E-mail: jig@aircas.ac.cn
Website: www.cjig.cn
Tel: 010-58887035
中国图象图形学报
JOURNAL OF IMAGE AND GRAPHICS

中图法分类号:TP391　　**文献标识码**: A　　**文章编号**: 1006-8961(2026)03-0880-16

论文引用格式:Li L, Xue H W, Zhu J C, Zhao Y and Liu X P. 2026. Real-time neural super-sampling rendering based on frame-recurrent structure. Journal of Image and Graphics, 31(3):0880-0895(李琳, 薛皓文, 朱纪春, 赵洋, 刘晓平. 2026. 帧循环结构的实时神经超采样渲染. 中国图象图形学报, 31(3):0880-0895)[DOI:10.11834/jig.250296]

帧循环结构的实时神经超采样渲染

李琳[1,2],薛皓文[1],朱纪春[1],赵洋[1*],刘晓平[1,2]

1. 合肥工业大学计算机与信息学院,合肥　230009; 2. 安全关键工业测控技术教育部工程研究中心,合肥　230009

摘　要:**目的**　实时渲染图形程序(如游戏、虚拟现实等)对高分辨率和高刷新率的要求越来越高,因此,针对渲染图像的实时超分辨率技术在实时渲染中非常必要。然而,现有的视频超分算法和实时渲染处于不同的数据处理管线之中,这导致其难以直接应用到实时渲染管线里。**方法**　对此,提出了一个基于帧循环结构的实时神经超采样方法。充分利用实时渲染管线中生成的低分辨场景几何数据,以提升超采样网络对于三维空间信息的感知力;将帧循环框架结合到超采样方法中,通过引入先前帧重建结果的特征来改善当前帧的重建结果,从而实现时间尺度上的稳定性;将重加权网络和注意力网络置于特征提取模块中,以提升提取到的特征的有效性。此外,本文还提出了一个面向神经超采样的实时渲染流程,该流程能够将超采样网络部署至图形计算管线之上,并与实时渲染管线相结合。**结果**　与同样能够实时且效果较好的基准方法面向实时渲染的神经超采样(neural super-sampling for real-time rendering, NSRR)比较,本文方法在速度少许提升的前提下,图像质量指标峰值信噪比(peak signal to noise ratio, PSNR)平均提升了0.4 dB,并在部署到实时渲染管线后,通过轻量化裁剪继续保持实时性且部分场景效果仍然优于非实时的部署后NSRR;在网络模块的消融实验中也证明了各个子模块对于神经超采样任务的有效性。**结论**　本文提出的神经超采样网络模型与搭建的神经超采样渲染流程,在取得更好效果的同时具有一定的实用价值。

关键词:实时渲染;帧循环神经网络;超采样;超分辨率(SR);卷积神经网络(CNN)

Real-time neural super-sampling rendering based on frame-recurrent structure

Li Lin[1,2], Xue Haowen[1], Zhu Jichun[1], Zhao Yang[1*], Liu Xiaoping[1,2]

1. School of Computer and Information, Hefei University of Technology, Hefei　230009, China; 2. Engineering Research Center of Safety Critical Industrial Measurement and Control Technology, Ministry of Education, Hefei　230009, China

Abstract: **Objective** The complexity of modern real-time rendering algorithms, such as ray tracing, is often closely related to the number of pixels on the target screen. However, with the increasing resolution and refresh rates of external displays and head-mounted displays, as well as the proposal of graphics algorithms that achieve more realistic visual effects, the computational cost of real-time rendering programs has increased. As a result, many users must balance rendering resolution, real-time performance, and rendering quality, even with the support of new parallel computing hardware. Although many video super-resolution methods now produce excellent reconstruction results, these methods cannot be

收稿日期:2025-07-14;**修回日期**:2025-09-16;**预印本日期**:2025-09-23

*通信作者:赵洋　yzhao@hfut.edu.cn

基金项目:国家自然科学基金项目(62277014,62272142);中央高校基本科研业务费专项资金项目(PA2025GDGP0028)

Supported by: National Natural Science Foundation of China(62277014, 62272142); Fundamental Research Funds for the Central Universities of China(PA2025GDGP0028)

directly applied to real-time rendering super-resolution. First, most of these methods require a network submodule to estimate the optical flow map between two adjacent frames, which inevitably increases the algorithm's time complexity and reconstruction error. In real-time rendering, the algorithm can directly obtain almost complete motion vector maps from the modern graphics rendering pipeline, effectively avoiding the time overhead and cumulative errors of generating optical flow maps. Second, the latest video super-resolution algorithms emphasize the importance of bidirectional propagation of video frame features for reconstruction quality. However, in real-time rendering, the algorithm cannot obtain feature information of future frames, making bidirectional propagation of frame features impossible. Lastly, real-time rendering pipelines often generate images that record scene geometry information, and failing to effectively utilize this data limits the final reconstruction results. **Method** In response to the above issues, this method proposes a real-time neural super-sampling method based on a frame cyclic structure. First, it can make full use of the low-resolution scene geometry data generated in the real-time rendering pipeline to enhance the three-dimensional spatial information perception of the super-sampling network. Second, the frame cyclic framework is integrated into the super-sampling method. Temporal stability is achieved by introducing the features of the previous frame's reconstruction results to improve the current frame's reconstruction. Finally, a reweighting network and an attention network are embedded in the feature extraction module to enhance the effectiveness of the extracted features. Additionally, this method presents a real-time rendering pipeline oriented to neural super-sampling, which can deploy the super-sampling network onto the graphics computing pipeline and integrate it with the real-time rendering pipeline. To further ensure the real-time performance of the real-time rendering pipeline, we performed lightweight processing on the proposed real-time neural supersampling method at different levels and compared the changes in various image evaluation metrics and real-time performance of the network before and after lightweighting in subsequent experiments. No public and universal dataset in the field of neural supersampling is available. Thus, we use Unity3D for the collection of the dataset. For low-resolution data, we set the output resolution to 320 × 180 pixels and disable anti-aliasing. For high-resolution data, we set the output resolution to 5 120 × 2 880 pixels with 8 × MSAA anti-aliasing enabled, then downsample it to 1 280 × 720 pixels and apply a sharpening operation with a 3 × 3 convolution kernel to the downsampled result to further improve the quality of the test data. **Result** Compared with NSRR, a benchmark method that also achieves good real-time performance, the proposed method shows a slight improvement in speed while increasing the peak signal-to-noise ratio by an average of 0.4 dB. Meanwhile, the structural similarity index is increased by an average of 0.015, and the learned perceptual image patch similarity is decreased by an average of 0.028. After being deployed to the real-time rendering pipeline, the method maintains real-time performance through lightweight pruning, and its performance remains superior to that of the non-real-time deployed NSRR in most scenarios. Subsequent experiments also demonstrate that the lightweight network can operate on the graphics computing pipeline and meet real-time requirements at various resolutions, as validated by experimental data. Additionally, through ablation experiments on the three submodules of the network, this method validates the effectiveness of each submodule in reconstructing the current frame. Moreover, the image evaluation metrics show a minimal decline when the scene normals are sufficiently simple or the camera movement amplitude is too large. **Conclusion** First, compared with previous super-resolution and super-sampling methods, the proposed neural super-sampling network model not only achieves better performance but also demonstrates superior real-time performance. Second, the constructed neural super-sampling rendering pipeline can operate on the graphics pipeline and meet real-time requirements at different resolutions. In the future, the regions of rendered images prone to artifacts can be further optimized by starting from the generation principles of rendered images. For instance, for dynamic shadows and specular reflections, ordinary motion vector maps cannot capture their changing trends as objects and light sources move. Thus, the rendering pipeline may need to be customized further to output vector maps that describe such changes to address this issue.

