

硕 士 研 究 生 读 书 报 告



题目 为角色运动的综合和编辑提供的深度学习框架

作者姓名 王晓瑶

作者学号 21651026

指导教师 梁秀波

学科专业 软件工程

所在学院 软件学院

提交日期 二○ 一七年四月

A Deep Learning Framework for Character Motion Synthesis and Editing

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: Xiubo Liang

By

Xiaoyao Wang

Zhejiang University, P.R. China

2017

# 摘要

该篇报告主要对A Deep Learning Framework for Character Motion Synthesis and Editing这篇论文进行阐述。该篇论文提出了一个基于高水平参数综合角色运动的框架，使得所产生的运动符合人体运动的多样性，并在大型运动捕捉数据集上进行训练。由卷积自动编码器的隐藏单元表示的学习运动流形可以组合以产生广泛复杂运动的稀疏分量中的运动数据。这个网络经过训练，可以从诸如角色应该遵循的地形上的曲线等参数或打孔和踢球的目标位置产生逼真的运动序列。前馈控制网络和运动多样被独立地训练，允许用户根据期望的接口容易地在前馈网络之间切换，而无需重新训练运动流形。一旦生成运动，就可以通过在运动流形的空间中执行优化来编辑它。这允许施加运动约束，或改变运动的风格，同时确保编辑的动作保持自然。因此，该系统可以产生平滑，高质量的运动序列，而无需对训练数据进行手动预处理。

**关键词：**深度学习，人类动作，角色动画，流形学习

# Abstract

This paper focuses on A Deep Learning Framework for Character Motion Synthesis and Editing. This paper presents a framework based on a high level of parametric integrated motion that allows the resulting motion to conform to the diversity of human motion and is trained on large motion capture data sets. The learning motion manifolds represented by the hidden units of the convolutional autocoder can be combined to produce motion data in the sparse components of the widely complex motion. This network is trained to generate realistic motion sequences from parameters such as curves on terrain that should be followed by the character or the target location of punching and playing. The feedforward control network and the movement are individually trained to allow the user to easily switch between feedforward networks according to the desired interface without having to retrain the motion manifold. Once the motion is generated, it can be edited by performing optimizations in the space of the motion manifold. This allows the application of motion constraints, or changes in the style of movement, while ensuring that the editing action remains natural. Thus, the system can produce smooth, high-quality motion sequences without the need for manual pre-processing of training data.

**Keywords:** deep learning, human motion, character animation, manifold learning

# 1.简介

数据驱动的运动合成允许从高水平参数中生成可信的角色动作，这种方法使得动画师无需通过关键帧提供低层次的细节，只需要提供高水平的指导即可。目前大部分的数据驱动方法都需要大量的手工数据预处理，包括运动分割、运动对齐和运动标记。任何一个阶段的错误都很容易导致最终动画的失败，因此这种预处理通常会通过大量的人工干预确保输出的动作流畅自然，这些使得系统的完全自动化实现困难，经常需要专门的技术开发人员进行维护。

该篇论文主要提出了一个基于深度学习框架的动画合成和编辑的模型，该模型无需手动数据预处理或者人工干预，使用大量人体运动数据以一种非线性流形学习的方式，学习嵌入的运动数据。我们在大量运动数据集上训练卷积自编码，这样可以让再现的运动数据和通过插值合成的新的运动作为输入，这种无监督的非线性学习过程不需要任何运动分割或对准，这使得该过程比以前的方法显着更容易，在这个自动编码器的顶部，我们堆叠另一个前馈神经网络，将高级别参数映射到低级别人体运动中。这样，用户可以从直观的输入（例如角色应该遵循的某个地形上的曲线）或用于打孔和踢球的末端执行器的轨迹，轻松地产生逼真的人体运动序列。由于前馈控制网络和运动模块是独立进行训练的，所以用户可以根据需要的接口轻松地交换和重新训练前馈网络。我们的方法本质上是并行的，这使得计算速度非常快，并且适合主流动画包。

该篇论文还提出了在运动多样化的空间中编辑运动数据的技术。卷积自动编码器的隐藏单元以稀疏和连续的方式表示运动，使得在该空间中调整数据保持运动的自然度和平滑度，同时仍然允许复制身体的复杂运动。这种编辑的示例性例子是通过最小化合成运动的隐藏单位和参考样式运动的Gram矩阵的差异来将一个运动的风格与另一运动的时间相结合。

