面向梯度域路径跟踪的时序网络重建方法

马鸣聪，徐延宁*，王璐，孟祥旭

(山东大学软件学院　济南　250101)
(xyn@sdu.edu.cn)

摘　要：针对现有的梯度域路径跟踪方法大多关注单帧的重建，缺乏帧间信息的重用，导致渲染结果帧间一致性难以保持的问题，提出一种结合历史帧信息的梯度域网络重建方法。该方法使用运动向量将历史帧信息映射到当前帧，通过重建网络实现高质量的帧图像重建；在重建网络中，设计了基于时序信息的损失函数项，实现在重建结果的空间可靠性和时间稳定性上的平衡；进而，设计了重加权模块，实现帧内重建的结果在帧间的合理传递，在提升视觉质量的同时，提升了时序连续性。在 Bedroom、Classroom 和 Bathroom 等多个测试场景中进行实验的结果表明，与主流的梯度域路径追踪网络重建方法 GradNet 相比，所提方法均获得更高的单帧重建质量，可在有效地消除噪声的同时保留更多的光照细节，并且有效地消除了连续帧的闪烁问题，RelMSE 降低 14.28%～32.56%，LPIPS 误差最多降低 50.95%。

关键词：梯度域渲染；机器学习；无监督网络；时序重建
中图法分类号：TP391.41　　**DOI：**10.3724/SP.J.1089.2025-00371

Temporal Reconstruction Network for Gradient-Domain Path Tracing

Ma Mingcong, Xu Yanning*, Wang Lu, and Meng Xiangxu

(*School of Software, Shandong University, Jinan*　250101)

Abstract: Aiming at the problem that most existing gradient domain path tracking methods focus on intra-frame reconstruction and lack the reuse of inter-frame information, resulting in difficulty in maintaining inter-frame consistency in rendering results, a gradient domain network reconstruction method combining historical frame information is proposed to address this issue. This method maps historical frame information to the current frame through motion vectors, and then achieves high-quality image reconstruction through the reconstruction network. In the reconstruction of the network, a loss function term based on temporal information was designed to achieve a balance between spatial reliability and temporal stability of the reconstruction results. Furthermore, a reweighting module was designed to achieve reasonable transmission of intra-frame reconstruction results between frames, improving both visual quality and temporal continuity. The experimental results show that compared with the mainstream gradient domain path tracking network reconstruction method GradNet, the proposed method achieves higher single frame reconstruction quality in multiple test scenes such as Bedroom, Classroom, and Bathroom. It can effectively eliminate noise while preserving more lighting details, and effectively eliminate the flicker problem of continuous frames. The RelMSE is reduced by 14.28% to 32.56%, and the LPIPS error is reduced by up to 50.95%.

Key words: gradient-domain rendering; machine learning; unsupervised network; temporal reconstruction

收稿日期：2025-09-09；修回日期：2025-12-05。**基金项目：**国家重点研发计划（2022YFB3303203）；国家自然科学基金（62272275）。
马鸣聪（1994—），男，博士研究生，主要研究方向为渲染图像重建；**徐延宁**（1975—），男，博士，副教授，硕士生导师，CCF 会员，论文通信作者，主要研究方向为计算机图形学；**王璐**（1981—），女，博士，教授，博士生导师，CCF 会员，主要研究方向为计算机图形学；**孟祥旭**（1962—），男，博士，教授，博士生导师，CCF 会员，主要研究方向为计算机图形学。

蒙特卡罗路径追踪方法因其灵活性和多样性在许多领域得到了广泛应用，但这些方法也存在收敛速度慢、在低每像素采样次数（samples-per-pixel，SPP）下图像噪声显著，以及在高 SPP 下资源消耗巨大等问题。梯度域渲染方法[1]使用空间相关性缓解低 SPP 下图像噪声过大的问题，其将渲染器的渲染结果作为输入，屏蔽泊松求解器，利用梯度信息约束渲染结果进行迭代拟合，重建出更高质量的图像。基于这项工作，越来越多的帧内辅助特征通道被用来进一步约束重建过程。

一些经过良好训练的网络模型[2-4]在重建质量和处理效率方面都表现出足够好的性能，取代了屏蔽泊松求解器。GradNet[2]降低了训练网络的成本，使得深度神经网络成为一种更具竞争力的方法。然而，现有的基于神经网络的梯度域方法未考虑使用帧间信息维持时间上的稳定性。Manzi 等[5]在梯度域渲染中使用帧间信息进行图像重建，在渲染连续的动画帧时额外保存了运动信息，可以整合几个相邻帧的渲染结果为单个帧的重建服务。该方法利用相邻帧的信息减少动画序列中的闪烁问题，然而其仍然存在噪声和鬼影，这是因为运动向量的求解并不是完全可靠的，偶尔的求解误差和失败映射导致帧间的时序信息丢失或被错误使用。

连续帧采样结果的时序复用对提升重建质量有重要的作用，但是简单地基于运动向量的时序复用结果是不够可靠的[5]。因此，如何高效且准确地对像素采样结果进行时序复用，并在像素映射后将历史帧数据与当前帧数据的以恰当的权重进行融合，仍然是一个挑战。

本文构建了一个基于时序的反馈网络，通过该网络，历史帧的数据可以辅助当前帧的重建，并将完成的重建结果整合到新的历史帧中辅助后续帧的重建；然后设计结合数据损失、梯度损失、一阶损失和时序损失的综合损失函数，以提高帧间连续性；进而设计重加权过程，稳定历史帧结果，从而更合理地利用历史帧信息，减少鬼影现象。实验结果表明，无论在摄像机运动下的连续动画帧中，还是在摄像机静止下的连续动画帧中，本文方法在重建结果的质量和时间一致性方面都优于现有方法。