Key words: real-time rendering; frame-recurrent neural network; super-sampling; super-resolution (SR); convolutional neural network(CNN)

0 引 言

现代实时渲染算法的复杂度通常与目标屏幕的像素点个数密切相关，如光线追踪算法。然而，随着外接显示屏和头戴式显示设备的分辨率与刷新率不断提高，以及能够实现更逼真视觉效果的图形算法的提出，实时渲染图形程序的计算成本相比之前已有了成倍增加。这使得即使在全新的并行计算硬件支持下，许多用户在运行高级图形程序时也不得不在渲染分辨率、渲染实时性和渲染质量之间做出权衡。

尽管现在已有许多视频超分辨率（super-resolution，SR）方法输出了令人印象深刻的重建结果（Sajjad等，2018），但这些方法并不能直接应用于实时渲染领域。首先，其中大部分方法都需要一个网络子模块来预估两个相邻帧之间的光流图，这必然会导致算法时间复杂度和重建结果误差的增加。而在实时渲染中，可以直接从图形渲染管线中获取几乎完整的运动矢量图，从而避免生成光流图的时间开销和累积误差。其次，最新的视频超分算法（Chan等，2022）强调了视频帧特征双向传播对重建结果质量的重要性，而在实时渲染管线中，由于算法无法获取未来帧的特征信息，因此无法执行帧特征的双向传播操作以提高当前帧的重建结果。

因此，针对实时渲染领域的超采样方法开始得到关注，这些方法旨在维持实时性的同时重建并输出更高分辨率的渲染图像。相比于视频超分辨率方法，它们更注重抗锯齿性能、算法实时性以及渲染管线数据的利用情况。其中，用于实时渲染的神经超采样（neural super-sampling for real-time rendering，NSRR）方法（Xiao等，2020）率先且成功地在确保超采样方法实时性的同时将渲染管线输出的场景深度数据融入到低分辨率渲染图像的重建过程中，并通过滑动窗口的方式在每一个循环批次中结合固定个数的历史帧的图像特征来生成当前帧的重建结果。

然而，该方法在重建结果的时间稳定性和视觉表现上仍然存在不足，且未直接接入现有的实时渲染管线中。因此，本文提出了一个基于帧循环结构的神经超采样方法，并在此基础上搭建了一个面向神经超采样的实时渲染流程。

本文的主要贡献如下：1）提出了一个神经超采样方法，能够充分利用渲染管线输出的低分辨率场景几何信息提高重建结果的视觉表现指标；能够将先前帧的重建结果特征引入到当前帧的重建过程以提升输出结果在时间尺度上的稳定性。2）将特征重加权和注意力网络置于不同的特征提取流程中，前者可以削弱由摄像机和物体动态遮挡所导致的无效历史帧图像特征的影响；后者能够削弱先前帧重建结果中由于误差累积导致的无效特征的影响。3）搭建了一个实时渲染流程，能够将训练出来的超采样网络部署至图形计算管线上，从而直接将其接入实时渲染管线的末端，在图形程序运行时进行实时神经超采样任务。

1 相关工作

1.1 视频超分方法

视频超分方法在工作流程与重建目标上与渲染超采样方法大体一致，都需要在未对齐的帧之间收集互补的特征信息以提高当前帧重建结果的质量。在视频超分方法中，一种流行的方法是基于滑动窗口框架，该框架往往以轮询的方式将固定长度的低分辨率源图像送入网络进行训练，这种方法更注重源训练数据在局部时间尺度上的特征信息，但却无法保证各个重建结果在时间尺度上所应该具有的连续性。视频亚像素卷积神经网络（video efficient sub-pixel convolutional network，VESPCN）（Caballero等，2017）方法由对齐网络和融合时空亚像素网络组成，能够有效利用时间冗余，提高重建精度，同时保持实时速度；增强深度残差网络（enhanced deep residual networks，EDSR）（Lim等，2017）方法提出了一种增强的深度超分辨率网络，它通过去除残差网络中不必要的组成模块，提高了图像超分辨的效果。与滑动窗口框架相比，基于帧循环结构的视频超分方法往往会进行单向或者双向的帧特征传播操作，该操作在一定程度上保证了重建结果在时间尺度上的连续性。最新的视频超分研究表明，使用双向传播框架的视频超分方法的重建结果相比于使用滑动窗口框架和单向传播框架的效果更好。帧循环视频超分辨率（frame-recurrent video super-resolution，FRVSR）方法（Sajjad等，2018）提出了一个端到端可训练的帧循环视频超分辨率框架，该框架使用先前

帧的重建结果估计后续帧的重建结果，从而使得重建结果保持时间尺度上的一致性；时间相关的对抗生成网络（temporally coherent generative adversarial network，TecoGAN）（Chu 等，2020）方法首次提出了一种对抗和循环训练方法，以监督空间高频细节和时间关系。在没有真值动态的情况下，时空对抗损失和循环结构可使该模型生成照片级真实度的细节，同时使帧与帧之间的生成结构保持连贯；递归反投影网络（recurrent back-projection network，RBPN）（Haris 等，2019）方法使用循环编码器—解码器模块集成来自连续视频帧的空间和时间上下文，该模块将多帧信息与传统的目标帧的单帧超分辨率路径融合在一起，从而获取目标帧的重建结果；基础视频超分辨率（basic video super-resolution，BasicVSR）（Chan 等，2021）方法对视频超分研究进行了梳理并重新审查了视频超分的 4 个基本模块，通过复用现有方案的模块，作者提出了一个轻量且高表现性能的视频超分框架；BasicVSR++方法（Chan 等，2022）则是在 BasicVSR 的基础上对传播模块和对齐模块进行了增强，使得网络模型重建结果的视觉指标更为优越。国内近年来也开始关注视频超分的研究（江俊君 等，2023），如引入各类图像特征方法进行帧间对齐来提升效果（靳雨桐 等，2022）。尽管目前的视频超分方法已经取得优异的超分结果，但由于其网络输入数据单一，很难充分利用渲染管线的中间数据以取得更好的重建结果，也很难与实时渲染管线相结合。

1.2 渲染超采样

与视频类似，实时渲染的图形程序也依赖于人类视觉暂留，在一定的时间域内以符合实时要求的速率播放离散的图像。然而，图形程序和视频在很多方面是不相同的。首先是对于实时刷新率的要求，尽管两者的图像帧在时间尺度上都是离散的，但对于视频来说，其每个图像帧中记录的信息是在一定时间段内由光学设备采集的，而每个实时渲染帧则是在某个时间点依据屏幕空间的像素进行点采样从而渲染得到的瞬时图像。因此，对于大部分用户来说，视频（如电影）只需要达到 24 帧/s 的刷新率就可以获得不错的流畅度，然而对于实时渲染图形程序来说，为了避免用户察觉到明显的帧缺失现象，刷新率至少需要达到 30 帧/s。其次，视频中的每一帧在录制时会由光学设备记录其周围环境的真实光照情况。这意味着在任一视频帧，环境光会被很好地保留到图像中。然而，在实时渲染中，由于无法直接对基于物理的渲染方程做积分操作，图形程序往往会对虚拟空间中的环境光做离散采样，从而拟合物体在真实环境光照下的效果。因此，实时渲染中的超采样任务往往比视频超分任务更具有挑战性。

在实时渲染中，渲染图像中每一个像素的值都是通过点采样的方式生成。因此，对于渲染图像的超采样，特别是在 4 × 4 的上采样因子时，重建的图像往往会因为采样率不足而出现严重的锯齿现象，因此超采样任务可以被认为是抗锯齿任务与超分任务的集合。

传统的超采样方法可以分为两大类，一类是基于图形空间的抗锯齿方法，其基本方法是以特定规则寻找图像中的欠采样区域，并通过模糊操作将目标区域中的锯齿状边缘变得更平滑。其中多重采样抗锯齿（multisample anti-aliasing，MSAA）（Microsoft，2017）方法通过提取像素点周围的颜色信息并通过混合颜色信息消除高对比界面所产生的锯齿。与 MSAA 方法不同的是，快速近似抗锯齿（fast approximate anti-aliasing，FXAA）（Lottes，2009）方法则是将像素颜色的提取和混合过程交由 GPU（graphics processing unit）内算术逻辑单元执行，从而加快了抗锯齿的速度并降低了显存消耗。MLAA（morphological antialiasing）方法（Reshetov，2009）首先对锯齿边缘执行重矢量化操作，然后以重矢量化线段覆盖的像素面积为混合系数，将锯齿两侧的像素按照混合系数进行混合。子像素形态学抗锯齿（subpixel morphological anti-aliasing，SMAA）（Jimenez 等，2012）方法则是在 MLAA 的基础上更精确了边界判断与边界搜索的过程。另一类方法是时间抗锯齿算法，该方法首先对虚拟摄像机施加抖动，然后通过一个时间累加器将多个历史帧的采样信息运用到当前帧的抗锯齿计算中，该类算法的代表为时域抗锯齿（temporal anti-aliasing，TAA）（Meinds 等，2003）方法。