这篇论文首先回顾基于内核的运动合成方法，这是通过混合运动捕捉数据合成运动的主要技术。接下来回顾交互式角色控制的方法，其中利用用户指令来知道运动数据库合成新的动作。最后，回顾深度学习中的方法以及它们如何应用于角色动画中。

这里主要介绍深度学习在角色动画中的应用。基于深度学习的技术目前是图像和语音识别领域最先进的技术，基于深度学习的框架的一个重要优点是它们自动学习数据集的特征。例如，当将卷积神经网络（CNN）应用于图像识别时，类似于Gabor滤波器的滤波器出现在底层，并且对应于不同对象的更复杂的滤波器出现在较高级别的层中。该篇论文主要就是将这样的优势应用于角色动画中。深度学习和神经网络也吸引了控制社区在机器人和物理动画等应用中的兴趣。Allen和Faloutsos使用NEAT算法，演化了用于控制双足字符的神经网络的拓扑。Mordatch等应用经常性神经网络（RNN）来学习用于铰接字符的近似最佳反馈控制器。虽然这些方法学习了如何在物理环境中控制角色的动态，但该篇论文的重点是从捕获的人体运动数据中学习特征并将其应用于动画制作。

已经有几种方法将深度学习应用于人体运动捕获数据，Du等人使用来自各种来源的大运动数据集构建分层RNN，他们的方法达到了较好的识别率。

该篇论文所使用的系统中使用来自大型人体运动数据库的数据，对卷积自动编码器进行了训练，因此发现了一般的运动流形，并且在这次训练之后，运动可以由网络的隐藏单元来表示，给定该表示，通过堆叠在卷积自动编码器顶部的前馈神经网络，在高级别控制参数和隐藏单元之间产生映射，这项工作中显示的高级别控制参数是地形上的人物轨迹以及末端执行器的运动。由于这些参数化可能包含歧义，因此该系统有另一个小型网络用于计算消除输入歧义的参数，只有与该任务相关的运动捕捉数据的子集用于训练这些前馈网络。一旦生成了运动，就可以在隐藏单元的空间中进行编辑，以便产生的运动满足诸如滑步处理的位置限制。使用这种技术，该论文描述了一种转换方法，这种转换方法使用短样式化剪辑作为参考，以此来转换角色运动的风格。

# 2.具体工作

## 2.1 数据获取

该论文描述了构建适合深度学习的大型运动数据库。首先我们==通过收集许多免费提供的运动捕获大型在线数据库来构建深度学习运动数据集，以及从我们的内部捕获中添加数据，并将其重新定位到具有单个刻度和相同骨长度的统一骨架结构。重定向是将源骨架结构中的任何对应关节角度复制到目标骨架结构，然后使得源骨架与目标骨架尺寸相同，最后执行全身逆运动学方案移动目标骨架的关节以匹配源骨架中的任何对应关节位置，一旦构建，最终的数据集大约是CMU运动捕捉数据库大小的两倍，并且包含以每秒120帧采样的单个字符的大约600万帧的高质量运动捕捉数据。

捕捉到的运动数据集需要转换成合适的格式来进行训练。该论文中将数据库中的所有运动都采样到每秒60帧，并将数据从原始数据集中的关节角度表示转换为3D关节位置格式。

## 2.2 建立动画流形

为了在人类运动空间上构建一个流形，论文中构建了一个自动编码卷积神经网络，并将在完整的运动数据库训练它。我们遵循Holden等人的做法，但采用不同的设置来优化运动合成网络。首先，论文中只使用单层编码运动，因为由于网络的池层减少了数据的时间分辨率，因此重建之后的多层合并/解除池可能导致重构后的模糊运动。论文中也改变了几个组件来改善网络训练的表现和基础质量。

