MuRF

> MuRF: Multi-Baseline Radiance Fields(CVPR 2024)

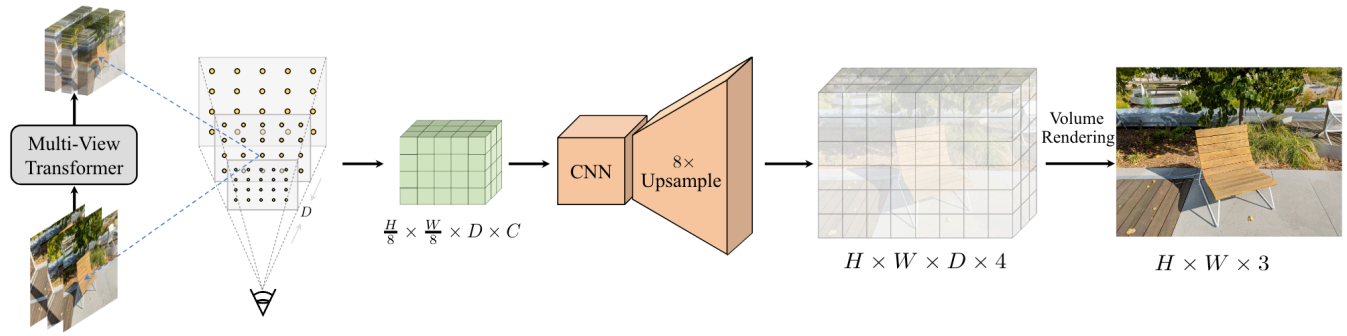

Figure 2. **Overview**. Given multiple input images, we first extract multi-view image features with a multi-view Transformer. To render a target image of resolution $H \times W$, we construct a target view frustum volume by performing $8\times$ subsampling in the spatial dimension while casting rays and sampling D equidistant points on each ray. For each 3D point, we sample feature and color information from the extracted feature maps and input images, which consists of the elements of the target volume $\boldsymbol{z} \in \mathbb{R}^{\frac{H}{8} \times \frac{W}{8} \times D \times C}$. Here, C denotes the channel dimension after aggregating sampled features and colors. To reconstruct the radiance field from the volume, we model the context information in the decoder with a (2+1)D CNN operating on low resolution and subsequently obtain the full-resolution radiance field with a lightweight $8\times$ upsampler. The target image is finally rendered with volumetric rendering.

3.1 Multi-View Feature Encoder

在MuRF（Multi-Baseline Radiance Fields）模型的3.1部分，即“Multi-View Feature Encoder”（多视图特征编码器），其目的是从输入的多视图图像中提取必要的特征，为MuRF的NeRF模型提供基础信息。

具体过程包括以下步骤：

1. **特征提取**：模型从(K)个输入图像中提取特征。这一过程由具有权重共享的二维卷积神经网络（CNN）和多视图Transformer共同完成。

- **CNN特征提取**：CNN包括6个残差块，每两个残差块间包含一个步幅为2的卷积层，实现特征的2倍降采样。因此，它生成1/2、1/4和1/8分辨率的特征图。
- **多视图Transformer**：将1/8分辨率的特征输入多视图Transformer以进一步处理。该Transformer基于GMFlow的双视图Transformer结构进行扩展，通过对所有(K)视图同时执行交叉注意力，使得处理多个输入视图更高效。

2. **特征聚合**：模型在体积重建步骤中对多尺度特征进行采样，并将采样后的特征进行拼接。这样得到的多视图特征可以为后续的体积渲染提供更多的细节和深度信息。

通过这一编码器结构，MuRF可以有效地从多视角中提取和融合信息，增强在小基线和大基线场景下的渲染效果。

3.2 Target View Frustum Volume

在3.2部分“Target View Frustum Volume”（目标视图椎体体积）中，MuRF方法通过构建与目标视图对齐的体积来进行图像渲染。

目标视图椎体体积的关键思想

与传统方法不同，MuRF方法不是在预定义的参考视图下构建体积，而是使用**目标视图的椎体体积**。这种空间对齐的椎体体积表示，能够从输入图像中有效地聚合信息，使渲染效果更加真实，尤其适用于输入图像之间存在较大视角差异的场景。

具体步骤

1. **椎体体积的构建**：
   为了渲染目标视图，MuRF会在目标视图的椎体内均匀采样。假设目标图像的分辨率为 $H \times W$，MuRF对空间维度进行 8 倍下采样，并沿每条光线均匀采样 (D) 个点。这种 8 倍下采样的方式能够在保持高分辨率的情况下，控制体积的计算复杂度。

2. **颜色采样**：
   由于进行了 8 倍下采样，仅采样一个点可能会导致信息损失。因此，MuRF在每个点的二维投影周围采样一个 9×9 的窗口，这样可以获得多个视角的颜色信息，提供更丰富的细节。

3. **特征采样和匹配**：
   从输入特征图中采样几何特征，以计算多视图一致性。模型通过计算多视图特征的**余弦相似性**来获取几何线索，从而预测体积密度。在观察到表面点的多视图特征往往具有高一致性的基础上，使用特征相似性来辅助几何信息的预测。

4. **多视图聚合**：
   采样到的颜色和特征会通过一个**学习的加权平均**方式聚合起来，以生成最终的椎体体积信息。聚合后的特征包含颜色、密度以及几何一致性的信息，这些信息随后会输入解码器以预测辐射场。

