

**研究生读书报告**

**阅读论文** **Neural Supersampling for**

**Real-time Rendering**

**姓名与学号** 钱佳乐 22251186

**指导教师**  李启雷

**年级与专业**  2022级软件工程研究生

**所在学院**  软件学院

**论文出处**  2020 SIGGRAPH

**提交日期**  2022.12.31

一、介绍

由于近年来增长的更高的分辨率和刷新率，以及在物理着色和实时光追领域的进步，实时渲染对于游戏和新兴的虚拟现实头盔来说已经变得越来越具有挑战性。为了满足这一需求，现代图形硬件和游戏引擎通常渲染低分辨率图像，然后通过上采样到原始分辨率以减少计算消耗。

现存用来解决该问题的大量技术，例如虚幻引擎利用像素颜色统计和时域积累的TAAU，微软通过降低高分辨率显示的着色复杂度来加速渲染的可变速着色，以及最接近本篇论文的英伟达的DLSS，它用神经网络实时上采样低分辨率的渲染内容。然而这些方法要么在上采样图中引入了明显的视觉伪影，要么依赖于专有技术和特定硬件。

在这篇论文中作者提出了专为实时应用中渲染内容的高质量上采样而定制的机器学习方法。它采用了丰富的渲染属性（即颜色、深度和运动矢量），并为实时应用而优化。该方法很容易整合到现代游戏引擎中而不需要特定的硬软件，使其能够广泛被采用，支持最高4×4上采样，能够实时生成高保真度和时间上稳定的结果，并在视觉保真度和图像质量的定量指标方面都明显优于先前的工作。

**二、相关工作**

2.1实时渲染

由于渲染图像的样本在时间和空间上是点样本，而摄影图像的样本则是区域积分，所以渲染图像往往高度混叠。这使得渲染图像的上采样既是一个抗混叠又是一个插值的问题，而不是计算机视觉中的去模糊问题。

实时渲染中一个重要问题是欠采样，欠采样的图像表现出大量混叠，当前主流方法分为使用单一渲染图像信息的纯空间抗锯齿方法，与使用多个渲染历史帧的时间抗锯齿方法。

空间抗锯齿有经典的MSAA、纹理过滤，除此之外大部分空间抗锯齿方法基于图像处理中的图像增强方法，找到图像中的不连续，然后模糊它们一次平滑锯齿边缘，如MLAA、FXAA、SMAA等。虽然这些方法可以为单个静止图像提供良好的质量，但用这些方法抗锯齿的帧之间的时间变化仍然包含可见的闪烁和其他类型的虚假像素运动。

时间抗锯齿使用以时间累积的缓冲区，使用某种形式的时间拒绝过滤来抗混叠。经典方法如分时渐进超采样将历史记录从一帧重投到另一帧以补偿运动；TAA使用边缘检测滤波器作为代理，通过较重的时间累积来抑制闪烁现象。最近TAA也被用于执行时间上采样，本文给出了与其的对比结果。DLSS与本文方法最为接近，它使用时间历史和神经网络来增强边缘和执行上采样。

除此之外，与本文的上采样插值问题任务有关的工作是最近出现将机器学习方法应用于实时低样本数重构和凹陷重构。

2.1图像和视频超分辨率

单一图像超分辨率(SISR)旨在从其低分辨率，通常是退化的版本中恢复高分辨率图像。在SRCNN使用三层CNN进行SISR后，诸多DL方法被提出，如VDSR学习高低分辨率图像之间的残差、SRResNet引入残差网络结构、EDSR使用更多残差块提高性能、LapSRN提出用于逐步重构高分辨率图像的子带残差的拉普拉斯金字塔网络等等。

视频超分辨率方法一般在SISR方法上利用相邻帧的时间相关性来改善重构。大部分方法的关键部分是帧之间的运动估计，如VESPCN引入了一个多分辨率空间变换器模块，用于联合运动补偿和视频超分辨率、SPMCVSR引入了一个子像素运动补偿层来融合多个帧以揭示图像细节。另一种方法是使用RNN，自然激励时间上的一致性结果，如FRVSR提出了扭曲之前估计的帧以方便后续帧的RNN、RBPN开发了一个递归的编码器-解码器架构，用于纳入从单幅图像和多帧模块中提取的特征。

**三、算法**

3.1核心思想

首要的问题是点采样的图像在阴影和几何边缘具有很多锯齿，而要插值的目标像素的信息则完全丢失。论文利用渲染中当前和过去帧的已知详细信息（像素颜色、深度图和运动矢量）以及采样方式解决该问题。