1 相关工作

1.1 梯度域渲染和图像重建

梯度域渲染方法除了使用颜色信息外，还引入空间信息帮助渲染结果收敛，空间信息也称为图像梯度。在 Lehtinen 等[6]提出梯度域梅特罗波利斯光传输的开创性工作之后，Kettunen 等[7]将梯度域方法扩展到路径跟踪中；其他相关技术与梯度域方法的结合也带来了不同的优势，如 Manzi 等[1]使用双向路径跟踪以优化光线路径。

在梯度域渲染中，图像重建通常需要使用 Bhat 等[8]提出的屏蔽泊松求解器或 Rousselle 等[9]提出的基于迭代的优化方法。为了获得更好的结果，后续工作通常使用额外的辅助特征作为新的约束，但同时会增加时间成本和内存开销[10-12]。

最近，Back 等[13]也使用梯度域重构的想法设计了一个无偏去噪核，而现有的有偏去噪器需要在偏差和方差之间取得平衡；Yan 等[14]采用滤波而非网络的方法首先对图像进行多尺度的重建，然后使用引导上采样对多尺度重建结果进行合并。

1.2 用于重建的深度学习

深度学习在计算机视觉等领域得到了很好的应用，在渲染中也进行了相当丰富的探索。文献[15-18]将深度网络应用于传统的蒙特卡罗路径跟踪图像去噪，取得了积极的效果。相比之下，深度网络在梯度域渲染中的应用相对较少，据本文调研，Kettune 等[3]的工作首次将深度学习与梯度域渲染相结合，并取得了超越传统蒙特卡罗路径跟踪图像去噪的效果，但是，这项工作也使用监督网络来实现去噪。Guo 等[2]提出首个取代梯度域渲染中屏蔽泊松求解器的无监督深度学习解决方案。无监督网络比传统的屏蔽泊松重建取得了更好的结果，且与有监督网络的质量相当。Xu 等[4]提出一种用于梯度域体光子密度估计的无监管卷积神经网络。

在此基础上，本文在无监督网络中添加时序信息，进一步优化网络的重建能力。据本文调研，这是首个使用时序信息进行梯度域渲染重建的无监督网络。

1.3 时序重建

引入时序信息可以在更高维度上有效地使用数据，获得比单独使用空间信息获得更好的结果。众所周知，时间抗锯齿技术[19-20]是通过将样本从单个时间帧扩展到几个连续的时间帧，并对不同帧的采样结果进行加权实现的，该技术不需要额外的时间成本，非平均加权机制和运动向量在一定程度上克服了闪烁和鬼影，因此具有广泛的应用，相关思想也影响了连续帧渲染的其他领域。

具有更高维度的时空信息也提供了更有效的

重建操作。Schied 等[21]在空间尺度上添加时序信息，并在 1 SPP 的超低采样次数下取得了显著的去噪效果。本文同样使用运动向量作为对齐不同帧数据的基础。由于运动向量可以更高效地利用时间信息，且额外的时间开销很小，因此 Zeng 等[22]提供的更可靠的运动向量具有相当重要的意义。Manzi 等[5]将时序信息引入到梯度域渲染中提出 TGPT，相较传统方法取得了更优的渲染质量。

2　研究背景

梯度域渲染的核心思想是使用基本路径高度相关的偏移路径估计相邻像素之间的有限差分，这些有限差分通常称为梯度。Manzi 等[1]将求解方程定义为

$$I=\arg\min_{\hat{I}}\left\|\alpha(\hat{I}-I_{\mathrm{b}})\right\|^{n}+\left\|\begin{pmatrix}H_{\mathrm{dx}}\hat{I}\\H_{\mathrm{dy}}\hat{I}\end{pmatrix}-\begin{pmatrix}I_{\mathrm{dx}}\\I_{\mathrm{dy}}\end{pmatrix}\right\|^{n}\quad(1)$$

其中，α 表示用于控制梯度和颜色之间的权重；I_{b} 表示梯度域渲染器生成的颜色图像通道；I_{dx} 和 I_{dy} 分别表示 I_{b} 的水平像素梯度和垂直像素梯度；H_{dx} 和 H_{dy} 表示水平和垂直方向上的有限差分算子。最终，输出图像 I 最小化了其与输入图像的像素差分和梯度差分。

在图像重建过程中，使用 L_2 范数时结果是无偏的，但它通常不能很好地处理异常数据，视觉体验不佳。虽然使用 L_1 范数重建的结果在视觉上要好得多，但它是有偏的[6-7]。

在涉及时间序列的梯度域渲染中，对式（1）增加了更多约束条件[3]。因此，时间序列下的求解方程为

$$I=\arg\min_{\hat{I}}\left\|\alpha(\hat{I}-I_{\mathrm{b}})\right\|^{n}+\left\|\begin{pmatrix}H_{\mathrm{dx}}\hat{I}\\H_{\mathrm{dy}}\hat{I}\\H_{\mathrm{d}t}\hat{I}\\H_{\mathrm{dx}\mathrm{d}t}\hat{I}\\H_{\mathrm{dy}\mathrm{d}t}\hat{I}\end{pmatrix}-\begin{pmatrix}I_{\mathrm{dx}}\\I_{\mathrm{dy}}\\I_{\mathrm{d}t}\\I_{\mathrm{dx}\mathrm{d}t}\\I_{\mathrm{dy}\mathrm{d}t}\end{pmatrix}\right\|^{n}。$$

