

硕 士 研 究 生 读 书 报 告



题目 Artemis: articulated neural pets with appearance and motion synthesis reading report

作者姓名 胡家豪

作者学号 22251098

指导教师 李启雷

学科专业 软件工程

所在学院 软件学院

提交日期 2023年1月

Reading Report of Artemis: articulated neural pets with appearance and motion synthesis

A Dissertation Submitted to

Zhejiang University

in partial fulfillment of the requirements for

the degree of

Master of Engineering

Major Subject: Software Engineering

Advisor: Qilei Li

By

Jiahao Hu

Zhejiang University, P.R. China

2023.1

# 摘要

**关键词：**

Abstract

NeRF models and improvements are used in robotics, urban mapping, automated navigation, virtual/augmented reality, and more. At present, many researchers regard the realization of virtual human by NeRF as an important research topic. In this paper, the authors propose a method named Few-shot Dynamic Neural Radiance Fields which can reconstruct and edit 3D faces from a small number of dynamic frames extracted from monocular video. In addition, a CFW module is introduced to perform different expression conditioned warping in 2D feature space to deal with the inconsistency between dynamic frames. This module also has identity adaptive and 3D constrained. The architecture also supports free edits of facial expressions, and enables video-driven 3D reenactment.

**Keywords：**3D face reconstruction, expression editing, NeRF, few-shot and

dynamic modeling

1 研究背景与论文简介

在实时渲染中，对于多毛动物的毛发和连续的行动模拟对于视觉特效工作室来说始终是一个难题，这些多毛动物的仿真模拟需要消耗大量的计算资源且需要艺术家的大量精力投入。然而，即使是在拥有充足计算资源的情况下，拥有高度真实感的，可以满足用户对动物的近距离观察和实时交互的动物仿真仍然面临很大的困难。

挑战来源于多个方面，冲突的核心在于对渲染实时性和对渲染仿真程度难以同时满足的冲突。一方面，如果以视频和照片作为参考，动物身上有着以数十万根毛发纤维覆盖的毛皮，以传统方式进行建模过程非常繁琐，需要高超的艺术技能和大量的精力投入。另一方面，手动渲染的过程也要消耗大量的时间，非实时的光线追踪算法可以满足半透明，光散射，体积阴影等多种物理效果的模拟来使得画面更加逼近现实中的动物，但该算法的渲染速度非常慢，即使使用最先进的图形硬件也远远无法满足实时渲染的要求。除上述问题以外，以实时交互速度和高逼真度为模型设置相应的反应动画也同样非常困难。因此，一个集成有渲染和动画组件的可迁移自动化工具能够为多毛动物的渲染带来很大的帮助。

本论文介绍了ARTEMIS，(ARTiculated neural pets with appEarance and Motion synthesIS)一种全新的神经建模和渲染框架，相比于现有的非实时动画和渲染系统，ARTEMIS支持实时的交互式运动控制，逼真的动画和多毛动物渲染。同时，作者们还把ARTEMIS拓展到了OpenVR，希望通过VR眼镜为用户提供如同在现实世界与动物接触和互动的逼真体验。示意图如下图所示