基于深度学习的超采样方法开始涌现，其中深度学习超采样（deep learning super sampling，DLSS）（Andrew，2020）与英特尔 Xe 超级采样（Intel® Xe super sampling，XeSS）（Guo 等，2021）使用预训练的超采样网络对低分辨渲染图像进行上采样，从而降低直接渲染高分辨图像的时间成本。尽管这些方法

已经在游戏领域得到应用，然而它们主要关注上采样因子小于等于2×2的重建任务，且有着严格的硬件限制。与DLSS和XeSS相比，NSRR则注重利用多个低分辨率的历史帧特征完成上采样因子为4×4的重建任务。Zhong等人(2023a)在NSRR的基础上添加了特征增强模块和特征缓存模块，特征增强模块要求在实时渲染管线中额外生成多幅与高分辨率大小一致的场景几何信息图，并在重建阶段融入由上述几何信息图中提取的特征信息来提高重建结果的视觉表现。特征缓存模块则优化了在滑动窗口框架中对低分辨率历史帧图像频繁的特征提取操作。Zhong等人(2023b)提出了一种高效的超采样方法，该方法同样控制渲染管线额外生成高分辨率的场景几何信息作为输入来预测高质量的上采样重建。之后，该方法引入了一个高效的H形网络架构来使得网络能够在多分辨率上对齐和融合特征。尽管上述两个方法都取得了不俗的重建表现，但它们都需要控制渲染管线来额外生成与重建结果分辨率大小一致的场景几何信息，而这些额外输出的高分辨率数据又无疑加重了渲染管线的运行负担，并且该负担会随着屏幕分辨率的变大而成倍提高，从而增加了方法与实时渲染管线相结合的难度。

2 神经超采样网络结构

图1展示了本文提出的神经超采样网络结构及各个子模块的设计。首先，将渲染管线输出的低分辨率(low resolution，LR)场景几何数据融入到神经超采样网络中，从而使得网络训练与评估时能够感知更丰富的三维场景信息。图2展示了渲染管线的中间数据。其次，将基于单向帧循环结构引入到网络的设计之中，该结构通过一个注意力加权模块(SR attention)提取先前帧重建结果中的有效特征，并将其加入到当前帧的重建过程中。然后，使用一个特征重加权模块过滤掉历史帧图像中由动态遮挡所带来的无效特征，最后，使用一个U形残差网络生成最终的重建结果。

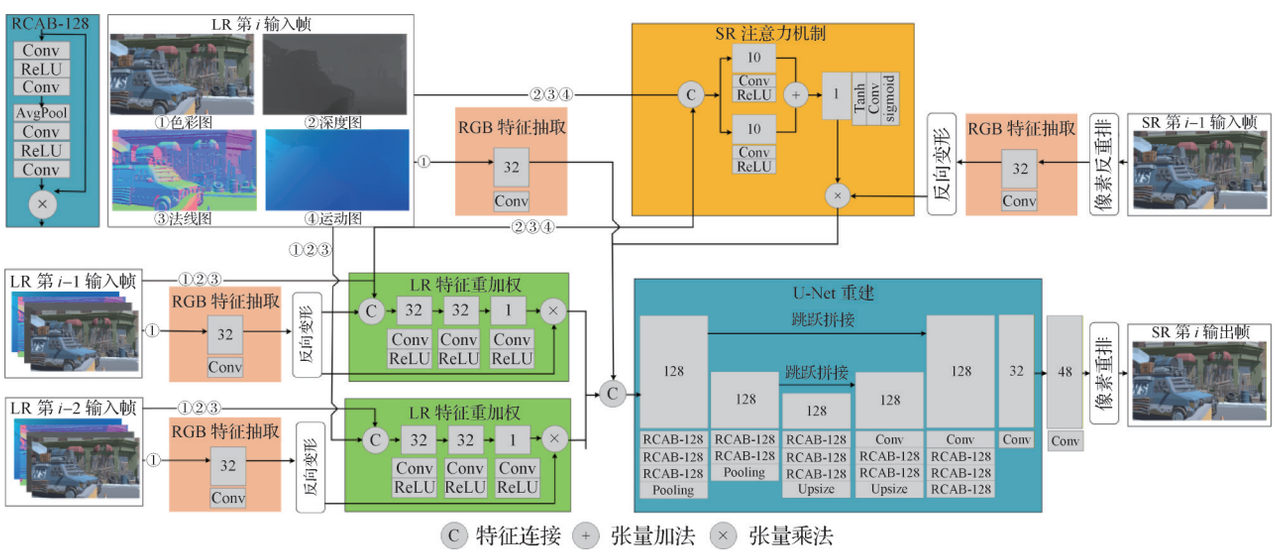

图1 帧循环结构的重建网络

Fig. 1 Reconstruction network with frame recurrent structure

2.1 帧循环结构

基于滑动窗口的超分方法将目标任务视为多个独立帧之间的超分问题。这些方法往往以固定长度的滑动窗口处理多个历史帧图像从而生成最终的结果。然而，这种方法的计算成本很高，因为每个历史帧都会在滑动窗口框架下被多次处理。除此之外，将每一个重建结果视为相互独立的帧输出会大大降低各个结果帧间的时间稳定性，从而导致不可预料的伪影与闪烁。为了提高网络输出结果时间稳定性，同时削减由帧循环误差累积带来的伪影效果，本文重新设计了帧循环网络框架，它由3个部分组成。

1)帧循环部分。该部分将先前帧的重建结果特征引入到当前帧的重建过程中。具体为

(a) 色彩图　(b) 深度图

(c) 法线图　(d) 运动图

图2　渲染管线生成的场景几何信息

Fig. 2　The scene geometry information generated by the pipeline ((a) color map; (b) depth map; (c) normal map; (d) motion map)

$$\tilde{\boldsymbol{I}}_{i-1}^{\mathrm{SR}} = f_{\mathrm{SRAttention}}(f_{\mathrm{BackWarp}}(\boldsymbol{F}(\boldsymbol{M}(\boldsymbol{I}_{i-1}^{\mathrm{SR}})))) \tag{1}$$

式中，$\tilde{\boldsymbol{I}}_{i-1}^{\mathrm{SR}}$是前一帧的重建图像，$\boldsymbol{M}(\boldsymbol{X})$表示将高分辨率图像映射到低分率空间中，$\boldsymbol{F}(\boldsymbol{X})$表示提取前一帧重建结果的特征信息，$f_{\mathrm{BackWarp}}(\boldsymbol{X})$表示将特征图扭曲投影到当前帧，$f_{\mathrm{SRAttention}}(\boldsymbol{X})$表示对特征图执行注意力加权操作，$\boldsymbol{I}_{i-1}^{\mathrm{SR}}$表示提取到的特征信息图。因此，帧循环部分首先对前一帧的重建结果进行降采样和特征提取操作，然后根据当前帧的运动矢量图对上述特征进行重投影操作，之后再通过注意力加权模块从特征图中提取到对重建结果有效的部分。

2)滑动窗口部分。该部分将固定数目的历史帧特征引入到当前帧的重建结果中。具体为

$$\tilde{\boldsymbol{I}}_{i-1}^{\mathrm{LR}} = f_{\mathrm{LRReweight}}(f_{\mathrm{BackWarp}}(\boldsymbol{F}(\boldsymbol{I}_{i-1}^{\mathrm{LR}}))) \tag{2}$$

$$\tilde{\boldsymbol{I}}_{i-2}^{\mathrm{LR}} = f_{\mathrm{LRReweight}}(f_{\mathrm{BackWarp}}(\boldsymbol{F}(\boldsymbol{I}_{i-2}^{\mathrm{LR}}))) \tag{3}$$

式中，$\tilde{\boldsymbol{I}}_{i-1}^{\mathrm{LR}}$和$\tilde{\boldsymbol{I}}_{i-2}^{\mathrm{LR}}$表示历史帧的低分辨率渲染图像，$f_{\mathrm{LRReweight}}(\boldsymbol{X})$表示特征重加权。滑动窗口部分每一次循环都会将固定个数的历史帧图像输入到网络中并提取这些图像的特征图，然后对特征图执行重投影操作。此后，特征重加权模块会对特征图做一次评估，从而减弱由动态遮挡带来的无效特征信息对重建结果的负面影响。