其中，$H_{\mathrm{d}t}$ 表示相邻帧之间的时间有限差分算子；$H_{\mathrm{dx}\mathrm{d}t}$ 和 $H_{\mathrm{dy}\mathrm{d}t}$ 表示混合时空有限差分算子，它们是时间算子和空间算子的简单连接，$H_{\mathrm{dx}\mathrm{d}t}=H_{\mathrm{dx}}H_{\mathrm{d}t}$ 和 $H_{\mathrm{dy}\mathrm{d}t}=H_{\mathrm{dy}}H_{\mathrm{d}t}$；与 I_{dx} 和 I_{dy} 类似，$I_{\mathrm{d}t}$，$I_{\mathrm{dx}\mathrm{d}t}$ 和 $I_{\mathrm{dy}\mathrm{d}t}$ 分别表示 I_{b} 的时间梯度、时空水平梯度和时空垂直梯度。

3　本文方法

3.1　问题定义

本文为当前帧和历史帧添加约束，并使用重加权过程隐式地调整 2 帧之间的权重。定义具有可训练参数 θ 的网络模型为

$$W\left\{I_{\mathrm{b}}^{i},I_{\mathrm{w}}^{i},I_{\mathrm{dx}}^{i},I_{\mathrm{dy}}^{i},\boldsymbol{F}^{i}\left(\boldsymbol{N}^{i},\boldsymbol{D}^{i},\boldsymbol{A}^{i}\right);\theta\right\}\rightarrow I_{i}。$$

其中，I_{b}^{i} 表示直接渲染的原始图像；I_{w}^{i} 表示来自 I_{i-1} 的历史帧；I_{dx}^{i} 表示水平方向梯度，I_{dy}^{i} 表示垂直方向梯度；$\boldsymbol{F}^{i}$ 是所有辅助通道的集合，包括法线 $\boldsymbol{N}^{i}$，深度 $\boldsymbol{D}^{i}$ 和反射率 $\boldsymbol{A}^{i}$，如图 1 所示。最终，第 i 帧的重建结果 I_{i} 通过网络模型 W 推断得到。本文方法的流程如图 2 所示。

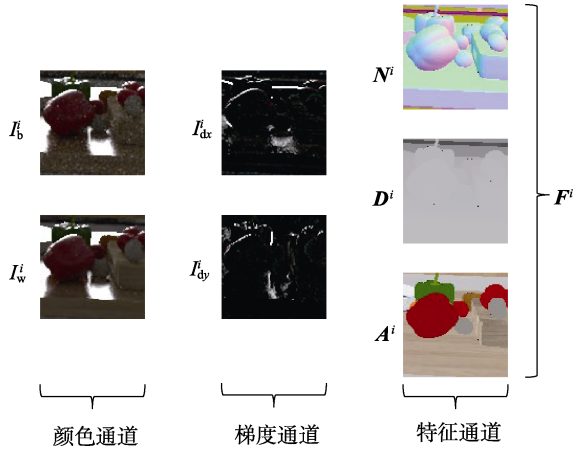


图 1　重建网络的输入通道

3.2　网络架构

本文网络架构如图 3 所示。网络处理后，当前帧的重建结果通过反馈机制进行时间重投影后成为下一帧输入数据的一部分。由于多分支网络结构的可靠性已经得到验证[2]，因此本文方法同样遵循类似的结构。

在 W 完成前一帧训练后，输出的 I_{i-1} 使用运动向量进行时间重投影为 I_{w}^{i}，从而形成反馈机制。I_{w}^{i} 和 I_{b}^{i} 使用 $\boldsymbol{D}^{i}$ 进行重加权，形成新的变量 I_{r}^{i}，最终将其反馈到网络中。随着网络的逐步优化，越来越稳定的 I_{i} 通过反馈机制影响下一帧的结果，直至迭代优化整个网络参数。

为了保证反馈机制在不破坏网络稳定性的前提下引入重投影结果，本文对网络入口处的卷积结构进行调整，允许网络接受额外的数据维度。鉴于已有的基于运动向量的重映射的工作[19-22]均考