算法为了有效利用渲染数据的特性，首先通过零上采样(zero-upsampling)将输入投射到高分辨率空间，然后再对过去帧扭曲，使其与当前帧对齐以减少所需的感受野和重构网络的复杂度。

值得注意的是，运动矢量定义为当前帧中可见的三维点在前一帧中的屏幕空间位置。它虽然以低成本提供了像素间的候选匹配，但具有三点限制。首先，由于它向后映射，所以前一帧的像素不能直接投射到当前帧。其次，它不考虑相邻帧之间的动态遮挡，当前帧可见的点在前一帧可能被遮挡。最后，它只提供了表面点和相机的运动，而没有考虑光照、阴影、视相关反射等。

由于运动矢量的限制，扭曲的过去帧会包含与当前帧不匹配的无效像素，而这会误导后期重构。为了解决该问题，论文在重建网络前加入了重加权机制以去除无效像素，并且与现有方法相比，权重将由神经网络学习。

最后，预处理过的前几帧（经过零上采样、扭曲和重加权）与当前帧（经过零上采样）堆叠在一起，并送入重建网络以生成所需的高分辨率图像。

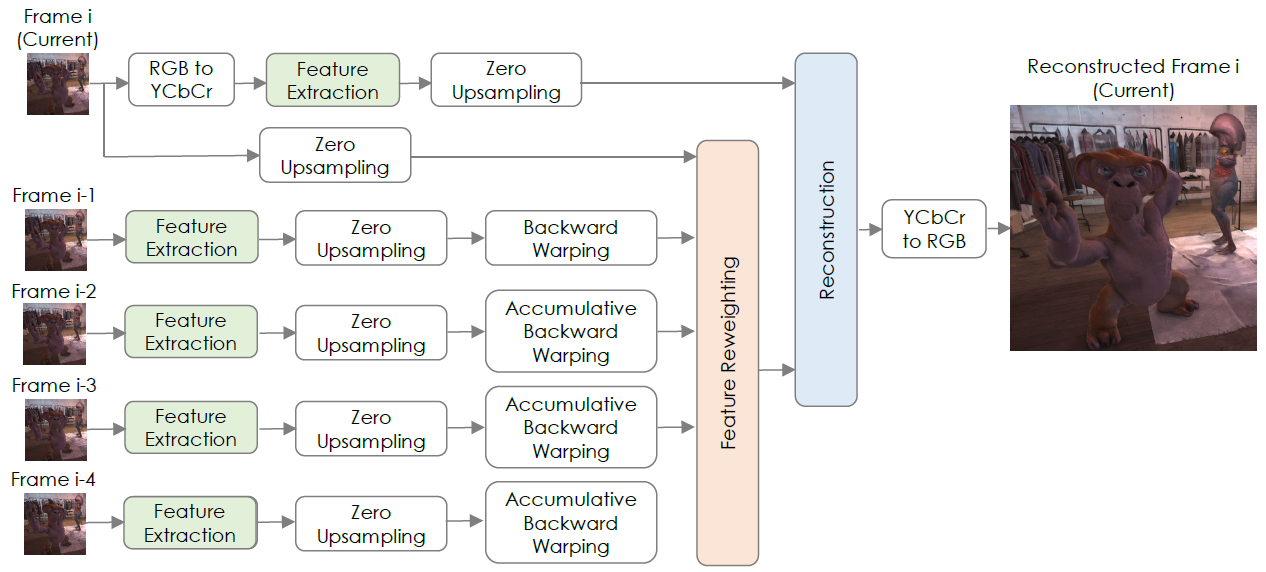
算法流程如下图3-1所示：

图3-1 算法流程

3.2网络结构

3.2.1 特征提取

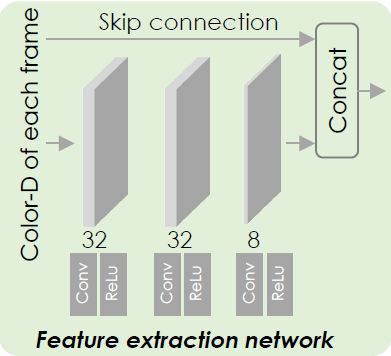
特征提取模块是一个三层CNN。该子网络单独处理每个输入帧，并在除当前帧外的所有帧中共享权重。对于每一帧，将RGB-D作为输入，产生8通道的学习特征，然后将学习特征与输入相连接，总共产生12通道的特征。网络结构如3-2所示：