3)重建部分。该部分将前两部分获得的特征图合并(⊕)在一起，并通过一个U形残差网络生成最终的重建结果，即

$$\boldsymbol{I}_{i}^{\mathrm{SR}} = f_{\mathrm{U\text{-}Net}}(\tilde{\boldsymbol{I}}_{i-1}^{\mathrm{SR}} \oplus \tilde{\boldsymbol{I}}_{i-1}^{\mathrm{LR}} \oplus \tilde{\boldsymbol{I}}_{i-2}^{\mathrm{LR}} \oplus \boldsymbol{F}(\boldsymbol{I}_{i}^{\mathrm{LR}})) \tag{4}$$

2.2　特征重加权

图1展示了特征重加权模块的内部结构。该模块的主要目标是处理由运动矢量图带来的无效特征。相比于视频超分方法中估计出的光流图，由渲染管线获得的运动矢量图在大部分情况下都更好地描述了屏幕空间中各个像素点在两帧之间的平移矢量，但它仍然存在一些限制。首先，它不会考虑到同一个物体在前一帧与当前帧的遮挡情况，即当前帧的可视物体由于相机或者其他物体的移动可能在上一帧被遮挡，此时运动矢量图会将在时间尺度上毫无关系的像素点错误地联系在一起，从而提取到会导致重建结果质量下降的无效特征。此外，它仅考虑屏幕空间下像素点的二维平移，对于三维空间中的其他变化，只靠一个运动矢量图是无法正确反映的。

为此，本文在特征重加权模块中引入了由渲染管线中生成的低分辨率的场景法线信息。相比仅使用着色图和深度图，场景法线信息不仅能表现出同一物体表面的不同细节信息，也能在同一深度下反映出不同物体之间的不同朝向。特征重加权模块对于每一个输入的历史帧特征都生成一个对应的加权矩阵，该矩阵反映了目标历史帧与当前帧之间物体的动态遮挡关系。通过将历史帧的特征信息与加权矩阵相乘，可以有效屏蔽可能出现的伪影。该模块采用3层卷积神经网络进行处理，将经过重投影操作的历史帧特征信息作为输入。网络为每一个历史帧预测加权，再用其乘以对应历史帧的特征图，从而作为重加权网络的输出结果。

2.3　注意力加权

对于在先前帧重建结果上提取与重投影的特征信息，由于其本身由网络评估得出，带有一定的误差，因此本文并不使用特征重加权模块对其进行二次评估，而是使用一个注意力网络对提取到的特征进行有效性评估。图1展示了注意力加权模块的内部结构，该模块先利用两帧之前的运动矢量图、场景法线矢量图和场景深度图生成一个加权图，然后将加权图与先前帧重建结果提取的特征图相乘，从而提取出有效特征。

2.4　损失函数

本文为网络训练所设计的损失函数为

$$\mathcal{L}(\boldsymbol{I}_{i}^{\mathrm{SR}}, \boldsymbol{I}_{i}^{\mathrm{HR}}) = 1 - f_{\mathrm{SSIM}}(\boldsymbol{I}_{i}^{\mathrm{SR}}, \boldsymbol{I}_{i}^{\mathrm{HR}}) + w \cdot \sum_{i=1}^{5} \left\| conv_i(\boldsymbol{I}_{i}^{\mathrm{SR}}) - conv_i(\boldsymbol{I}_{i}^{\mathrm{HR}}) \right\|_2^2 \tag{5}$$

式中，$\boldsymbol{I}_{i}^{\mathrm{SR}}$表示第$i$帧的重建结果，$\boldsymbol{I}_{i}^{\mathrm{HR}}$表示第$i$帧的真实高分辨图像，$w$的值设置为0.1。损失函数由两个

部分加权组成，其一是结构相似性指数(structural similarity index measure，SSIM)(Wang等，2004)，另一个则是从预训练的VGG-16(Visual Geometry Group)网络计算的感知损失(perceptual loss)(Johnson等，2016)。

SSIM可以衡量两幅图像的相似程度，从而评价图像的重建情况。相较于传统方法中使用的图像品质衡量指标，结构相似性属于感知模型，它在图像品质衡量上的指标更能符合人眼对图像品质的判断，而这一特性也符合我们设计算法的目的。感知损失是图像风格迁移中的损失函数。与传统的均方误差(mean squared error，MSE)损失函数相比，感知损失更注重图像的感知质量，更符合人眼对图像质量的感受。利用预训练的VGG-16网络能够将重建图像卷积得到的高频特征与真实图像卷积得到的高频特征进行比较，使得其高频信息相近。

2.5 数据集构建

目前，实时神经超采样领域并没有公开且通用的数据集，因此本文使用Unity3D构建了一个数据集。在Unity3D中收集了覆盖室内和室外的三维模型场景，并编写了数据收集脚本和数据预处理脚本。其中，数据收集脚本负责生成网络所需的低分辨率训练数据与高分辨率(high resolution，HR)测试数据。低分辨数据包括最终着色图、运动矢量图、深度图和记录像素点在世界空间下的法线图。生成前，将输出分辨率设置为320×180像素并关闭所有的抗锯齿效果。而对于高分辨数据，仅输出最终的着色图，生成前将输出分辨率设置为5 120×2 880像素并开启8倍的MSAA抗锯齿功能。数据预处理脚本则首先使用降采样将之前采集的高分辨着色图像的分辨率降低至1 280×720像素，然后再对降采样后的结果使用卷积核大小为3×3的锐化操作来进一步提高测试数据的质量。

通过数据生成脚本，总计生成了50个数据集用例，每个用例都是帧连续的100个图像集序列，而每个图像集中又包含了超采样算法所需的低分和高分辨率图像数据，总计25 000幅训练图像。训练时，按照34∶8∶8的划分方法划分出训练集、验证集和测试集。图3展示了所构建的数据集用例。

在网络训练与验证时，将低分辨图像划分成32×32像素的图像块以丰富训练数据，高分辨率图像则划分为128×128像素。在网络测试时，则在完整分辨率的图像上执行网络评估操作。

本文网络使用Pytorch进行训练。使用默认超参数的自适应矩估计(adaptive moment estimation，ADAM)优化器方法进行训练优化，学习率设置为0.000 01，批次大小为16，运行了100个epoch。网络在1个RTX3060 GPU上训练大约需要一天。

(a) 色彩图　(b) 深度图　(c) 法线图　(d) 运动图　(e) 参考图

图3　数据集用例展示

Fig. 3　Demonstration of dataset use cases ((a) color map; (b) depth map; (c) normal map; (d) motion map; (e) reference map)

3　实时神经超采样渲染流程

为了提高本文方法的应用价值，本节搭建了一个面向实时神经超采样的渲染流程，该流程通过利用基于计算着色器实现的神经网络推理框架 Sentis（https://github.com/Unity-Technologies/Sentis-samples）将神经超采样网络部署至图形计算管线之上，图 4 展示了该流程。

3.1　网络轻量化策略

在将由 PyTroch 训练好的超采样网络部署至 Sentis 上后，出现了明显的运行速率下降。通过后续的实验发现，Sentis 的二维卷积算子的运行效率相比于 PyTorch 会随着输入通道个数的增加而更显著地降低，这限制了网络在图形计算管线上运行的实时性。对此，本节首先对超采样网络进行了轻量化处理。

轻量化后的神经超采样网络仍采用基于单向帧循环的网络结构，以保证重建结果在时间尺度上的稳定性。对于神经网络中的滑动窗口结构，做了部分取舍。首先，对于第 i 帧图像的重建过程来说，由于在单向帧循环的结构中，已经将第 $i-1$ 帧重建结果的特征信息进行了提取并融入到当前帧的重建过程中，因此，轻量化网络舍去了滑动窗口结构中对第 $i-2$ 帧的低分辨率渲染图像的特征提取操作，仅保留对第 $i-2$ 帧渲染图像的特征提取；其次，对低分辨率渲染图像的特征提取模块的输出维度进行了删减，减少了一半的输出维度；最后，由于特征输出模块的维度减少，对神经超采样网络后续的特征重加权以及 U 形残差网络重建模块的输入维度做了同步删减。在后续实验中，对比了轻量化前后的超采样网络的运行时间指标与视觉表现评估指标。