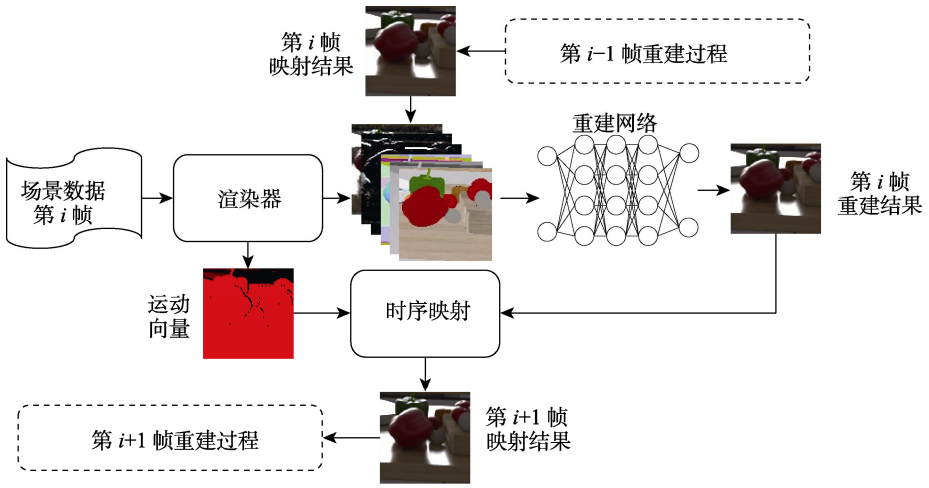


图 2 本文方法流程图

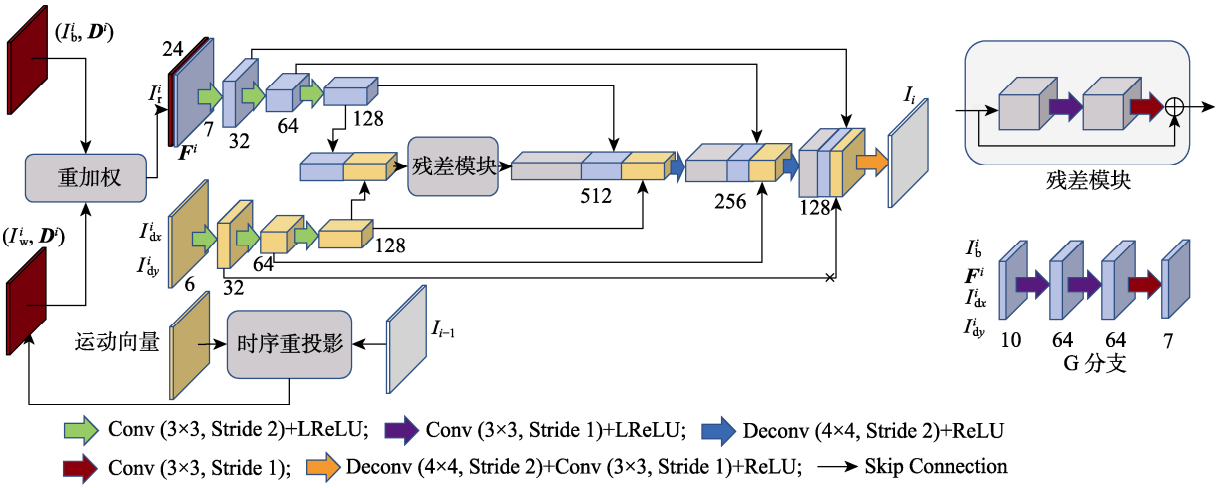


图 3 本文网络架构

虑了运动向量本身的可靠性，本文也采取相应措施提升反馈的稳定性。如果像素的运动向量丢失或重映射到合法范围之外，则丢弃像素的反馈结果；当像素的命中点是镜面反射时，使用其下一个非镜面反射命中点的运动向量作为合法运动向量。

每个分支的结构类似于 U-Net[23]，多分支在上采样阶段前通过残差结构连接。数据分支旨在从噪声图像中提取低频内容，因此本文将时间重投影获得的历史帧引入这个分支中。虽然历史帧的质量高于直接渲染的图像，但是随着网络逐渐完成训练，两者的质量逐渐接近。由于梯度分支用于提取高频细节，因此只有梯度图像参与训练。如图 3 所示，G 分支使用简单的多层卷积结构，训练结果作为训练网络所需的参数。

3.3 损失函数

本文方法侧重于动画帧之间的连续性，因此损失函数需要对时间连续性进行额外设计。基于表现良好的损失函数设计[2]，本文通过增加时序损失，在帧内的重建质量和连续帧差异之间取得平衡。

时序损失采用基于 VGG16 网络[24]的感知损失方案，并提取第 3，8，15 和 22 层的输出特征图作为比较的数据源；经过 BatchNorm 操作后，计算该数据源的均方误差（mean-square error，MSE），形成最终的时序损失 L_{Temp}。

本文中，网络训练的最终损失函数由数据损失 L_{Data}、梯度损失 L_{Grad}、一阶损失 L_{1st} 和时序损失 L_{Temp} 4 部分组成。总损失函数为

$$L_{\text{Total}} = L_{\text{Data}} + L_{\text{Grad}} + \alpha \cdot L_{\text{1st}} + \beta \cdot L_{\text{Temp}}。$$

其中，L_{Data}、L_{Grad} 和 L_{1st} 采用与 GradNet[2]类似的设计；权重参数 α 将随着训练周期的增加而缓慢增加，最终稳定在 2；权重参数 β 用于调整帧内与帧间的质量平衡，经过多次网络参数调整和尝试后，本文最终选择 β=0.95。

3.4　重加权过程

在输入端，本文网络从不同来源获得颜色通道的 2 个版本：（1）直接来自当前帧，其噪声相对较大；（2）从先前帧的重投影生成的历史帧，由于在先前帧已经完成了重建，因此历史帧的噪声相对较小。

为了有效地组合 2 个颜色通道，本文参考传统的时间抗锯齿技术，并设置一个加权参数对当前帧和历史帧颜色通道进行加权。但是，简单的全局统一权重参数是不合理的，像素自适应的权重参数同样表现不佳。最终，与 Xiao 等[25]的方法类似，本文基于网络隐式地重加权当前帧和历史帧，而不是使用显式的权重参数。

本文的重加权过程如图 4 所示。其中，不同的重加权模块用相似的方式分别处理 $(I_{\mathrm{b}}^{i}, \boldsymbol{D}^{i})$ 和 $(I_{\mathrm{w}}^{i}, \boldsymbol{D}^{i})$，形成最终的重加权特征，再由图 3 中的主网络进行处理；引入深度信息 $\boldsymbol{D}^{i}$ 可以有效地帮助颜色通道进行特征和权重提取，更稳定地将历史帧数据与当前帧结合。