图3-2 特征提取网络

3.2.2 时序重映射

时序重映射使用运动矢量将像素样本和之前每一帧的学习特征投射到当前帧。为了充分利用运动矢量，算法在时序重映射前使用零上采样将像素样本从低分辨率投射到高分辨率，在高分辨率空间执行扭曲能够减少过程中低通插值的影响，保护渲染的点样本中包含的高频信息。零上采样把每个输入像素分配给其在高分辨率下的相应像素，并将其周围的所有缺失像素置零。零上采样不仅效率高，而且提供了哪些样本是有效的信息。

对于低分辨率的运动矢量图，算法利用其分段平滑的特征，通过双线性上采样将其映射到高分辨率空间。然后使用上采样后的运动矢量对零上采样后的帧进行扭曲以对齐当前帧。因为运动矢量是一组相邻帧下的定义，所以为了扭曲多个过去帧，论文迭代的应用扭曲处理直到每个过去帧都被扭曲到了当前帧上，本次实验中过去帧数量至多为4。

3.2.3 特征重加权

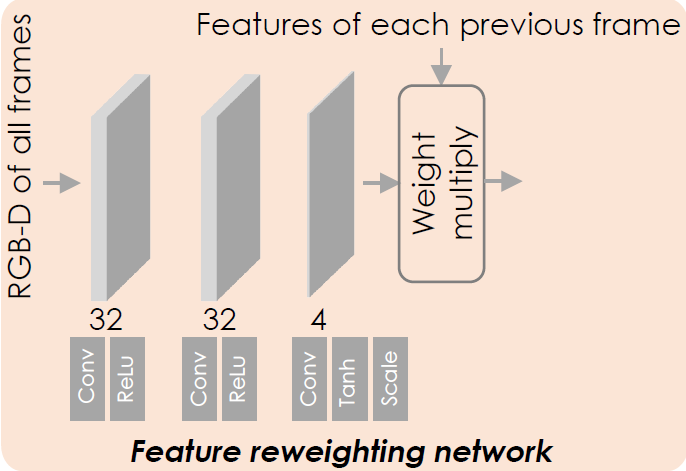
如3.1中所述，由于运动矢量不反映帧之间的动态遮挡或阴影变化，所以被扭曲的帧会包含无效像素。特征重加权网络就是为解决该问题而提出的。它是一个三层CNN，将零上采样后的当前帧的RGB-D以及零上采样、扭曲后的过去帧作为输入，为每个过去帧生成一个像素级的加权图，其值在0到10之间，10为根据经验得出的超参数。加权图将与对应过去帧的特征做乘积以得到重加权特征。网络结构如3-2所示：

图3-3 特征重加权网络

3.2.4 重建

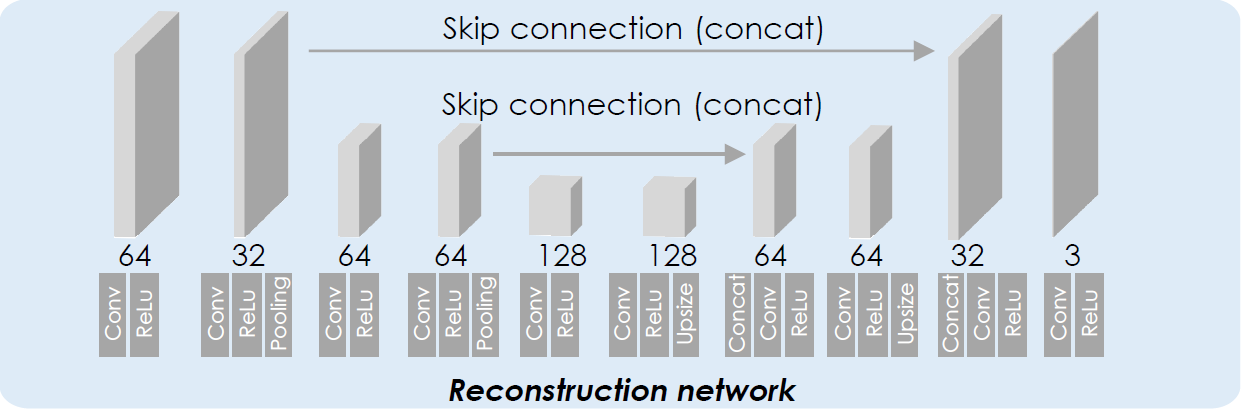
重建网络是一个3级10层且具有跳跃连接的U-Net，将当前帧的特征和过去帧的重加权特征拼接起来作为输入，输出当前帧修复后的高分辨率图像。重建网络如下图3-4所示：