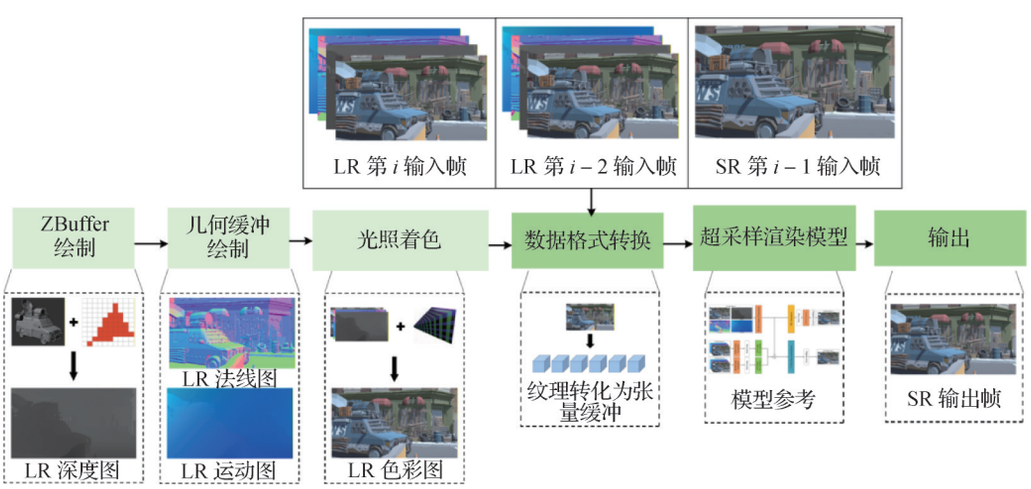

图 4　实时神经超采样渲染流程

Fig. 4　Real-time neural supersampling on render pipeline

3.2　实时渲染流程

本节基于图形渲染与计算管线搭建了面向实时神经超采样的渲染流程。

首先，采用延迟渲染方法搭建图形渲染管线，尽管该方法会增加图形渲染管线的带宽消耗，但该方法不仅可以生成后续实时神经超采样模块进行网络推理所需的各类低分率的场景几何与材质数据，还可以进一步提高图形渲染管线渲染一帧图像的速率。其次，利用 Sentis 将先前轻量化处理后的实时神经超采样网络部署至图形计算管线之上，在渲染流程上处于延迟渲染管线的末端。

延迟渲染方法将物体的渲染任务拆分为 3 个子渲染阶段，首先是 ZBuffer 绘制阶段，该阶段会预先生成一幅低分辨率的场景深度图；其次是几何缓冲绘制阶段，该阶段会遍历三维场景中的所有非透明物体并将这些物体的几何（如位置信息以及法线信息等）与材质（如基础反照率、金属度以及粗糙度等）信息渲染到多幅与输出分辨率大小一致的纹理上；最后是光照着色阶段，该阶段会利用自定义的着色算法生成一幅场景颜色图像，对于该颜色图像上的每一个像素点，该阶段都会从几何缓冲阶段绘制的纹理中进行采样，以获取这些像素点所对应的几何与材质信息，从而为着色算法提供数据。

本节利用 Sentis 神经网络推理框架将轻量化的

神经超采样网络部署至图形计算管线上，并以完整的浮点数精度进行网络推理。为了对齐延迟渲染管线模块与网络推理模块的数据格式，两个模块之间额外添加了格式转换阶段。该转换阶段是将渲染管线输出的纹理图像进行通道平铺，将图像的各个分量按其所处通道中的顺序依次存放在张量缓冲中，这是因为Sentis推理框架是基于缓冲的数据格式而非纹理进行神经网络推理，图5展示了纹理图像转换为张量缓冲数据的过程。在网络推理完成后，会重新通过逆变换将缓冲类型的数据转换为目标分辨率下的纹理图像，从而输出最终的重建结果。此外，在部署神经网络时，额外生成了用于面对不同分辨率情况的神经超采样网络模型。

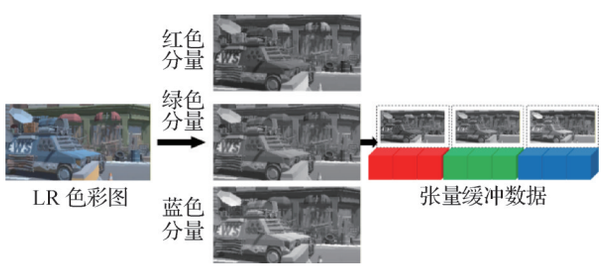

图5　数据转换示意图

Fig. 5　Diagram of data transformation

4　实验结果与分析

本节对神经超采样方法进行对比实验与分析。4.1节中与其他超分辨率与超采样方法进行效果评估对比，4.2节对网络的各个子模块进行消融实验。

4.1　效果评估

本节在Unity3D的数个三维场景中录制了由摄像机捕获到的连续渲染帧，并在其上对本文方法与之前的超采样和超分辨率方法进行结果评估。

轻量化策略上，根据轻量化程度的不同，额外生成了两个轻量化超采样网络，分别为FRNSRR-64以及FRNSRR-32，用来与本文算法进行对比。超采样方法中，选择NSRR进行对比。超分辨率方法中，基于方法实时性选择VESPCN和EDSR进行对比，基于循环网络框架选择TecoGAN和RBPN进行对比，方法的实现来源于共享代码，每个方法都在本文搭建的数据集上重新训练。评估指标中，优先选择了峰值信噪比(peak signal to noise ratio，PSNR)和SSIM这两项指标。此外，添加了学习感知图像块相似度(learned perceptual image patch similarity，LPIPS)(Zhang等，2018)这项评估指标，作为一种衡量图像相似度的指标，它通过深度学习模型评估两个图像之间的感知差异。LPIPS认为，即使两个图像在像素级别上非常接近，人眼也可能将它们视为不同。LPIPS使用预训练的深度网络(如AlexNet)提取图像特征，然后计算这些特征之间的距离，以评估图像之间的感知相似度。相比于传统的图像评估指标，如PSNR和SSIM，LPIPS更符合人眼视觉的感知情况，LPIPS的值越低表示两幅图像越相似，反之，则差异越大。

在表1—表3中，与其他方法比较了上述3个评估指标的平均值。其中的每一个场景都使用连续渲染的25帧图像进行效果评估，记录各方法面对这些

表1　不同方法在不同三维场景下重建结果的PSNR指标

Table 1　PSNR results of super-resolution using different methods in different 3D scenes

/dB

方法	广告牌	坦克基地	黑市	汉堡店口	十字路口
VESPCN	21.231 6	24.196 3	23.195 7	23.950 8	24.584 4
EDSR	21.659 8	24.829 6	23.474 9	<u>24.514 0</u>	24.984 2
TecoGAN	21.682 5	24.827 0	23.535 0	24.434 8	25.007 3
RBPN	21.812 0	**25.069 0**	<u>23.594 3</u>	24.456 4	25.177 2
NSRR	21.381 9	24.372 0	23.482 1	24.046 8	25.009 7
FRNSRR(本文)	**21.837 4**	<u>24.879 4</u>	**23.645 3**	**24.555 1**	**25.386 0**
FRNSRR-64	<u>21.830 8</u>	24.669 5	23.584 9	24.260 7	<u>25.279 5</u>
FRNSRR-32	21.547 5	24.288 9	23.241 7	23.967 9	24.833 7

注：加粗、下划线字体分别表示各列最优、次优结果。

表 2　不同方法在不同三维场景下重建结果的 SSIM 指标

Table 2　SSIM results of super-resolution using different methods in different 3D scenes

方法	广告牌	坦克基地	黑市	汉堡店口	十字路口
VESPCN	0.721 9	0.777 5	0.795 7	0.782 2	0.774 7
EDSR	0.756 6	0.814 1	0.825 6	0.818 0	0.809 0
TecoGAN	0.741 8	0.799 3	0.810 2	0.807 0	0.793 4
RBPN	0.770 4	0.827 9	0.836 4	0.824 4	0.824 9
NSRR	0.770 0	0.829 2	0.840 1	0.825 7	0.825 2
FRNSRR(本文)	**0.788 9**	**0.844 7**	**0.851 8**	**0.841 0**	**0.843 4**
FRNSRR-64	0.779 6	0.828 6	0.843 8	0.830 3	0.834 0
FRNSRR-32	0.767 8	0.818 0	0.834 9	0.818 8	0.821 8