虽然本文使用一个小型网络重加权历史帧和当前帧，但是该网络并不独立于主网络，因此本文网络仍然是端到端的，不需要额外的单独训练。

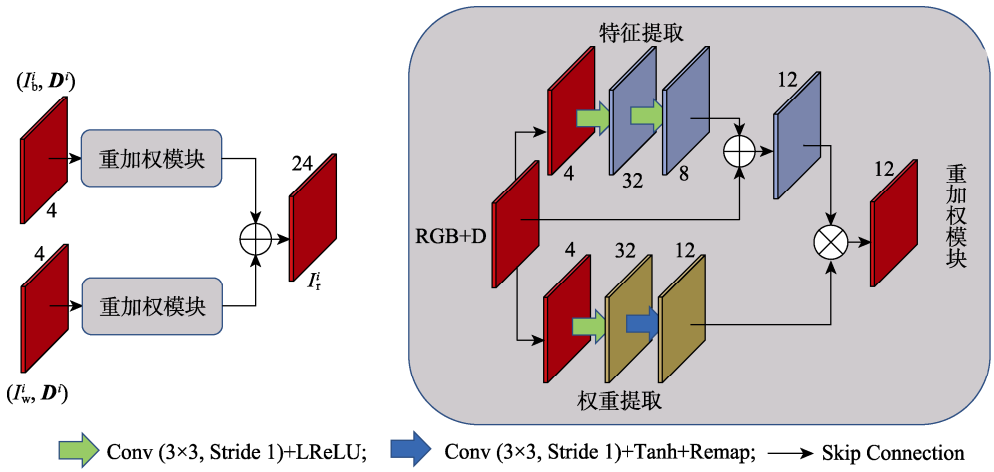


图 4　重加权过程

3.5　数据集准备

由于无监督网络在训练时不需要真实图像，因此只需要生成相对低质量的训练样本。本文从场景集[26]中选择了 8 个场景，并以 64 SPP 的采样次数为每个场景生成了 100 个连续帧，最终获得了 800 帧图像（1 280×720 像素）；每帧使用扩展后的梯度域路径追踪渲染器[5]，以 64 SPP 渲染颜色通道和相应的梯度通道，以及法线、深度、反射率和运动向量 4 个额外的缓冲区。其中，法线通道记录路径上的第 1 个非镜面命中点，数值归一化到[0,1]；深度通道同样记录命中点的位置，并使用场景位置信息完成归一化到[0,1]；反射率通道记录相同命中点的位置，并归一化到[0,1]；运动向量基于与 Schied 等[21]相似的方法实现。法线、深度和反射率通道在进入网络之前均已归一化。为了增加训练样本的数量，本文随机选择 30 组样本子集，每组包含 100 个样本（256×256 像素）；同一组中，所有样本在基础图像中具有相同的像素坐标。在极少数情况下，数据集中会出现非数值（NaN）异常，本文将这些异常值强制设置为 0。需要注意的是，每帧的运动向量都需要重新排序以便于后续的网络训练，保证历史帧可以直接从当前帧的数据直接推导出来，如图 3 所示。

本文从场景集[26]中选择 Bedroom、Classroom 和 Bathroom 这 3 个未用于生成训练集的场景，并在不同的 SPP 下对它们进行渲染以构成验证集。此外，还额外选择了 Livingroom 和 Diningroom 这 2 个未用于生成训练集的场景，但仅在特定 SPP 下渲染，作为验证集的补充。

3.6　网络训练

本文网络基于 PyTorch 实现，使用小批量 SGD（stochastic gradient descent）进行训练，并应用 Adam 求解器[27]。基础学习率设为 0.000 1，并在每个训练周期完成后将其乘以 0.95；网络权重使用 He 等[28]的方法进行初始化；训练样本的 90%作为训练集，剩余的 10%用作测试集。

4 实验结果与分析

为了评估本文方法，首先从运行效率和单帧图像质量的角度将其与现有方法进行比较；然后评估多帧连续图像的质量；最后通过消融实验，分析该方法中各种设计对最终重建结果的影响。实验中，本文使用 NVIDIA GTX 4090，所有图像使用统一的 1 280×720 像素分辨率。

本文使用光线追踪方法以 81 920 SPP 渲染图像作为真实图像，并用这些真实图像构建验证集；这些真实图像仅在验证集中作为判断各种方法重建质量的依据，不会出现在训练集中。

4.1 运行效率

首先在单帧图像上，将本文方法与现有方法的运行效率进行对比实验，包括 GPT[1]、TGPT[5]、GradNet[2]和 IDUW[13]，所有图像的平均运行时间如表 1 所示。其中，GPT[1]是经典方法，使用屏蔽泊松求解器且只考虑单帧图像；TGPT[5]中引入了时间变量，但同样使用屏蔽泊松求解器；GradNet[2]使用无监督网络而不是屏蔽泊松求解器进行重建，但也只考虑单帧图像；IDUW[13]使用输入相关核无偏地对图像进行去噪，但该方法需要 4 个独立的输入。可以看出，本文方法的时间消耗与现有的重建方法保持在相同的数量级；作为去噪方法，IDUW[13]比重建方法耗时明显更多；与同样使用时间信息的 TGPT[5]相比，本文方法在时间方面更具优势。虽然本文方法比 GradNet[2]和 GPT[1]方法稍慢，但这 2 种方法未利用时间信息。考虑本文方法不面向实时或交互式渲染，这一时间成本是可以接受的。

表 1 在单帧图像上 5 种方法的运行时间对比 ms

方法	时间	方法	时间
GPT[1]	211.9	IDUW[13]	2 183.0
TGPT[5]	368.4	本文	277.1
GradNet[2]	263.1		

4.2 对比实验

4.2.1 与重建方法对比

将本文方法与 GPT[1]、TGPT[5]和 GradNet[2]方法的重建能力进行对比实验。

图 5 所示为从未出现在训练集中的场景，本文使用这些新数据来更好地说明网络的泛化能力。可以看出，与现有方法相比，本文方法在最终重建结果中噪声更少，保留了更多的细节，主观上的视觉效果也更好。

