

# Real-time Monte Carlo Denoising with the Neural Bilateral Grid

Dezeming Family

2023 年 7 月 1 日

正常字体：表示论文的基本内容解释。

粗体：表示需要特别注意的内容。

红色字体：表示容易理解错误或者混淆的内容。

蓝色字体：表示额外增加的一些注释。

绿色字体：表示额外举的一些例子。

## 目录

一 基本介绍	1
二 相关工作	1
三 问题陈述	1
四 去噪器架构	2
参考文献	2

## 一 基本介绍

我们开发了一种高效的卷积神经网络架构，以学习在数据相关的双边空间中对噪声输入进行去噪。我们的神经网络学习生成一个引导图像，用于首先将有噪声的数据泼溅 (splating) 到网格中，然后对其进行切片 (slicing) 以读取去噪数据。为了将双边网格无缝集成到我们的可训练去噪管道中，我们利用了一种可微分的双边网格，称为神经双边网格，它可以实现端到端训练。此外，我们还展示了如何使用多尺度双边网格的层次结构来进一步提高去噪质量。

我们的实验结果表明，这种方法可以以实时帧速率（每帧几毫秒）对 1-spp 噪声输入图像进行鲁棒去噪。在如此低的采样率下，我们的方法在质量和速度方面都优于基于核预测网络的最先进技术，并且与最先进的特征回归技术相比，它显著提高了质量。

我们方法的核心思想是利用神经双边网格 [2] 进行去噪。我们将有噪声的图像数据拼接 (splat) 到数据相关的双边网格中，并使用神经网络实现图像像素到网格中的数据相关映射。直觉是，与图像本身去噪相比，可以使用更简单的网络来学习【有效的数据相关飞溅到双边网格中】(effective data-dependent splatting into the bilateral grid)，并且通过双边网格去噪非常快，因为其复杂性不取决于所得到的去噪滤波器的支持。

在实践中，我们的网络预测引导图像，以控制图像数据如何飞溅到双边网格中。在双边空间中，通过简单的三角滤波 (tent) 来融合飞溅的样本。由于网格的低分辨率，这有效地对数据进行了去噪。作为最后一步，我们重用神经网络中的引导图像，从双边网格中提取数据，以重建去噪结果，这被称为切片 (slicing)。

值得注意的是，我们将双边网格作为一个可微分组件来实现，这样我们就可以将其与神经网络无缝集成，进行端到端训练。此外，我们提出了一种多尺度双边网格的层次结构，在略微增加计算成本的情况下，进一步提高了去噪质量。最后，为了对视频中的连续帧进行去噪，我们应用时间增强 (temporal accumulation) 步骤来确保时间稳定性。

受 [1] 的启发，我们在流水线中应用连续帧的时序累积，以提高有效采样率并抑制时序伪影。根据综合实验，我们的轻量级神经网络利用神经双边网格，能够以实时帧率（每帧几毫秒）对有噪声的输入帧进行去噪，同时为 1-spp 数据提供与最先进的实时去噪方法相当甚至更高的质量。此外，通过更深的神经网络架构，我们可以进一步提高去噪质量，同时仍然确保每秒数十帧的交互速率。总之，我们的方法做出了以下贡献：

- 一种新颖、实用且稳健的神经网络去噪方法。实验证明，在非常低的采样率下，去噪质量比现有方法更好。
- 一种使用可微神经双边网格的去噪方法，该方法允许端到端训练和高效推理。
- 双边网格的多尺度金字塔，用于处理多个级别的数据，并通过最佳加权组合进行融合。

## 二 相关工作

双边网格已经应用于 colorization 和 semantic segmentation，见论文 The fast bilateral solver。

此外，Gharbi 等人训练神经网络来学习仿射颜色变换，以进行图像增强。它们将低分辨率输入作为系数嵌入到双边空间中，并使用 slicing 系数来执行逐点变换。

受双边网格的边缘感知优势的启发，我们将神经双边网格纳入我们的去噪管道。与 Gharbi 等人不同的是，我们通过将多个相邻像素累积到特定的网格元素中，将有噪声的辐射度数据嵌入到双边网格中，并直接从网格中 slicing 以获得去噪数据。我们的方法避免了大核的每像素权重的显式计算，并通过快速、空间均匀的滤波器实现了高质量的去噪。

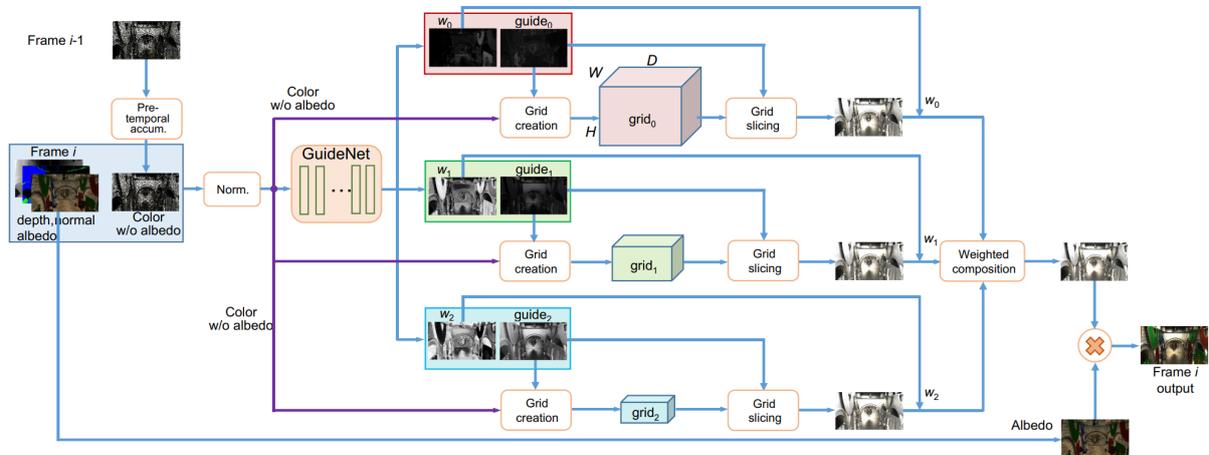
## 三 问题陈述

生成  $N$  对数据帧  $i \in [1, N]$ ，每个数据帧表示为  $((r_i, \mathbf{f}_i), t_i)$ ，三个字母分别表示有噪声 radiance、辅助特征和 ground truth。

神经双边网络  $M_i$ ，输入是 3D 空间的位置，输出是一个辐射度值。首先，我们将有噪 radiance 给 embedding 到网格中，设像素  $j$ ， $r_i(j)$  为其有噪输入。

论文公式 (1) 意味着可能有多个像素映射到 3D 网格同一个位置，像素映射操作  $g_i$  是一个神经网络。公式 (2) 是一个 3D 滤波器，作用就是去噪。之后，就通过公式 (3) 的 slicing 从网格中获得去噪结果。

## 四 去噪器架构



从图中可知，去噪管线分为了三种分辨率。GuideNet 来预测权重值和引导参数，即在三维网格中的位置。该位置最后还用于 slicing。权重值是用来组合不同分辨率结果的。这里的分辨率指的是网格分辨率。

可以想象的是（这也是去噪器有效的关键保证），这个去噪器如果网格分辨率比较低，那么滤波时就可以覆盖更大的区域，同时限制网格的长宽和深度是有必要的。

由于该去噪器没有时序去噪，所以时序效果应该不是很好。

## 参考文献

- [1] Deep adaptive sampling for low sample count rendering
- [2] Deep bilateral learning for real-time image enhancement