图1.1 数字动物在VR环境中的展示

# 2 实现方法

**2.1 ARTEMIS框架整体概览**

ARTEMIS框架主要由两个核心组件组成，整体结构如图2.1所示

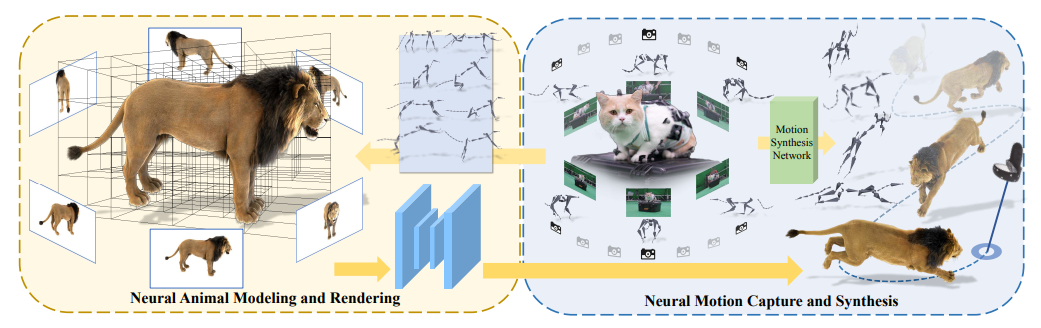


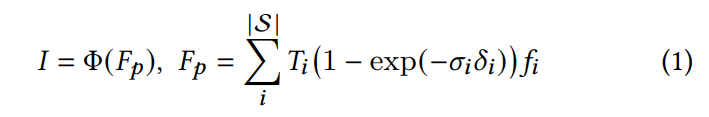
图2.1 ARTEMIS框架整体概览

对于第一个模块（建模与渲染模块），根据该模块给定的CGI的动物资产和与该资产相关的骨架数据，蒙皮权重，具有代表性的多种姿势的RGBA多视图，框架开发者构建了基于动态八叉树的神经网络来满足动态动物的实时骨骼动画绘制和实时渲染，并支持实时交互。

对于第二个模块（动作捕捉与动作合成模块），开发者构建了一个带有多视图RGB和VICON相机的混合动物运动捕捉系统，以重建真实的3D骨骼姿势，这支持训练神经运动合成网络，以使用户能够交互式地指导神经动物的运动。ARTEMIS系统进一步集成到现有的消费者级VR头戴式耳机平台中，用于神经生成动物的沉浸式VR体验。

**2.1神经不透明度辐射场NeORF实现过程**

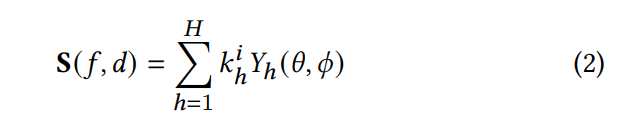
该模块通过神经网络对动物毛发进行模拟，传统方法一般直接对数十万计的毛发纤维进行模拟，而后再模拟这些毛发的运动。本模块采用神经不透明度辐射场（Neural Opacity Radiance Fields）的方式来对动物毛发的半透明性进行模拟。经由NeORF产生的毛发相比于自然真实的毛发存在透明度嘈杂和不够连续的问题，最近的ConvNeRF通过在特征空间中处理图像而不是直接在RGB颜色中处理图像来解决这个问题，神经不透明度辐射场表示如下：



Φ是不透明度合成网络，𝐹𝑝 是聚集的图像特征，S是采样点的集合，𝜎𝑖, 𝛿𝑖 分别是采样点之间的密度和距离。这里的关键是功能的使用𝑓𝑖 以表示每个空间点处的容量，而不是直接使用密度。

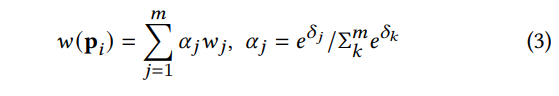
**2.2 八叉树特征构建**

以CGI动物角色的阿尔法蒙版数据作为输入来构建八叉树，得到的八叉树包含一个三维像素阵列P，而后分配一个数组Features Look up Table（FLUT）用于存储原始PlenOctree中的特征和密度值。这里FLUT用于在任意3D位置高效地查询特征，以加速训练和推理。对于不透明度特征值的建模同样采用球谐函数：

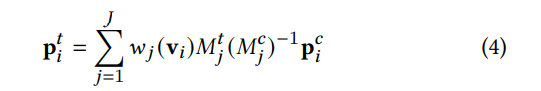


**2.3 骨骼绑定和变形**

为使得八叉树适用于动物模型，把八叉树装配于相应的特征姿势所属的骨骼上。暴力方法是根据体素到骨骼的距离装配，本框架根据给定的网格顶点和相应的蒙皮权重，通过混合周围最近的m个顶点的蒙皮权重的方式来生成该顶点的权重

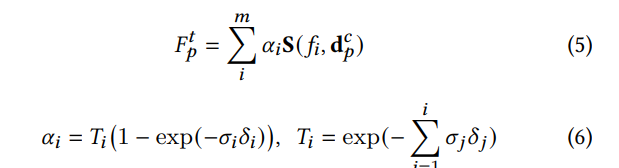


在获得了标准骨骼下的体素集，蒙皮权重和骨骼数据后，可按照LBS对骨骼从标准姿势状态变换到目标姿势状态



**2.4 动态体渲染**

将八叉树变换为目标姿势后，即可通过体渲染来渲染神经网络动物，然后通过端到端训练。对于给定像素起点和射线方向的射线，可以根据目标姿势的骨骼数据计算对应的视角特征值



到这一步已经完成了动体向目标姿势的神经外观特征和不透明度的光栅化。最后一步采用ConvNeRF的额外U-Net架构Φ来执行图像渲染把动体转化为保留细节特征的彩色图像即可完成。