在图 5 所示的 3 个场景中，将本文方法与 3 种对比方法进行对比实验，RelMSE（relative MSE）、结构相似性指数（structural similarity index measure，SSIM）、可学习感知图像块相似度（learned perceptual image patch similarity，LPIPS）和视频多方法评估融合（video multi-method assessment fusion，VMAF）的结果如表 2 所示。可以看出，本文方法的 RelMSE 和 SSIM 优于对比方法，说明其在像素级和结构级重建上表现优异；仅在 Bedroom 场景中，GradNet[2]的 LPIPS 略优于本文方法，但本文方法在其他场景中表现更好，表明其在感知或视觉维度上同样具有优势；本文方法在 VMAF 指标上也表现更优，该指标从视频连贯性的角度评估了连续帧的质量。

4.2.2 与去噪方法对比

将本文方法与基于梯度域的去噪方法 IDUW[13]进行对比实验，结果如表 3 和表 2 所示。可以看出，与 IDUW[13]相比，在任意 SPP 下，本文方法都具有更好的可视化效果；在较低的 SPP 下，本文方法的 RelMSE 最低，但在较高的 SPP 下，其 RelMSE 优势不明显；在场景 3 中，IDUW[13]的 VMAF 稍微高于本文方法，但综合考虑输入限制和时间代价，这个结果是可以接受的。

4.2.3 连续帧重建结果对比

鉴于本文方法和 TGPT[5]都使用时序数据，因此针对运动向量引起的鬼影问题额外进行了一些实验。在 Diningroom 场景中，将本文方法与 TGPT[5]进行对比实验，结果如图 6 所示。可以看出，TGPT[5]重建的区域存在严重的偏色问题和黑色像素区域；本文方法在相同区域上更加接近真实图像，且没有黑色像素区域。实验结果表明，本文方法在时序数据的使用上更加可靠，并有效地减少了重影鬼影的产生。

本文通过固定摄像机位置并连续渲染多帧相同内容，对随着时间的积累，不同方法在重建结果质量上的变化进行实验。4 种方法不同帧数的 RelMSE 对比如图 7 所示。可以看出，随着帧数的增加，GPT[1]的结果出现了较大的误差波动；相比之下，TGPT[5]和 GradNet[2]方法的误差相对稳定；本文方法的误差在前几帧后显著降低，并在后续帧中持续保持较低水平。尽管本文方法与 TGPT[5]都



a. 场景1. Bedroom（16 SPP）

b. 场景2. Classroom（64 SPP）

c. 场景3. Bathroom（256 SPP）

图 5　在 3 个场景中 5 种方法的重建结果对比

表 2　在 3 个场景中 5 种方法的性能对比

场景	方法	RelMSE ↓	SSIM ↑	LPIPS ↓	VMAF ↑
1	输入	0.033 4	0.972 9	0.131 9	
	GPT[1]	0.018 0	0.990 1	0.017 9	81.668 1
	TGPT[5]	0.015 4	0.991 8	0.013 8	82.400 6
	GradNet[2]	0.012 6	0.992 7	**0.010 6**	88.629 6
	IDUW[13]	0.012 2	0.993 0	0.013 7	92.011 6
	本文	**0.010 8**	**0.993 1**	0.010 8	**92.478 3**
2	输入	0.100 7	0.865 6	0.252 8	
	GPT[1]	0.036 1	0.969 1	0.036 1	64.843 5
	TGPT[5]	0.031 2	0.975 9	0.030 4	73.146 0
	GradNet[2]	0.028 8	0.978 4	0.029 1	78.038 4
	IDUW[13]	0.031 6	0.976 4	0.033 3	75.421 3
	本文	**0.022 3**	**0.979 0**	**0.028 8**	**84.844 9**
3	输入	0.024 4	0.975 9	0.144 3	
	GPT[1]	0.007 6	0.991 9	0.087 9	72.446 0
	TGPT[5]	0.006 2	0.996 2	0.052 3	72.932 5
	GradNet[2]	0.004 3	0.998 3	0.045 7	79.978 4
	IDUW[13]	**0.002 9**	0.988 4	**0.022 4**	**84.715 3**
	本文	**0.002 9**	**0.998 7**	**0.022 4**	82.434 9

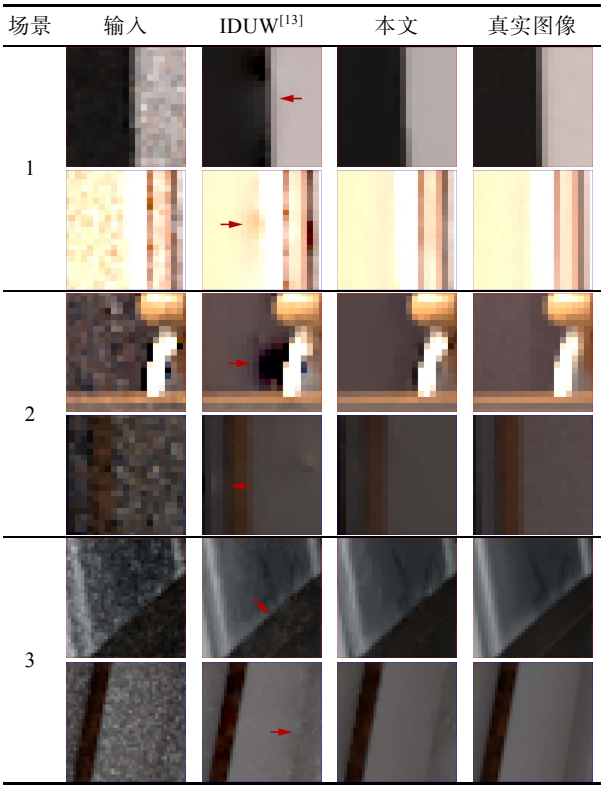
注：粗体表示最优值。

使用时序数据，但是 TGPT[5]需要在渲染完成后通过组合前后多帧合成单个帧，而本文方法只使用历史帧。因此，TGPT[5]没有表现出随着帧数的增加误差逐渐降低的特性，而本文方法中这种下降趋势较为明显。