注：加粗、下划线字体分别表示各列最优、次优结果。

表 3　不同方法在不同三维场景下重建结果的 LPIPS 指标

Table 3　LPIPS results of super-resolution using different methods in different 3D scenes

方法	广告牌	坦克基地	黑市	汉堡店口	十字路口
VESPCN	0.297 4	0.233 2	0.225 9	0.292 1	0.258 3
EDSR	0.224 6	0.158 6	0.153 0	0.205 3	0.187 4
TecoGAN	0.192 8	**0.133 2**	0.137 2	0.186 8	0.155 3
RBPN	0.196 0	0.144 6	**0.128 6**	0.181 6	0.157 4
NSRR	0.226 2	0.160 3	0.159 4	0.216 5	0.170 3
FRNSRR(本文)	**0.189 9**	0.137 8	0.136 1	**0.181 0**	**0.147 8**
FRNSRR-64	0.209 9	0.161 9	0.147 3	0.200 6	0.164 5
FRNSRR-32	0.235 7	0.179 1	0.172 7	0.228 1	0.184 0

注：加粗、下划线字体分别表示各列最优、次优结果。

连续帧执行超采样操作所获得的重建结果的指标平均值，标记出各个指标中位于前 3 的值。

表 1—表 3 表明，本文方法在大多数情况下优于其他方法，特别是在与 NSRR 对比时，在 3 项指标中均表现出更优异的结果。其中 PSNR 平均提升 0.4 dB，SSIM 平均提升 0.015，LPIPS 平均降低 0.028。此外，相较于轻量化前，FRNSRR-64 的 PSNR 指标平均下降了 0.15 dB，SSIM 指标仅平均下降 0.01，LPIPS 指标平均上升了 0.018。FRNSRR-32 的 PSNR 指标平均下降 0.49 dB，SSIM 指标平均下降 0.021，LPIPS 指标平均上升 0.041。尽管指标有所下滑，但在一些复杂场景下，通过利用渲染管线的中间数据，轻量化后的网络仍能优于其他方法。

表 4 展示了与其他方法的网络模型参数大小与在 PyTorch 上的以半浮点数精度进行推理的时间对比。表 5 展示了超采样方法内各个模块在 PyTorch 上以半浮点数精度进行推理的时间，输入大小为 320 × 180 像素，上采样因子 4 × 4。从表 4 和表 5 可以看出，相比其他方法，本文方法在取得显著效果的同时，也有效地减少了网络的推理时间，因此更具实时应用价值。

表 6 和表 7 中的各方法利用计算着色器提供的 GPU 端的并行计算能力在 Unity3D 开发的 Sentis 上进行网络推理。表 6 展示了各方法在 Sentis 上以全浮点数精度推理所需的时间（Sentis 不支持半浮点数精度推理），输入大小 320 × 180 像素，上采样因子 4 × 4，全精度。由于先前的实验证明了 TecoGAN 和 RBPN 的实时性不足（未达到 30 帧/s），故选择其他实

表4　PyTorch上不同方法的模型参数大小和运行时间

Table 4　Model parameter sizes and running time of different methods on PyTorch

方法	参数/M	运行时间/ms
VESPCN	0.878	**16.27**
EDSR	1.517	24.46
TecoGAN	3.834	34.83
RBPN	15.304	1 293.23
NSRR	0.805	26.42
本文	4.544	<u>23.06</u>

注：加粗、下划线字体分别表示最优、次优结果。

表5　各个模块在PyTorch上的推理时间

Table 5　Running time of each module

模块	推理时间/ms
FeatureExtraction	0.87
Warping	3.21
LR Reweighting	1.64
SR Attention	1.68
Reconstruction	15.66
总计	23.06

时性较强的方法。可以看到，将网络部署至Sentis后，各方法的运行效率均有不同幅度下降，以及本文针对Sentis中二维卷积算子处理较大输入通道个数时运行效率降低的问题所做的网络轻量化策略的有效性。表7则展示了本文方法在Sentis下面对不同分辨率时的运行效率，Sentis的运行参数为上采样因子4×4，全精度。可以看到，在轻量化处理之后，网络的运行效率有了较大的提升，并且能够在常用分辨率下达到最低实时性的需求。

图6对不同方法的重建结果进行视觉对比，从连续渲染的帧图像序列中抽取一帧进行。可以看到测试场景选择时，考虑到了室外、室内以及不同时间段的情况。第1列为实验的5个场景，其中红色框部分进行了放大便于细节观察；第2列和最后一列为红色框部分的低质量输入和高质量对比图，其余为表3中不同方法的局部视觉效果对比。当三维场景中的物体堆放比较简单时，本文方法得到的重建结果的评估指标值相比于其他方法并没有特别大的优势。不过，随着场景复杂度的增加，超采样网络能够充分利用由渲染管线所提供的低分辨率的场景深度图和场景法线矢量图感知三维场景的空间信息，从而重建出更优秀的结果。除此之外，也可以直接从与其他方法重建结果的对比中看出，本文方法在边缘处的抗锯齿性能和图像整体的清晰度方面有着显著的效果提升。

表6　不同方法在Sentis上的推理时间

Table 6　Running time of different methods on Sentis

方法	推理时间/ms
VESPCN	26.57
EDSR	61.55
NSRR	63.44
FRNSRR(本文)	53.31
FRNSRR-64	23.46
FRNSRR-32	**12.43**

注：加粗字体表示最优结果。

图7展示了场景测试在SSIM评估指标下的箱型图。每一个箱型图中的元素均表示对一组连续25帧的渲染图像的超采样结果评估的集合。从箱体的大小和位置可以看出，相比于其他超分和超采样方法，本文方法在整体获得更优异的评估指标的同时，其数据也相对更加集中，特别是与其他实时方法的对比。这进一步验证了本文在网络设计时引入

表7　不同方法在Sentis上以不同分辨率进行推理的时间

Table 7　Running time of different methods at different resolutions on Sentis

/ms

方法	分辨率/像素			
	1 280×720	1 600×900	1 920×1 080	2 560×1 440
FRNSRR	53.31	88.65	125.49	205.67
FRNSRR-64	23.64	32.81	48.64	82.72
FRNSRR-32	**12.43**	**19.62**	**28.40**	**49.68**