图3-4 重建网络

3.3损失函数

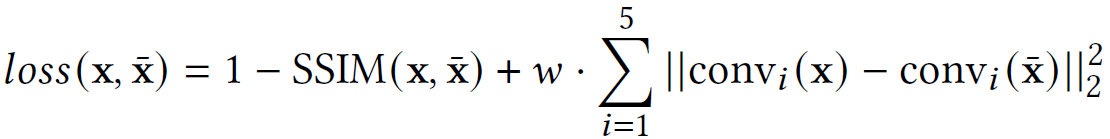
损失函数是是预训练VGG-16网络计算的感知损失和结构相似性指数(SSIM)的加权组合，如下图3-5所示

图3-5 损失函数

3.4数据集与实现

论文收集了Unity中有代表性的动态场景，构建了一个大规模数据集生成程序来渲染训练和测试数据。该程序回放从VR头盔的用户研究中捕获的头部运动，并渲染每一帧的颜色、深度和运动矢量。每个场景渲染100个包含60帧的视频。每个视频的相机以随机位置开始，并按照预先捕获的头部运动路径移动，该路径从一个大型候选池中随机选择。

对于参考图像，论文首先使用8倍MSAA将图像渲染为4800×2700，然后使用3×3箱盒滤波器将图像缩小到1600×900，以进一步减少混叠。对于低分辨率输入图像，关闭MSAA并调整纹理采样的mip级别偏差，以将选定的mip级别与全分辨率图像相匹配。

**四、结果分析**

4.1运行时性能

论文给出了不同目标分辨率下的耗时，最大的1080p也只耗时24.42ms。并且为了了解网络复杂度和输出图像之间的权衡关系，论文还训练了一个减少重建网络中间层50%参数的轻型网络。轻型网络相比于主网络在质量上有轻微下降，时间上却减少25%。

4.2质量评估

论文对比了数种既有方法，包括SISR以及视频超分辨率方法，论文在相同数据集下重现并以相同方法训练所有算法，并适当调整算法以保持一致性。评比指标为峰值信噪比(PSNR)、结构相似性(SSIM)以及时空熵差(STRRED)。在10个测试视频的平均结果下，论文算法的表现要远远优于既有方法。

4.3分析

从渲染效率看，使用Titan RTX显卡渲染有高质量光追全局光照效果的1600×900分辨率图像耗时140.6ms，而渲染400×225分辨率图像只需要26.4ms，再将其用4\*4上采样到原始分辨率耗时17.68ms，总计44.08ms。这之间有三倍多的渲染性能改进，并且结果能保持高保真度。

从泛化角度看，论文对比了专门为一个场景训练的网络与联合训练于四个场景的网络，结果显示在PSNR上有0.05-0.4dB的下降，但依然优于其它方法专门训练于一个场景的结果，证明网络在不同场景间具有一定泛化能力。但由于联合训练的场景只有四个，所以还需要更多测试佐证。

从过去帧的数量看，论文测试了从1到4的过去帧数量，结果显示随着使用的过去帧增多，图像质量有所增加，但会轻微增加运行时间。

从超采样比例看，论文保持目标分辨率不变，改变输入分辨率，测试了从2×2到6×6的比例，结果显示低比例有着更好的重建质量。

从附加输入看，由于论文算法远远超出对比方法，所以作者试图理解算法中附加的深度和运动矢量输入对于图像质量有何影响。因此论文修改了VESPCN算法，将运动矢量取代了原始算法中的光流估计模块，将深度信息附加在颜色输入上，并命名为VESPCN+。结果显示VESPCN+在PSNR上有1.1-1.3dB的提升，但依然比论文算法低2.2-3.1dB。这说明附加输入与专门定制的网络设计都有重要作用。

从零上采样和扭曲看，论文为了验证零上采样和在目标分辨率空间扭曲的效果，设计了用双线性上采样取代零上采样以及在输入分辨率空间扭曲的消融实验。结果显示，在目标空间扭曲比在输入空间扭曲在PSNR上提升了1dB，零上采样比双线性上采样在PSNR上提升了0.3dB。这说明论文算法可以更有效利用特定于渲染的输入。