## 2.3 将高级参数映射到人类运动

论文中使用前馈卷积神经网络学习高水平参数和角色运动之间的回归。高级参数是用于描述运动的抽象参数，例如投影到地形上的根轨迹或终端执行器（例如手和脚）的轨迹，是一个可以应用于各种类型的高级参数和动画输出的一般框架，由于在输出中存在巨大的歧义和多模态，因此在低维，高水平参数与全身运动之间产生映射是一项艰巨的任务。有许多不同的有效运动可以遵循高水平的参数来执行，例如当字符被指示沿着一条线行进时，运动的时间是完全不变的：一个角色可能以不同的步长行走，或者与在相同的轨迹上执行的另一个运动不同步，混合这些不同步运动导致输出的平均值，使得角色看起来沿着路径浮动。论文还描述了前馈网络的结构以及如何进行训练，然后介绍了运动任务的高级参数的细节。

## 2.4 隐藏单位空间中的运动编辑

该篇论文中描述了如何编辑或转换隐藏单元空间中的运动风格，这是自动编码器学习的运动数据的抽象表示。因为运动是在隐藏单元的空间中进行编辑的，所以在有效的运动中参数化了多维数据，它确保即使在编辑之后，运动仍然平滑自然。论文中将约束作为隐藏单位价值的成本。作为最小化问题的运动编辑的这种形式通常是十分方便的，因为它指定了编辑的期望结果，而不推断任何实现它的技术。论文中首先描述应用运动学约束的方法，然后就是关于调整隐藏单元空间中运动的风格。

# 3.实验结果

论文中展示了训练和合成角色运动的一些实验结果。

关于地形上的运动，论文中训练前馈网络，使得在地形上绘制的曲线被用于产生人物的实际运动。在数据库中的数据中，使用不同类型的具有不同速度和步进模式的运动数据来训练系统，使用训练数据，将根的轨迹投影到地面上，以产生要用作输入的地形曲线。

关于隐藏单位空间中的运动编辑，论文中将前馈网络生成的约束应用到前后的运动。论文中还将隐藏单元空间中编辑的运动与通过逆运动学在笛卡尔空间中编辑的运动进行比较。前者产生更平稳的结果，因为运动在精简的流形上编辑。

关于运动风格的转换，论文中展示了一个使用单独的运动剪辑来改变角色运动风格的例子，给出了（1）僵尸风格（2）沮丧风格（3）老人风格（4）受伤风格，角色的风格数据通过自动编码器计算格式矩阵来得到。运动数据取自数据集，而风格数据是内部捕获的组合。

关于人群动画, 论文中允许使用GPU在时间轴上进行并行计算，这样允许一次创建许多角色的运动，利用这个功能，使用地形曲线框架来应用论文中的系统以此来合成大量的角色动画。

运动合成的许多其他方式可以归类为时间序列方法，但论文中的运动合成方法是一种程序性的方法，因为它不需要逐步进行计算，并且可以在任意时刻按需求产生单个帧，这使得这种合成方法非常适合如Maya一类的动画制作软件，可以使动画师方便的跳到时间轴上的任意一点，也不需要担心影响本地卷积窗口之外的生成动画的更改，这也使，得系统可以实现高度并行化，因为可以独立地计算所有帧的运动，并在运行时间内以较快的速度在GPU上生成许多角色的轨迹，跨时间线并行生成运动需要帧之间的连续性，论文中的框架以两种方式处理，通过生成脚部接触信息来提供高电平连续性，或者通过流形确保低电平连续性。

论文中的框架需要仔细选择前馈网络的输入参数，使得高电平参数和输出运动之间的歧义很小。模糊度是机器学习中的常见问题，其中当多个输出对应于训练数据中的相同输入时，回归器的输出被平均化。在某些情况下，可能需要解决歧义的附加数据，这可以由用户直接提供，或者使用额外的模型来发现，如我们的脚部接触模型。

# 4.结论

该篇论文提出了一个深度学习框架，将高级参数映射到输出运动，首先使用大量运动数据库学习运动流形，然后生成输出运动和输入数据之间的映射，同时还提出了在同一框架下编辑和改变运动风格的一些方法。

目前，论文中的自动编码器只有一层，因为深层堆叠自动编码器在脱模过程中会出现模糊。在系统中，组合和抽象低级特征的作用由叠加在其上的前馈网络覆盖，然而如果使用堆叠的深度自动编码器来学习运动流形，则可以使用更为简单的前馈网络，这样更容易进行训练。

# 参考文献

[1] Holden D, Saito J, Komura T. A deep learning framework for character motion synthesis and editing[J]. Acm Transactions on Graphics, 2016, 35(4):1-11.