注：加粗字体表示各列最优结果。

图 6　场景测试对比图

Fig. 6　Comparison of scene test results

的帧循环结构使得连续帧的超采样结果间的时间稳定性有了显著的提高。

4.2　消融实验

针对本文设计的实时神经超采样网络，总共进行了下述 3 组消融实验：

1）验证帧循环结构针对实时神经超采样任务的有效性。通过在当前帧的重建过程中去除从先前帧重建结果提取的有效特征输入来验证。

2）验证渲染管线输出的低分辨率的场景法线矢量图对特征重加权模块和注意力加权模块的影响。通过在对应模块中去除法线矢量图来验证。

3）验证注意力加权模块对先前帧重建结果特征提取的有效性。通过将网络结构中的注意力加权模块替换为由历史帧特征重加权模块改进而来的先前重建结果特征重加权来验证。

通过表 8 和图 8 可知，在大部分情况下，通过提

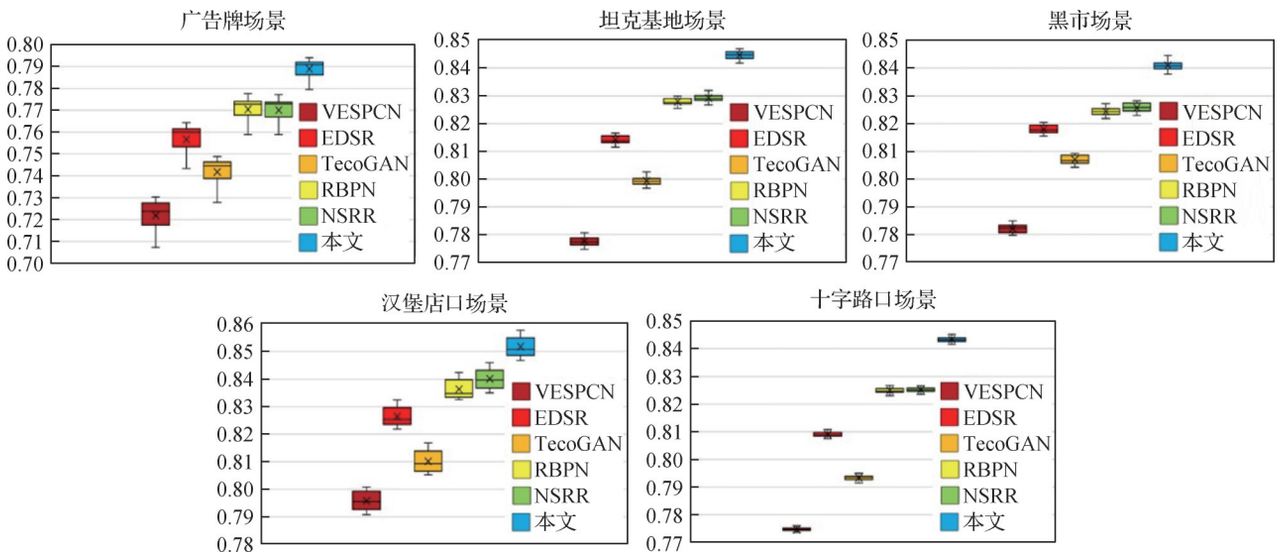

图7　SSIM箱型图

Fig. 7　SSIM indicator box chart for scene testing

表8　验证帧循环的有效性

Table 8　Verify the effectiveness of the frame-recurrent

帧循环	PSNR/dB				
	广告牌	坦克基地	黑市	汉堡店口	十字路口
−Recurrent	21.783 5	24.843 5	**23.679 4**	24.353 3	25.254 7
+Recurrent	**21.837 4**	**24.879 4**	23.645 3	**24.555 1**	**25.386 0**

注：加粗字体表示各列最优结果。

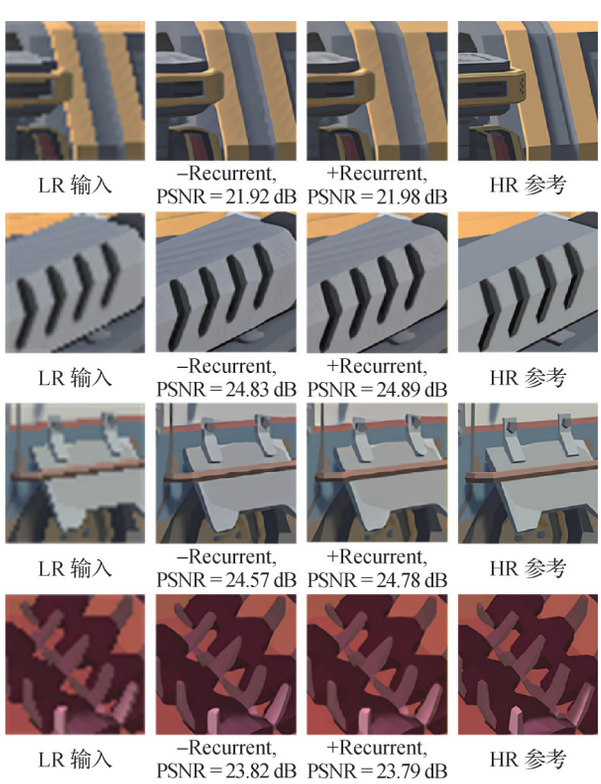

图8　帧循环的消融实验可视化

Fig. 8　Ablation experiment based on frame-recurrent

取和引入先前帧重建结果的特征，可以有效地提高当前帧重建结果的评估指标值。但是，在消融实验中也观察到，在黑市场景下，引入先前帧重建结果的特征反而会对当前帧的重建流程造成不良影响。这主要是因为当每一个帧中的漫游摄像机的平移距离和转角幅度过大时，从先前帧重建结果中逐像素提取的特征信息很多已与当前帧中的像素值没有了时间上的连续关系，因此，引入这些无关的特征信息后，反而可能会导致当前帧的重建结果质量的下降。

然而，由于网络后续的特征重加权模块的存在，从先前帧中提取的无关特征对当前重建流程的不良影响会被尽可能地减小。除此之外，帧循环网络也有效地提高了重建结果的时间稳定性。

表9和图9可以验证，引入渲染管线中输出的低分辨率场景几何信息可以一定幅度地提高重建结果的评估指标。这是因为，相比于仅使用深度图，添加额外的法线图能更好地还原出屏幕像素点对应的三维信息。然而，在消融实验中也发现，当测试

表 9　验证法线图的有效性

Table 9　Verify the effectiveness of the normal map

法线图	PSNR/dB				
	广告牌	坦克基地	黑市	汉堡店口	十字路口
−Normal	21.651 9	24.462 7	**23.655 1**	24.159 4	25.192 3
+Normal	**21.837 4**	**24.879 4**	23.645 3	**24.555 1**	**25.386 0**

注：加粗字体表示各列最优结果。

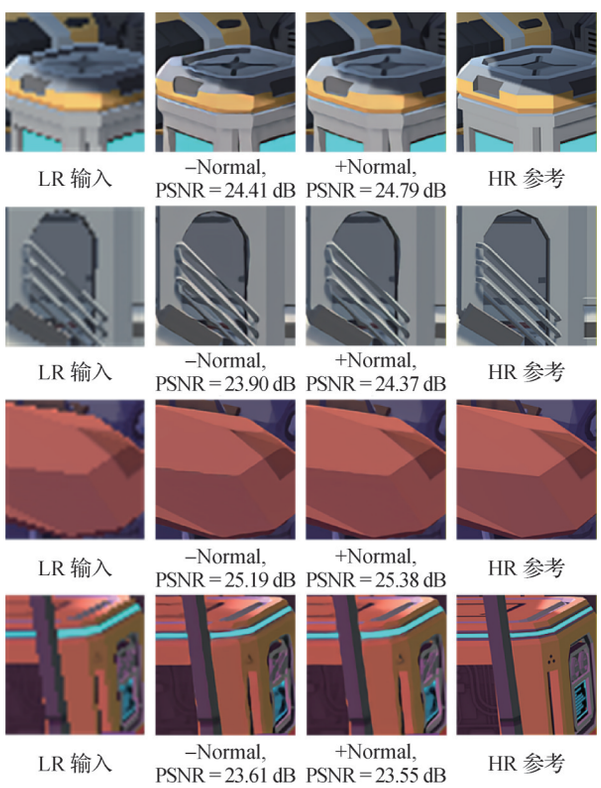

图 9　法线图的消融实验可视化

Fig. 9　Ablation experiment based on normal map

表 10　验证注意力加权（SR attention）的有效性

Table 10　Verify the effectiveness of SR attention

注意力加权	PSNR/dB				
	广告牌	坦克基地	黑市	汉堡店口	十字路口
−SR attention	21.622 4	24.497 1	23.600 1	24.261 6	25.286 8
+SR attention	**21.837 4**	**24.879 4**	**23.645 3**	**24.555 1**	**25.386 0**

注：加粗字体表示各列最优结果。

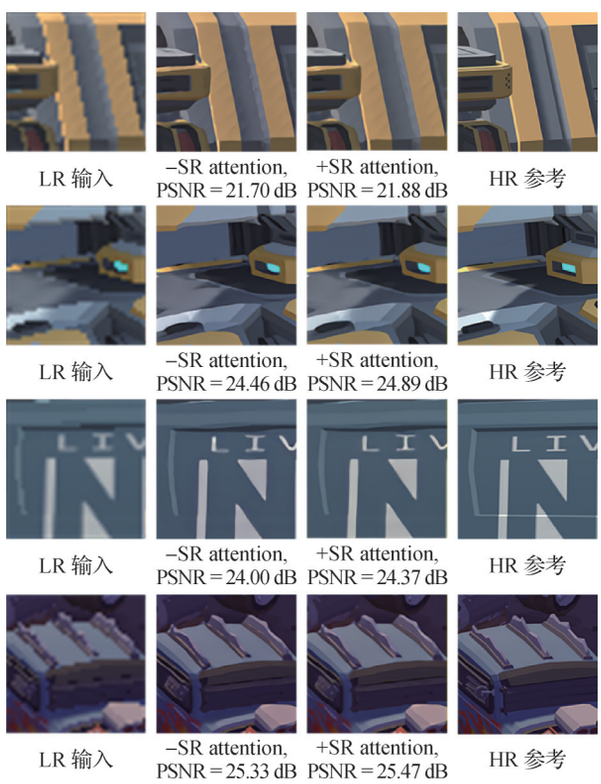

图 10　注意力加权的消融实验可视化

Fig. 10　Ablation experiment based on SR attention

的三维场景过于简单，如场景中大多数物体的表面法线方向单一，或者场景中缺少足够多的物体，都可能会导致评估指标的提升幅度降低。甚至，在黑市场景中，部分物体通过反射率贴图获得了复杂的视觉表现，但是由于其本身建模粗糙导致缺少足够精细的表面法线描述，因此，超采样网络在执行低分辨率历史帧图像特征重加权操作时，会被其粗糙的表面法线误导，从而给出错误的重加权值，最终导致重建结果的质量评估指标下降。