后续的部分对于神经网络的相关损失函数和混合渲染等进行了具体说明，本报告不对此做更多展开

**2.5 动作捕捉与动作合成**

动物的运动数据来源于动物捕捉，论文撰写者构建了两种类型的动物运动捕捉系统，第一种由用于大型动物的RGB摄像机阵列组成，第二种组合了用于温顺和小型宠物的RGB摄像机和维康摄像机

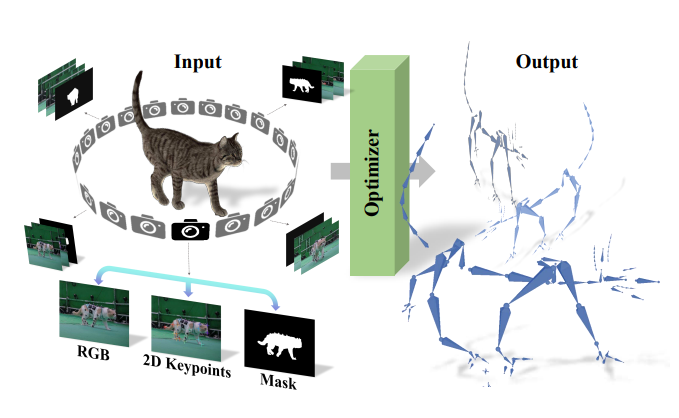
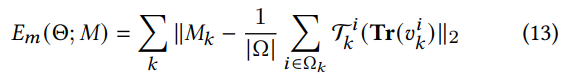


图2.2 用于进行动作捕捉的相机示意图

尽管四足动物在物种间具有相似的骨骼结构，但它们的形状和规模却截然不同。捕获适合所有类型四足动物的mocap数据集是不可能的。因此，论文撰写者们从温顺和小宠物身上收集运动数据并尝试把相关经验转移到老虎和狼等大型动物身上。在与马戏团合作再次收集相关数据后，后者的数据精度也得到了相关提升

采集足够数据后，采用参数化SMAL动物姿势数据模型对动物姿势进行预测和估计，为满足多视角要求，添加关键点约束和mocap约束





以上是动作捕捉的相关过程，接下来是动作合成。框架的主要目的是使得神经动物可以与使用者产生实时交互，因而根据使用者的命令合成相关动作显得至关重要，文章采用由门控网络和运动预测网络组成的LMP框架，如图2.3所示

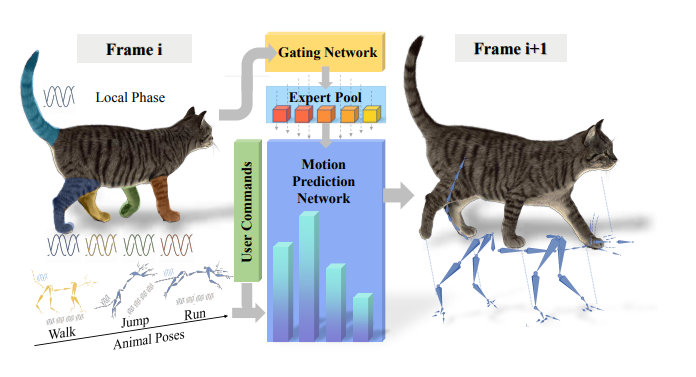


图2.3 LMP框架

门控网络将当前帧和过去帧的运动状态（关节位置、旋转和速度）作为输入，进入专家池，然后将专家池结果与运动状态动态混合并发送到运动预测网络。利用来自用户的专家权重和控制信号，运动预测网络将能够计算未来帧的运动状态。该过程可以表示为：