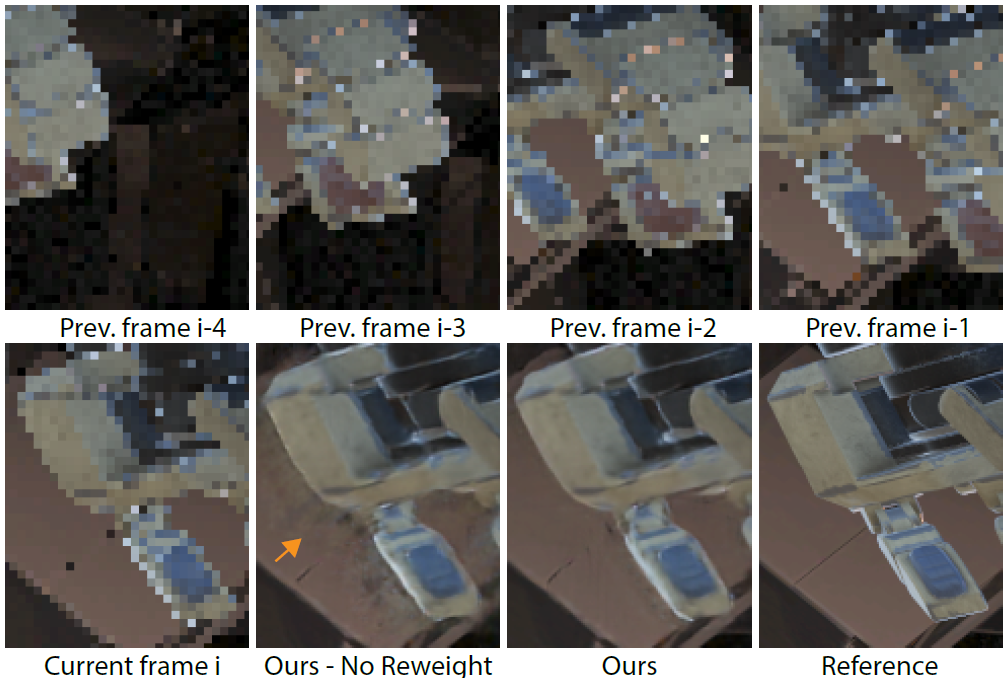
从特征提取和特征重加权模块看，论文为了分析这两个模块的贡献，设计了去除这些模块的消融实验。结果显示，在数值上只有微小下降，特别是特征重加权模块。但是这并不意味着该模块没用，反而是由于受该模块影响的图像区域（受去遮挡和不匹配像素影响的区域）占比相对较小，所以在数值上没有太大的波动，但是从视觉图像上却可以看出特征重加权模块的重要作用，如下图4-1所示，没有该模块的结果鬼影显著。

图4-1 特征重加权模块的消融实验结果

**五、总结**

论文算法学习了低分辨率多帧输入和高分辨率输出图像之间的映射，使用了图像对图像的损失，试图尽可能地匹配每一帧，在时间稳定性方面有了显著的改善。输出图像质量远远超出现有方法，在训练域以外图像质量不会急剧下降，并且实时性能良好，开放的网络设计也为后续图形学上的神经超采样工作铺平了道路，提供新的思路。

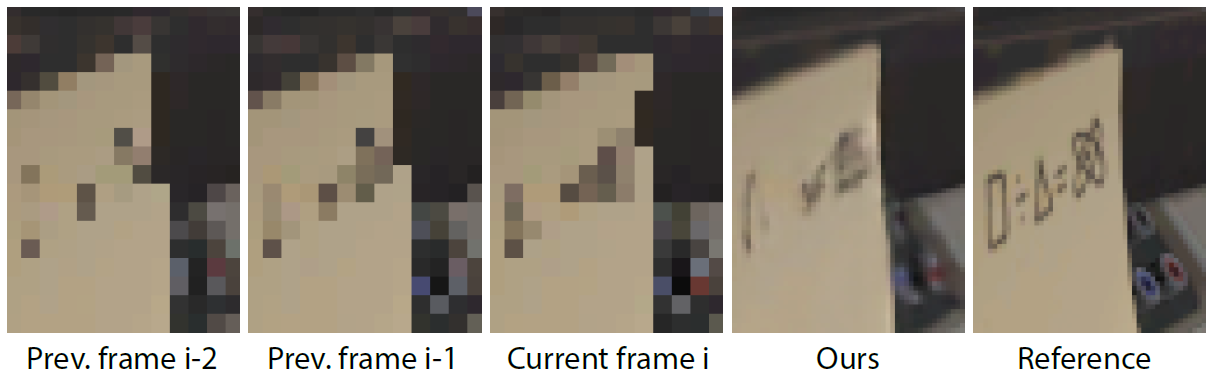
但目前算法依然具有限制。如果多帧输入由于极度混叠而包含一个场景区域的太少信息，算法就会失效。如下图5-1所示：

图5-1 失败案例

此外，算法虽然为神经超采样提供了重要改进，但当前实现在高分辨率显示应用上依旧代价高昂，并且实时性对于硬件要求极高。不过通过进一步的网络优化、硬件加速有望来加快计算。