4.3　消融实验

在基础的时序反馈机制网络上，通过逐步增加时序损失和重加权过程进行消融实验，比较各设计对重建结果的影响，并分别测试这些网络的重建能力。在 64 SPP 下的 2 个场景中，消融实验结果如表 4 所示。可以看出，当简单地引入时序反馈机制时，漫反射表面出现明显的模糊，在镜面反射中则会有明显的重影鬼影残留；当增加时序损失后，漫反射表面上的模糊问题被大致消除，镜面反射的重影鬼影也有所减弱，对应的 RelMSE 分别减少 58.28%和 23.5%；增加重加权后，镜面反射中的重影鬼影得到了进一步消除，对应的 RelMSE 评分分别减少 19.12%和 13.73%。

表 3　2 种方法的重建结果对比



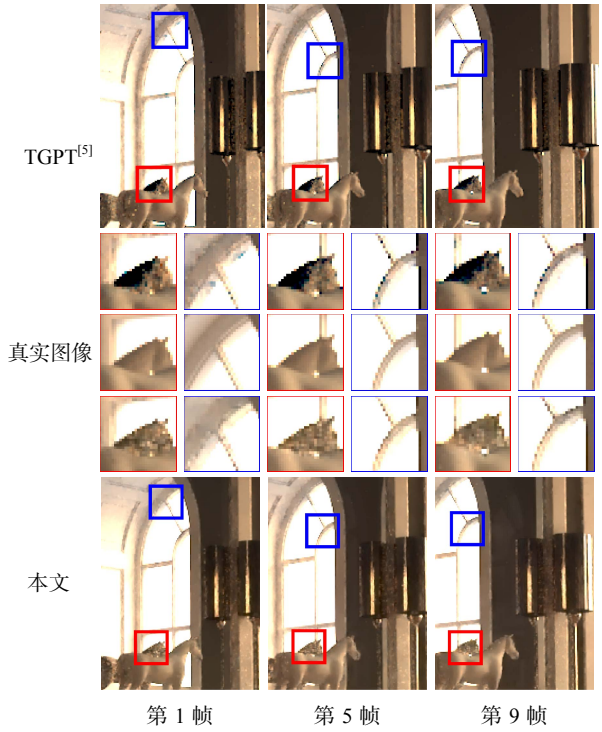


图 6　在 Diningroom 场景中 2 种方法的连续帧重建结果对比

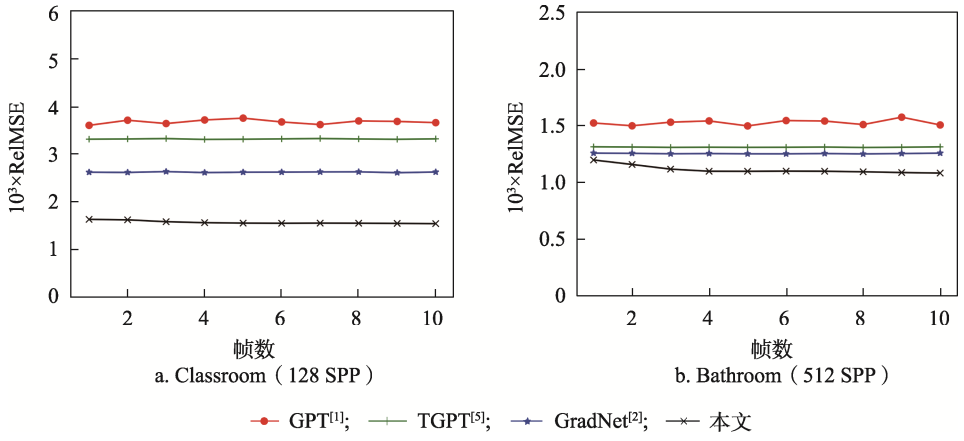
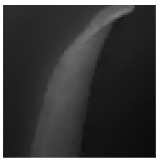
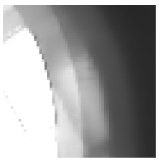
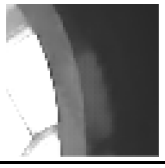
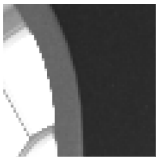


图 7　2 个场景中帧数增加时 4 种方法 RelMSE 对比

表 4　在 64 SPP 下 2 个场景的消融实验结果

方法	时间反馈机制	时序损失	重加权	Diningroom		Livingroom	
				重建结果	RelMSE	重建结果	RelMSE
本文-F	√				0.016 3		0.020 0
本文-FT	√	√			0.006 8		0.015 3

续表

方法	时间反馈机制	时序损失	重加权	Diningroom		Livingroom	
				重建结果	RelMSE	重建结果	RelMSE
本文	√	√	√		0.005 5		0.013 2
真实图像							

5　结　语

本文提出一种面向梯度域路径跟踪的时序网络重建方法。实验结果表明，无监督深度网络和时序重建方法相结合，可以有效地重建梯度域路径跟踪方法的渲染结果，获得优于文中对比方法的图像质量；此外，引入时序损失和重加权过程是必要的，两者都有助于提高最终的图像质量和时间稳定性。

本文方法的局限性主要是场景的镜面反射区域中，相关区域的噪声和重影鬼影无法完全消除，而这通常是由运动向量导致的。未来，将进一步简化网络结构，或将现有的整体端到端网络拆分为多个单独的子网络，以优化时间性能。

参考文献（References）：

[1] Manzi M, Kettunen M, Aittala M, *et al*. Gradient-domain bidirectional path tracing[C] //Proceedings of Eurographics Symposium on Rendering - Experimental Ideas & Implementations. Aire-la-Ville: Eurographics Association Press, 2015: 65-74