不过LMP产生的动作的神经动物模型由相关模型师创作，该模型与拍摄时的动物模型并不匹配，因此要根据正向运动学和反向运动学计算目标状态，对于不可能的姿势进行相关的细微调整。这些经过变换和调整的数据可以把运动状态从一个神经动物转移到另一个神经动物。

至此，关于该系统的两个模块的陈述基本完成，论文后续对整个流程进行了简单总结并阐述了该框架在消费级VR眼镜中的应用

# 3 结果分析

文章通过**ARTEMIS**系统对多毛动物进行了仿真模拟，进行模拟后文章也与其他系统进行了横向对比，对比结果如下图所示

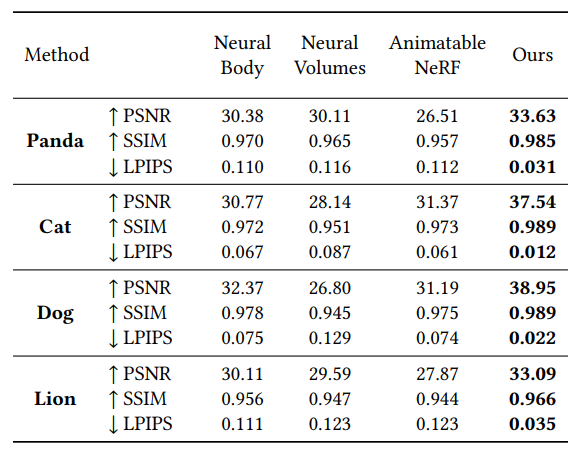


图 3.1 动态神经网络渲染横向对比

表中采用了峰值信噪比（PSNR），结构相似性指数（SSIM）和学习感知图像补丁相似性（LPIPS）作为评估渲染的指标，从指标来看，该系统确实在保留渲染细节方面具有一定优势