表 10 和图 10 验证了注意力加权模块对先前帧重建结果中有效特征提取的可行性。在消融实验中，当将该模块替换为修改分辨率后的重加权模块（LR feature reweighting）时，质量评估指标出现了明显的下降。

5　结　论

本文提出了一个基于帧循环结构的实时神经超采样方法以及搭建了一个面向实时神经超采样的渲染流程。主要结果与不足如下：1）本文提出的方法能够将渲染管线中的低分辨场景渲染特征融入到当前帧的超采样流程中，较之现有方法在实时帧率下能达到更好的视觉效果，对游戏仿真场景下，低分辨率设备上输出高分辨率画面有提升作

用。2)本文所提出的渲染流程首次将实时神经超采样方法在工程上进行了尝试,使用Senits将超采样模型部署至实时渲染管线之上,达到真实应用的目标,使用的轻量化策略较之其他方法可以有效兼顾性能和效果。3)目前方法在特殊的渲染场景下,如阴影和镜面反射区域还是存在一定的伪影,这是因为渲染管线中存在的运动矢量图只能保存物体本身运动,无法保存这些区域随着其他物体和光源的移动而变化的趋势,需要从渲染机理上进一步思考。

参考文献(References)

Caballero J, Ledig C, Aitken A, Acosta A, Totz J, Wang Z H, et al. 2017. Real-time video super-resolution with spatio-temporal networks and motion compensation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, USA: IEEE: 2848-2857 [DOI: 10.1109/CVPR.2017.304]

Chan K C K, Wang X T, Yu K, Dong C and Loy C C. 2021. BasicVSR: the search for essential components in video super-resolution and beyond//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville, USA: IEEE: 4945-4954 [DOI: 10.1109/CVPR46437.2021.00491]

Chan K C K, Zhou S C, Xu X Y and Loy C C. 2022. BasicVSR++: improving video super-resolution with enhanced propagation and alignment//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, USA: IEEE: 5962-5971 [DOI: 10.1109/CVPR52688.2022.00588]

Chu M Y, Xie Y, Mayer J, Leal-Taixé L and Thuerey N. 2020. Learning temporal coherence via self-supervision for GAN-based video generation. ACM Transactions on Graphics, 39(4): #75 [DOI: 10.1145/3386569.3392457]

Andrew Burnes. 2020. NVIDIA DLSS 2.0: A Big Leap In AI Rendering [EB/OL]. [2025-07-14]. https://www.nvidia.com/en-gb/geforce/news/nvidia-dlss-2-0-a-big-leap-in-ai-rendering/

Guo J, Fu X H, Lin L Q, Ma H J, Guo Y W, Liu S Q, et al. 2021. ExtraNet: real-time extrapolated rendering for low-latency temporal supersampling. ACM Transactions on Graphics, 40(6): #278 [DOI: 10.1145/3478513.3480531]

Haris M, Shakhnarovich G and Ukita N. 2019. Recurrent back-projection network for video super-resolution//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, USA: IEEE: 3892-3901 [DOI: 10.1109/CVPR.2019.00402]

Jiang J J, Cheng H, Li Z Y, Liu X M and Wang Z Y. 2023. Deep learning based video-related super-resolution technique: a survey. Journal of Image and Graphics, 28(7): 1927-1964 (江俊君, 程豪, 李震宇, 刘贤明, 王中元. 2023. 深度学习视频超分辨率技术综述. 中国图象图形学报, 28(7): 1927-1964) [DOI: 10.11834/jig.220130]

Jimenez J, Echevarria J I, Sousa T and Gutierrez D. 2012. SMAA: enhanced subpixel morphological antialiasing. Computer Graphics Forum, 31(2pt1): 355-364 [DOI: 10.1111/j.1467-8659.2012.03014.x]

Jin Y T, Song H H and Liu Q S. 2022. Super-resolution video frame reconstruction through lightweight attention constraint alignment network. Journal of Image and Graphics, 27(10): 2984-2993 (靳雨桐, 宋慧慧, 刘青山. 2022. 轻量级注意力约束对齐网络的视频超分重建. 中国图象图形学报, 27(10): 2984-2993) [DOI: 10.11834/jig.210345]

Johnson J, Alahi A and Li F F. 2016. Perceptual losses for real-time style transfer and super-resolution//Proceedings of the 14th European Conference on Computer Vision. Amsterdam, the Netherlands: Springer: 694-711 [DOI: 10.1007/978-3-319-46475-6_43]

Lim B, Son S, Kim H, Nah S and Lee K M. 2017. Enhanced deep residual networks for single image super-resolution//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, USA: IEEE: 1132-1140 [DOI: 10.1109/CVPRW.2017.151]

Lottes T. 2009. FXAA [EB/OL]. [2025-07-14]. https://developer.download.nvidia.com/assets/gamedev/files/sdk/11/FXAA_WhitePaper.pdf

Meinds K, Stout J and van Overveld K. 2003. Real-time temporal antialiasing for 3D graphics//Proceedings of 2023 Vision, Modeling, and Visualization Conference (VMV 2003). Munchen: Germany: AKA GmbH: 337-344

Microsoft Corporation. 2017. Multisample Rendering [EB/OL]. [2025-07-14]. https://learn.microsoft.com/zh-cn/previous-versions/windows/drivers/display/multisample-rendering

Reshetov A. 2009. Morphological antialiasing//Proceedings of 2009 Conference on High Performance Graphics. New Orleans, USA: ACM: 109-116 [DOI: 10.1145/1572769.1572787]

Sajjadi M S M, Vemulapalli R and Brown M. 2018. Frame-recurrent video super-resolution//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 6626-6634 [DOI: 10.1109/CVPR.2018.00693]

Wang Z, Bovik A C, Sheikh H R and Simoncelli E P. 2004. Image quality assessment: from error visibility to structural similarity. IEEE Transactions on Image Processing, 13(4): 600-612 [DOI: 10.1109/TIP.2003.819861]

Xiao L, Nouri S, Chapman M, Fix A, Lanman D and Kaplanyan A. 2020. Neural supersampling for real-time rendering. ACM Transac-

tions on Graphics, 39(4): #142 [DOI: 10.1145/3386569.3392376]

Zhang R, Isola P, Efros A A, Shechtman E and Wang O. 2018. The unreasonable effectiveness of deep features as a perceptual metric// Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition.586-595 [DOI:10.1109/CVPR.2018.00068]

Zhong Z H, Chen G L, Wang R and Huo Y C. 2023a. Neural super-resolution in real-time rendering using auxiliary feature enhancement. Journal of Database Management, 34(3): 1-13 [DOI: 10.4018/JDM.321544]

Zhong Z H, Zhu J S, Dai Y X, Zheng C K, Chen G L, Huo Y C, et al. 2023b. FuseSR: super resolution for real-time rendering through efficient multi-resolution fusion//Proceedings of 2023 SIGGRAPH Asia Conference Papers. Sydney, Australia: ACM: #8 [DOI: 10.1145/3610548.3618209]

作者简介

李琳，女，副教授，主要研究方向为虚拟现实与人机交互。E-mail：lilin_julia@hfut.edu.cn

赵洋，通信作者，男，教授，主要研究方向为图像重建与图像处理。E-mail：yzhao@hfut.edu.cn

薛皓文，男，硕士研究生，主要研究方向为实时渲染超采样。E-mail：2908903586@qq.com

朱纪春，男，硕士研究生，主要研究方向为图像重建与美学评估。E-mail：1159238128@qq.com

刘晓平，男，教授，主要研究方向为计算机图形学与辅助设计。E-mail：lxp@hfut.edu.cn