[2] Guo J, Li M T, Li Q W, *et al*. GradNet: unsupervised deep screened Poisson reconstruction for gradient-domain rendering[J]. ACM Transactions on Graphics, 2019, 38(6): Article No.223

[3] Kettunen M, Härkönen E, Lehtinen J. Deep convolutional reconstruction for gradient-domain rendering[J]. ACM Transactions on Graphics, 2019, 38(4): Article No.126

[4] Xu Z L, Sun Q, Wang L, *et al*. Unsupervised image reconstruction for gradient-domain volumetric rendering[J]. Computer Graphics Forum, 2020, 39(7): 193-203

[5] Manzi M, Kettunen M, Durand F, *et al*. Temporal gradient-domain path tracing[J]. ACM Transactions on Graphics, 2016, 35(6): Article No.246

[6] Lehtinen J, Karras T, Laine S, *et al*. Gradient-domain metropolis light transport[J]. ACM Transactions on Graphics, 2013, 32(4): Article No.95

[7] Kettunen M, Manzi M, Aittala M, *et al*. Gradient-domain path tracing[J]. ACM Transactions on Graphics, 2015, 34(4): Article No.123

[8] Bhat P, Curless B, Cohen M, *et al*. Fourier analysis of the 2D screened Poisson equation for gradient domain problems[C] //Proceedings of the 10th European Conference on Computer Vision. Heidelberg: Springer, 2008: 114-128

[9] Rousselle F, Jarosz W, Novák J. Image-space control variates for rendering[J]. ACM Transactions on Graphics, 2016, 35(6): Article No.169

[10] Manzi M, Vicini D, Zwicker M. Regularizing image reconstruction for gradient-domain rendering with feature patches[J]. Computer Graphics Forum, 2016, 35(2): 263-273

[11] Back J, Yoon S E, Moon B. Feature generation for adaptive gradient-domain path tracing[J]. Computer Graphics Forum, 2018, 37(7): 65-74

[12] Ha S, Oh S, Back J, *et al*. Gradient outlier removal for gradient-domain path tracing[J]. Computer Graphics Forum, 2019, 38(2): 245-253

[13] Back J, Hua B S, Hachisuka T, *et al*. Input-dependent uncorrelated weighting for Monte Carlo denoising[C] //Proceedings of the SIGGRAPH Asia 2023 Conference Papers. New York: ACM Press, 2023: Article No.9

[14] Yan D F, Zheng S K, Yan L Q, *et al*. Filtering-based reconstruction for gradient-domain rendering[C] //Proceedings of the SIGGRAPH Asia 2024 Conference Papers. New York: ACM Press, 2024: Article No.69

[15] Bako S, Vogels T, McWilliams B, *et al*. Kernel-predicting convolutional networks for denoising Monte Carlo renderings[J]. ACM Transactions on Graphics, 2017, 36(4): Article No.97

[16] Chaitanya C R A, Kaplanyan A S, Schied C, *et al*. Interactive reconstruction of Monte Carlo image sequences using a recurrent denoising autoencoder[J]. ACM Transactions on Graphics, 2017, 36(4): Article No.98

[17] Vogels T, Rousselle F, McWilliams B, *et al*. Denoising with kernel prediction and asymmetric loss functions[J]. ACM Transactions on Graphics, 2018, 37(4): Article No.124

[18] Suppan C, Chalmers A, Zhao J H, *et al*. Neural screen space rendering of direct illumination[C] //Proceedings of Pacific Graphics 2021-Short Papers, Posters, and Work-in-Progress Papers. Los Alamitos: IEEE Computer Society Press, 2021: 37-42

[19] Yang L, Nehab D, Sander P V, *et al*. Amortized supersam-

pling[J]. ACM Transactions on Graphics, 2009, 28(5): 1-12

[20] Salvi M. An excursion in temporal supersampling[OL]. [2025-09-09]. https://developer.download.nvidia.cn/gameworks/events/GDC2016/msalvi_temporal_supersampling.pdf

[21] Schied C, Kaplanyan A, Wyman C, *et al*. Spatiotemporal variance-guided filtering: real-time reconstruction for path-traced global illumination[C] //Proceedings of the High Performance Graphics. New York: ACM Press, 2017: Article No.2

[22] Zeng Z, Liu S Q, Yang J L, *et al*. Temporally reliable motion vectors for real-time ray tracing[J]. Computer Graphics Forum, 2021, 40(2): 79-90

[23] Ronneberger O, Fischer P, Brox T. U-Net: convolutional networks for biomedical image segmentation[C] //Proceedings of the 18th International Conference on Medical image computing and computer-assisted intervention. Heidelberg: Springer, 2015: 234-241

[24] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[C] //Proceedings of the 3rd International Conference on Learning Representations. Cambridge: OpenReview Press, 2015: 1-12

[25] Xiao L, Nouri S, Chapman M, *et al*. Neural supersampling for real-time rendering[J]. ACM Transactions on Graphics, 2020, 39(4): Article No.142

[26] Bitterli B. Rendering resources[OL]. [2025-09-09]. https://benedikt-bitterli.me/resources

[27] Kingma D P, Ba J. Adam: a method for stochastic optimization[C] //Proceedings of the 3rd International Conference on Learning Representations. Cambridge: OpenReview Press, 2015: 1-13

[28] He K M, Zhang X Y, Ren S Q, *et al*. Delving deep into rectifiers: surpassing human-level performance on ImageNet classification[C] //Proceedings of the IEEE International Conference on Computer Vision. Los Alamitos: IEEE Computer Society Press, 2015: 1026-1034