

FoVolNet: Fast Volume Rendering using Foveated Deep Neural Networks

Dezeming Family

2023年07月18日

DezemingFamily系列文章和电子书**全部都有免费公开的电子版**，可以很方便地进行修改和重新发布。如果您获得了DezemingFamily的系列电子书，可以从我们的网站[<https://dezeming.top/>]找到最新的版本。对文章的内容建议和出现的错误也欢迎在网站留言。

目录

一 Introduction	1
二 RELATED WORK	1
2 1 Advanced Volume Rendering	1
2 2 Foveated Rendering	1
2 3 Deep Learning for Image Denoising	2
2 4 Optical Flow	2
三 METHODS	3
3 1 Foveated Rendering	3
3 1.1 直接采样	3
3 1.2 直接采样	3
3 1.3 Stream Compaction	3
3 2 Reconstruction Network	4
四 结论	4
参考文献	4

abstract

我们介绍了FoVolNet，这是一种显著提高体数据可视化性能的方法。我们开发了一种具有成本效益的中央凹渲染管线，该管道对焦点周围的体进行稀疏采样，并使用深度神经网络重建整个帧。

Foveated渲染是一种围绕用户焦点对渲染计算进行优先排序的技术。这种方法利用了人类视觉系统的特性，从而在用户视野的外围渲染数据时节省了计算资源。我们的重建网络结合了直接预测和核预测方法，以产生快速、稳定和令人信服的输出。凭借纤薄的设计和量化的使用(slim design and the use of quantization)，我们的方法在端到端帧时间和视觉质量方面都优于最先进的神经重建技术。

我们对系统的渲染性能、推理速度和感知特性进行了广泛的评估，并与竞争对手的神经图像重建技术进行了比较。我们的测试结果表明，FoVolNet在保持感知质量的同时，始终比传统渲染节省了大量时间。

— Introduction

例如，先前的工作提供了渲染不适合系统主内存的超大体积的解决方案。通过引入流技术（如核心外(out-of-core)渲染），这些数据的可视化变得可行，这消除了整个数据集存在于内存中的需要。这种方法有助于我们减少对数据特定特征的依赖（在这种情况下，是数据的大小）。

在视觉质量方面，最先进的渲染技术缺乏类似的手段。一种技术产生的视觉质量越高，通常需要更多的计算资源来计算它。尽管多年来正在进行的研究已经产生了更有效的方法，但我们仍然受到光线数量、采样率或照明类型等因素的限制。因此，在交互式帧率下使用高质量着色技术可视化体积数据仍然具有挑战性，尤其是对于沉浸式可视化等要求苛刻的应用程序。

这项工作揭示了渲染技术和计算成本之间的紧密耦合。我们介绍了FoVolNet——一个完整的体渲染管线，旨在放松技术和成本之间的关系。通过跳过大多数屏幕空间像素处理，代之以恒定时间的神经图像重建，我们可以在不牺牲视觉质量的情况下大幅提高性能。

我们的灵感来自于关于foveated渲染和基于深度学习的图像去噪的文献。先前的工作已经表明，在渲染数据时考虑人类视觉系统（HVS）可以在没有明显质量损失的情况下产生出色的性能结果。利用HVS的特性是我们设计的关键部分，因为它可以让我们集中计算资源。

因此，FoVolNet渲染具有密集凹陷中央凹区域的稀疏图像。神经图像重建网络可以恢复丢失的视觉信息，使我们能够跳过屏幕空间的大部分像素处理，并用恒定的时间推断步骤代替它。这使得以比传统渲染方法低得多的计算成本以高质量体可视化成为可能。反过来，这种解耦使我们能够在各种渲染设置中实现更快、更一致的帧速率。我们进行了全面的测试，包括系统的整体渲染性能、图像质量和其他特性，如有效压缩率，以评估我们的方法。结果表明，FoVolNet以传统方法产生相同输出所需时间的一小部分忠实地重建全帧。

二 RELATED WORK

我们的工作涉及体渲染、光流估计以及计算机图形和图像处理中的深度学习方法。在本节中，我们将讨论这些领域的相关工作。

2.1 Advanced Volume Rendering

2.2 Foveated Rendering

眼睛跟踪技术的最新进展以及市场对增强现实和虚拟现实应用的推动，加强了对中心凹渲染技术的研究。计算能力对于高保真渲染应用程序和当今大多数沉浸式内容至关重要。因此，有效分配资源至关重要。

人类视觉系统感知高细节的能力仅限于相对较小的焦点区域。中央凹是视野中视力最高的区域，仅占眼睛光轴周围约5.2度。Foveated渲染方法通过将计算资源集中在这些区域来利用这一事实。Guenther等

人是最早开发这种方法的人之一，通过在焦点周围的同心圆中渲染多个细节级别的场景。后来的方法使用可变采样率来确定重点区域的优先级。Weier等人将这种方法与帧重投影相结合，以减少外围闪烁。