效果

这种构建目标视图椎体体积的方法可以有效地聚合来自多个视图的信息。相比传统的参考视图体积构建方法，MuRF的目标视图体积能够更好地捕捉到多视图输入中的场景几何结构，从而在大基线情况下提升渲染效果。

3.3 Context-aware Radiance Field Decoder

在3.3部分“Context-aware Radiance Field Decoder”（上下文感知的辐射场解码器）中，MuRF通过卷积网络来解码目标视图的椎体体积，以生成辐射场，从而实现高质量的渲染效果。

上下文感知的关键思想

为了在小基线和大基线的输入视角下获得更好的渲染效果，MuRF的解码器需要捕捉不同3D点之间的**上下文信息**。通过这种上下文信息的建模，解码器可以从数据中学习到有用的归纳偏置，从而增强场景结构的准确性。

具体步骤

1. **卷积网络结构**：
   为了有效建模3D空间中的上下文信息，MuRF采用了卷积网络来解码椎体体积。直接使用3D卷积虽然可以实现，但计算成本较高，因此MuRF使用了一种分解的卷积方法，以实现更高效的内存和参数使用。
   - 3D卷积分解为**2D空间卷积**和**1D深度卷积**，即（2+1）D卷积。这种分解的卷积方式既保持了空间和深度信息的上下文建模能力，同时减少了计算开销。
2. **解码器结构**：
   解码器主要由两部分组成：
   - （2+1）D CNN：在低分辨率的椎体体积上进行操作，以解码初步的辐射场信息。
   - **轻量级上采样器**：对解码后的结果进行8倍的上采样，以生成最终的全分辨率辐射场。

3. **颜色和密度预测**：
   解码器最终输出的辐射场包括每个3D点的**颜色和密度**。颜色和密度预测通过两个线性层实现，分别对应颜色和密度的3个和1个输出通道。

效果

与传统的MLP和Ray Transformer方法相比，MuRF的卷积解码器能够更有效地建模3D空间中的上下文关系，从而生成更清晰的场景结构。

3.4 Hierarchical Volume Sampling

在3.4部分“Hierarchical Volume Sampling”（分层体积采样）中，MuRF引入了分层采样策略来提高渲染质量。这种策略通过在粗略模型的基础上增加细化采样步骤，能够更准确地生成细节，从而改善最终的渲染效果。

分层体积采样的主要思想

分层体积采样通过两个阶段来提升采样效率和渲染质量：首先生成一个粗略的体积模型，然后基于粗略模型的结果在重要区域进行进一步的细化采样。这种方式使得计算更高效，因为在细化阶段仅需关注重要的体积区域。

具体步骤

1. **粗略模型（Coarse Model）**：
   - 在粗略阶段，每条光线上均匀采样 (D) 个点（例如 64 个点）。粗略模型会对这些采样点进行密度预测。
   - 基于密度预测结果，MuRF能够大致确定每条光线上更有可能包含重要信息的区域。
2. **概率分布函数（PDF）生成**：
   - 使用粗略模型的密度预测结果，计算每条光线上采样点的概率分布函数（PDF），并将其归一化。这些PDF用于指导细化采样阶段的点选择。
3. **细化模型（Fine Model）**：
   - 在细化阶段，根据粗略模型生成的PDF，在每条光线上进行新的点采样（例如 16 个点），从而更精准地采样关键区域。
   - 细化模型会预测这些采样点的颜色和密度，由于粗略阶段已去除空白空间或被遮挡的区域，细化采样更加高效。
4. **高效体积构建**：
   - 在细化阶段，MuRF在全分辨率下直接构建目标视图椎体体积，并减少体积的通道维度。这一优化让细化模型的计算开销更小。

效果

分层体积采样策略不仅提升了渲染的精度，还减少了计算资源的消耗。粗略模型帮助模型有效定位场景的几何结构，而细化模型则专注于有重要细节的区域，从而在减少冗余计算的同时提升渲染质量。

3.5 Training Loss

在3.5部分“Training Loss”（训练损失）中，MuRF定义了一个综合性的损失函数，以确保渲染结果与真实图像保持一致。这些损失函数包括多种常见的图像质量评价标准，能够有效提升模型的渲染质量和细节还原能力。

损失函数的组成

MuRF的训练损失由以下几部分组成：

1. **L1损失**：
    - 计算渲染图像与真实图像之间的L1损失（绝对值误差），从而确保预测颜色接近真实颜色。这种损失对于控制整体亮度和色彩匹配非常有效。
2. **结构相似性（SSIM）损失**：

- 使用SSIM损失来保持图像的结构信息。SSIM损失能够量化图像之间的结构相似性，因此在优化过程中，模型可以更加关注图像的局部细节和结构。

3. **感知损失（LPIPS）**：
   - 使用LPIPS（Learned Perceptual Image Patch Similarity）损失来提升视觉感知质量。LPIPS是一种基于深度神经网络的感知损失函数，可以使生成的图像更加符合人类视觉的真实感。

综合损失

以上三种损失函数的加权和构成了MuRF的总体训练损失函数。通过这种综合性损失，MuRF能够在多个指标上同时优化，生成细节丰富且视觉上接近真实的渲染图像。

效果

MuRF的训练损失通过多种评价标准来提升渲染结果，尤其是在结构和感知质量上，使得生成的图像在清晰度和真实感上达到了较高水准。