用计划抽样策略。该网络将最后一个时间步长的预测姿态作为概率p的输入，否则，它将从数据集中获取ground truth作为输入。

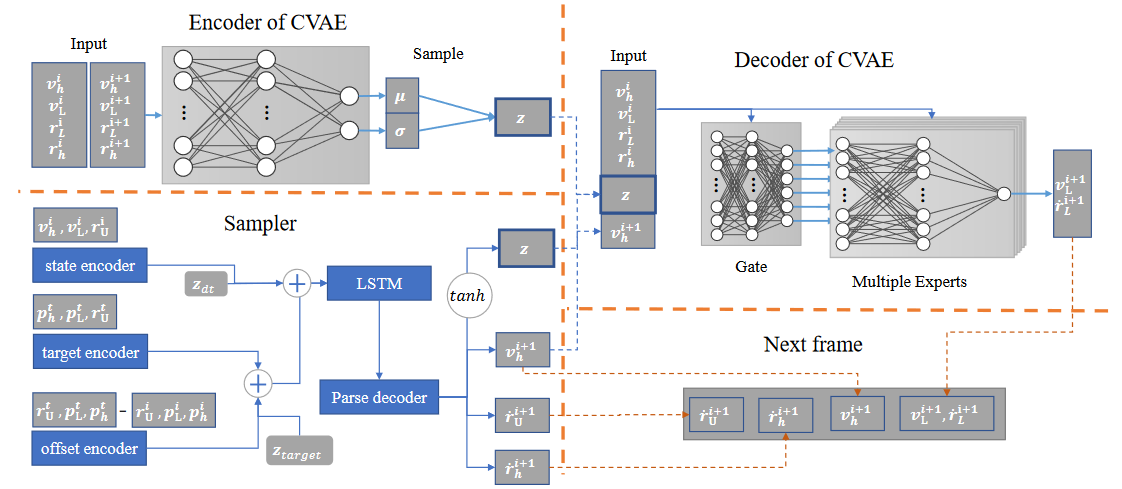


图2-2 网络结构

## 2.3 过渡采样

为了能够连续地生成运动，需要对z和进行采样，并在和的约束下，在给定的情况下生成。所以网络本质上是一个从学习过的流形中采样帧的采样器。

采样器考虑目标编码器和偏移编码器的约束，分别对目标帧和当前帧与目标帧之间的偏移进行编码。关键输出是下一帧的z和。此外，当用于解码下一帧时，z和将通过CVAE的解码器被导出，进行流形投影来细化姿态。同时向编码矢量添加随时间变化的噪声，从方差等于0.5的高斯分布零中心采样。它的振幅在接近目标帧时减小，因此采样器的注意力只在接近目标时集中在目标上。

接下来，循环神经网络利用所有潜在向量来预测下一个状态。解码器解析状态生成采样(z, )和上面的关节。

当将帧传递到编码器时，该方法用髋关节速度、下部关节的速度和上部关节的旋转来表示S，以降低与完整状态相比的维数(，，)，并使用下部关节的位置，来计算偏移。为了平衡对下部关节和上部关节的注意力，该方法在将传递到偏移编码器之前对其应用z-score归一化。所有三个编码器都是两层前馈网络，第一层隐藏层有512个单元，第二层有256个单元。每一层都是PLU激活。解析解码器有三层，第一层隐藏层512个单元，第二层256个单元，其次是ELU激活。z是CVAE对下一帧采样的关键输入，为了计算z，该方法应用tanh函数并将输出按4.5比例缩放，以确保正态分布的良好覆盖。

# 3. 创新点及应用

（1）文章提出了一个新的学习框架，由一个新的自然运动流形模型和一个新的过渡采样器组成，用于实时生成补间运动。运动流形模型分别处理不同的身体部位，重点关注运动的可控性和质量，而过渡采样器确保在用户控制层面上产生自然的运动，如图3-1所示。

（2）文章的模型能够生成高质量的运动以减轻滑步问题，可以用于离线动画和在线游戏。文章介绍的方法在不可见的控制信号下也是通用的。它优于其他解决方案和最先进的方法。

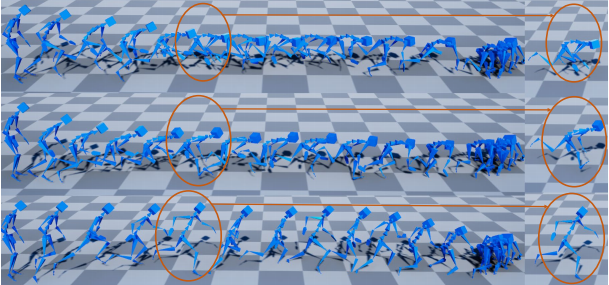


图3-1 自然奔跑动作

# 4. 局限性

（1）生成结果受到训练数据的限制。它不能生成与训练数据差异太大的运动。如果目标帧与数据、空间、时间或两者差异太大，该方法就不能保证100%实现目标帧。然而，框架本身仍然是有效的，当引入更多样化的数据时，上述问题可以很容易地克服。

（2）缺乏运动多样性是文章提出的方法的另一个局限性。作为基于CVAE的网络，该方法的模型可以对同一控制产生不同的运动，但是所产生的运动之间差异很小，尤其是下半身。

**参考文献**

[1]. Tang, Xiangjun et al. Real-time controllable motion transition for characters. ACM Transactions on Graphics (TOG) 41 (2022): 1 - 10.