2.3 Deep Learning for Image Denoising

深度学习在计算机图形学领域的一个突出用途是图像去噪。它是细化噪声图像的过程，噪声图像通常是每像素采样数较低的蒙特卡罗（MC）渲染的结果。提高图像质量的一种基本方法是增加SPP。然而，这种方法每帧需要更长的处理时间。最近的工作利用深度学习在没有这种计算开销的情况下细化低SPP图像。

早期的方法已经使用卷积神经网络取得了令人印象深刻的结果。这些技术确立了当今图像去噪的两种基本原理。一方面是直接预测，作为网络推理的直接结果产生去噪图像的一种方法。另一方面，核预测方法使用卷积神经网络来产生图像滤波器。通过在单独的步骤中将这些滤波器应用于输入图像来执行去噪操作。

随后的工作紧随其后，进一步挖掘了这两个概念的潜力。Wong等人为直接预测网络引入残差连接，以提高单帧图像质量。为此，Xu等人 and Lu等人进行了使用对抗性网络训练直接预测模型的实验。最近，Hofmann等人将直接预测应用于体路径跟踪领域。类似地，Weiss等人探索了直接预测在重建自适应体ray marching中的效用。除了Kettunen等人的梯度空间去噪和Wong等人的ResNet方法等工作外，他们还研究了各种辅助输入特征对最终图像质量的影响。遵循核预测技术，Vogels等人的工作通过将相邻帧合并到训练中来扩展该方法，以促进时序稳定性。Hasselgren等人基于这一概念，使用预测自适应采样和时间混合创建时间稳定的图像序列。Gharbi等人使用带有飞溅的内核预测方法来解决涉及运动模糊和景深的问题。

这些项目中使用的神经结构各不相同。然而，也存在一些可识别的趋势。U-Net架构最初是为医学图像分割任务开发的，已被证明是去噪任务的实用选择。许多作品的设计基于U-Net的图像编码器-解码器原理。扩展通常包括跳过连接和循环反馈，这往往会提高图像质量和时间稳定性。其他工作使用更传统的CNN或RNN模型。有趣的是，所选择的架构（U-Net、CNN、RNN）和去噪方法（直接预测、内核预测）之间没有明显的联系。一些作品也使用了一个额外的评估网络来进行对抗性训练。W-Net以其形状命名，是U-Net的延伸，由Thomas等人引入，它按顺序包括两个U-Nets，一个用于特征提取，而另一个用于生成卷积滤波器并将其应用于输入。这种设计便于通过选择性量化进行优化，而没有显著的图像质量损失。

为了进一步阅读该主题，我们请感兴趣的读者参考霍等人关于基于深度学习的图像去噪技术的综述。我们的网络设计基于W-Net架构。我们引入了一种结合直接预测和核预测的混合方法，以实现稀疏图像输入的最佳结果。这与传统的图像去噪方法的不同之处在于，缺失的视觉信息需要由网络生成。因此，纯内核预测网络不适合此任务，因为内核仅对现有像素值进行操作。DeepFovea[3]是目前仅使用直接预测进行稀疏帧重建的最先进技术。我们的方法将这一最初的想法转化为科学渲染领域，并显著提高了视觉质量和性能。

2.4 Optical Flow

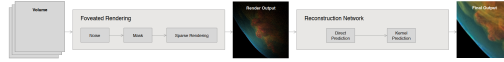
视频序列中感知到的运动是由场景中对象位置的增量变化或相机移动引起的。估计相邻帧中元素的光流是一个活跃的研究领域，近年来已经提出了各种方法。

Farneback等人的早期方法引入了一种运动估计算法，该算法将像素邻域表征为多项式，并使用这些多项式来寻找帧之间的映射。他们提出了一种使用先验运动估计的多尺度方法。这允许算法通过考虑不同大小的搜索窗口来迭代和细化估计。这提高了结果的稳健性和质量。随后的工作倾向于模仿这种迭代的、分层的方法。最近，Hanika等人介绍了一种基于该方法的方法。与以前的方法不同，这种算法为了速度而牺牲了一些质量。它还优雅地处理掉了遮挡。

在我们的工作中，我们利用Hanika等人的方法在训练期间重新投影帧。通过warp前一帧，我们可以获得更多关于当前图像的视觉信息，这些信息可以用于构建损失函数，该函数需要网络匹配重新投影的帧。该附加信息有助于提高图像质量并支持保持帧之间的时序稳定性。

三 METHODS

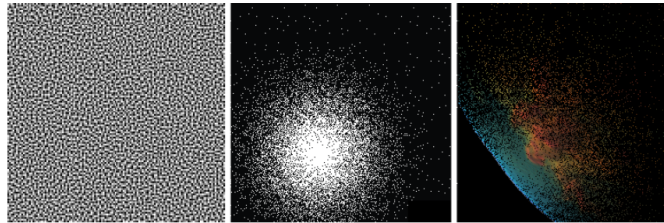
FoVolNet是一个完整的光线行进体渲染管线，由神经网络支持（图2）。整个渲染过程由两个关键步骤组成。首先，需要渲染体。我们不渲染整个帧，而是选择性地渲染像素的子集。然后使用神经网络来从该子集重建整个帧，以下部分详细介绍了我们的方法。在此，我们讨论了FoVolNet各个阶段的实现和设计。