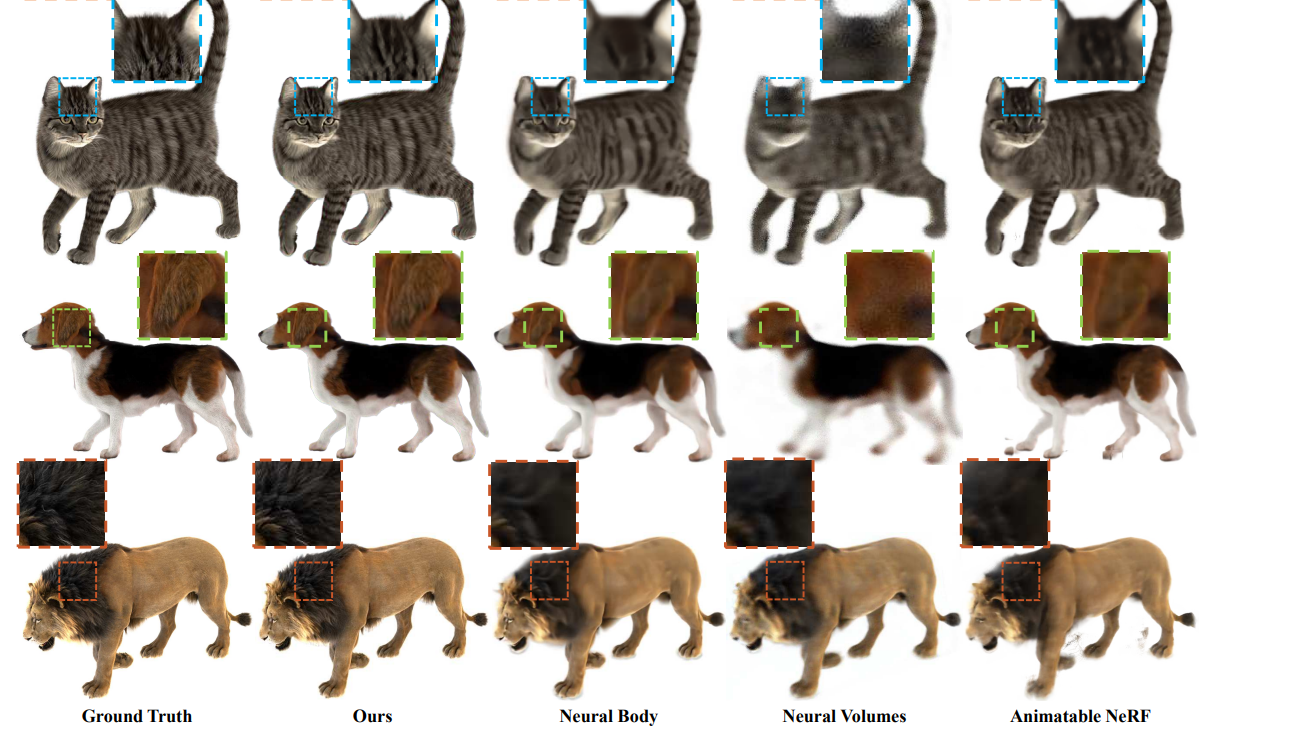


图3.2 动态渲染横向对比

除表格对比外，文章也提供了实际渲染结果的横向对比，第二列为文章所用渲染框架的渲染结果，图片中文章所采用的框架的猫的前额毛发的仿真细节明显优于右侧三个系统的猫的前额毛发细节，因此文章结果具有一定程度的说服力。

上述部分为动态渲染时会采用的神经网络，文章还与其他静态渲染在拟真程度方面进行了相关比较

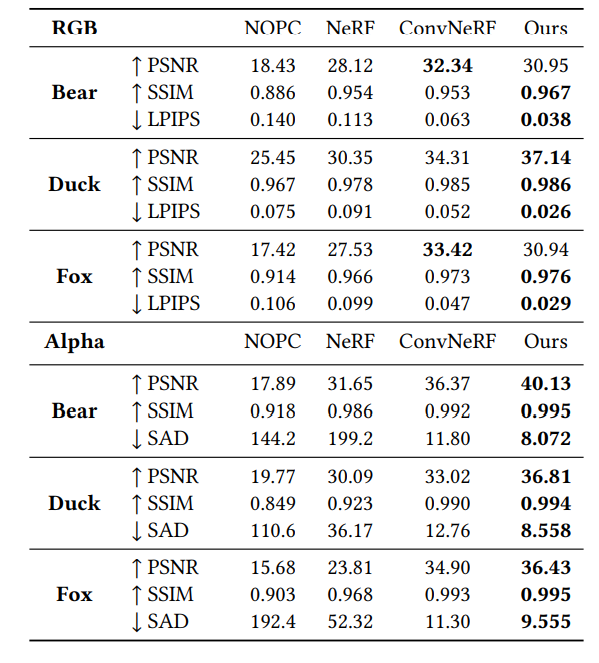


图3.3 静态RGBA渲染横向比较

RGB方面采用的评价参数与动态渲染相同，在不透明度方面采用了SAD（Sum of Absolute Distance）绝对距离之和作为评估参数。上图显示了该系统对于高质量毛发和不透明度渲染的有效性。

整体而言该系统确实在渲染质量和渲染实时性上都取得了一定程度的成果。

# 4 局限性分析

文章展示了该系统交互式运动控制，实时动画和神经生成（NGI）多毛动物的强大渲染能力，且该系统也可应用于VR设备。但该系统仍然存在一定的局限性，首先，该系统生成的动物严重依赖预定义的骨骼数据和相应CGI的蒙皮权重数据，文章相关作者已经考虑在未来结合非刚性跟踪技术减少对骨骼数据的依赖。其次，动作捕捉动物获得的相关数据仍然需要相关的3D动作模型师对动作数据进行相关处理和调整，无法完全自动生成神经动物。此外，当前的NGI动物只能在渲染CGI资产时已经烘焙过的照明环境中生成结果，因此与传统渲染引擎相比，无法在新环境中快速采用照明。从ARTEMIS系统方面来看，使用者目前与NGI动物的互动仍然受到严重依赖人类动作捕捉和预定义的基于规则的策略的限制。

但文章也已经对上述问题阐述了相关原因和可考虑的解决方案，该系统的改进方向整体上也较为明确。

# 5 参考文献

[1] Haimin Luo, Teng Xu, Yuheng Jiang, Chenglin Zhou, Qiwei Qiu, Yingliang Zhang, Wei Yang, Lan Xu, and Jingyi Yu. 2022. Artemis: articulated neural pets with appearance and motion synthesis. ACM Trans. Graph. 41, 4, Article 164 (July 2022), 19 pages. https://doi.org/10.1145/3528223.3530086