该渲染器支持全局ray marched阴影。对于阴影计算，我们每个采样step投射一条阴影光线，使用主采样率的1/4。

3.1 Foveated Rendering

此渲染器使我们能够随着整体像素密度的降低而减少渲染时间。我们的foveated渲染技术基于从噪声模式生成的二进制采样图，噪声模式决定了体渲染器应该对屏幕空间中的哪些像素进行采样。也就是说，我们的采样密度与之前的中央凹渲染不同，而是在像素上随机，根据噪声来进行选择去渲染哪些像素



我们使用时空蓝噪声（STBN）来生成时间稳定的样本模式，同时保留传统蓝噪声的感知优势（论文图3的(c)和(f)）。与独立的2D蓝噪声或3D蓝噪声模式的序列不同，STBN不仅在空间域中是blue的，而且随着时间的推移，每个像素都是blue的。这种性质是可取的，因为它有助于我们的重建网络产生稳定和感知上干净的图像序列。

噪声图是可以被无缝平铺，以覆盖任意大尺寸的图像。我们试过 16×16 以及 256×256 ，没什么差别。比如我们要生成 1024×1024 的图像，就把噪声图平铺到 1024×1024 。

为了生成一个采样Mask。需要假设一个阈值，如果噪声图上某噪声值小于阈值，就设mask在该像素为0，否则为1。不过这个阈值是可以根据空间来变化的，比如中央凹区域比较下，周围区域比较大。

3.1.1 直接采样

在OptiX和CUDA中渲染器的简单实现中，无效像素将在内核级别上被丢弃。然而，每个全尺寸帧缓冲区像素使用一个内核线程只会产生微不足道的性能增益。这是因为GPU内核调用按warps分组。只有在warps中的所有线程结束后，才能获得结果。我们开发了两种方法来规避这个问题。

3.1.2 直接采样

直接采样就是渲染时根据一个基于采样概率的函数 P 来生成像素位置。启动内核只在一片小区域，但是映射到全分辨率。就管线集成而言，直接采样相对不具侵入性。我们不是在整个图像维度上启动CUDA内核，而是在较小的缓冲区内调用它。当访问每个GPU线程中的屏幕空间坐标时，函数 P 充当代理。这种方法的缺点是 P 不能可靠地产生唯一的采样位置。尽管直接采样的性能比简单的方法要好，但我们经常会遇到重复的坐标，这会使渲染器重新计算现有的采样，从而丧失潜在的压缩能力。

3.1.3 Stream Compaction

流压缩就是先生成采样Mask，然后压缩到更小frame，然后渲染，然后再恢复到原始frame尺寸坐标上。

3.2 Reconstruction Network

整体设计：我们设计的核心思想是将重建过程分为两个步骤。两个步骤是通过按顺序运行两个网络来实现的（论文图7）。这种混合体系结构同时采用了直接预测和内核预测。在没有初始重建步骤的情况下，由于缺乏从中提取视觉信息的丰富像素邻域，内核预测方法无法执行。另一方面，如我们在评估中所示，仅执行直接预测步骤将导致低于标准的图像质量。

直接预测部分：网络 D （论文图7）使用直接预测重建图像。也就是说，其输出 O_d 可以直接解释为图像。该步骤重建帧的整体特征，并填充输入中有效像素之间的空白点。除此之外，我们保留解码器的隐藏状态 H_d 以供进一步处理。

核预测部分：如果没有直接预测作为输入，而是直接把渲染图像作为输入，那么结果就不会很好，所以两个步骤都是需要的。

在网络的多个部分引入循环连接来积累状态。聚集的信息有助于重建时间稳定的图像序列。网络 D 中的解码器块由循环连接来连接，该循环连接在下一训练步骤中将块的输出隐藏状态传递回其输入。在更广泛的范围内，网络 D 的输出 O_d 被传递回其输入层，作为后续运行的数据的一部分。使用沿着信道维度的级联操作来组合当前状态和循环状态，并应用适当的上采样或下采样以使所有输入与后续操作兼容。

四 结论

因为论文很简单，所以就暂时不再介绍了。

参考文献

- [1] Bauer D, Wu Q, Ma K L. Fovolnet: Fast volume rendering using foveated deep neural networks[J]. IEEE Transactions on Visualization and Computer Graphics, 2022, 29(1): 515-525.
- [2] Kaplanyan A S, Sochenov A, Leimkühler T, et al. DeepFovea: Neural reconstruction for foveated rendering and video compression using learned statistics of natural videos[J]. ACM Transactions on Graphics (TOG), 2019, 38(6): 1-13.
- [3] DeepFovea: neural reconstruction for foveated rendering and video compression using learned statistics of natural videos