**技术交底书**

**发明名称：一种基于少视角协作的可泛化神经辐射场方法**

**本专利发明人：倪张凯 杨佩琪 王瀚漓 程久军 ­**

**技术交底书撰写人及技术联系人：\_倪张凯**

**电话： 13646019080 FAX: \_ \_ E-MAIL：\_zkni@tongji.edu.cn \_**

注意：

1.**第2部分和第4部分**，**一定要写的全面、清楚**。

2.英文缩写要有中文译文和英文原词，最好在术语解释部分给出。

3.**全文对同一事物的叫法应统一，避免出现一种东西多种叫法。**

4.应该阐述发明目的是通过什么技术方案来实现的，不能只有原理，也不能只做功能介绍。

**术语解释：**

解释一些跟技术方案有关的专业术语

**1、本发明要解决的技术问题是什么？**（务必明确一个最主要的技术问题）

少视角输入下的高质量可泛化新视角图像合成方法

**2、详细介绍技术背景,并描述已有的与本发明最相近似的实现方案**（与本发明最接近的技术方案的说明（对于方法，应说明现有方法的步骤，对于装置，应当说明结构组成及其关系））

随着计算机视觉和图形学的飞速发展，三维重建和渲染技术已经成为各个领域中不可或缺的重要工具。三维重建是指利用一系列图像或传感器数据，恢复出场景或对象的三维几何形状和外观特征。这种技术不仅可以提供具有空间感的数字模型，为虚拟现实、增强现实、计算机辅助设计等应用提供基础支持，而且还可以为医学影像处理等领域提供有力支持。渲染则是将三维场景转化为逼真的二维图像的过程，通过模拟光线的传播和交互，使得渲染结果与真实场景无法区分。其中，由二维图像生成三维内容是进行三维重建的重要方法，也是和渲染过程互为逆的逆渲染过程。

随着计算机性能的提高和深度学习技术的兴起，基于神经网络的三维重建方法逐渐成为研究的焦点。神经辐射场（Neural Radiance Fields，NeRF）是其中引起广泛关注的一个重要代表。传统的三维重建方法通常依赖于多视角摄影、激光扫描等传感器数据，这些方法通常需要昂贵的设备、大量的时间和大量的计算资源。相比之下，神经辐射场利用神经网络模型来学习场景的密度和颜色信息，而无需先前的摄影或扫描数据。这种方法的关键思想是，通过训练一个神经网络来预测在不同视角和深度位置的像素颜色和密度，可以实现高质量的三维场景重建。

神经辐射场的具体做法是，通过一个隐式的五维神经辐射场构建过程生成新的视图图像，其具体表示为：，其中 表示三维的位置； 表示二维的目标视角方向；输出 表示 值；表示体密度，可以理解为射线在给定粒子处终止的概率。神经辐射场通过体渲染来生成最终的目标二维图像：

， （1）

， （2）

其中，和 表示沿着目标射线距离处的采样点的颜色和体密度； 表示从 到 沿射线的累计透射率，即射线从 到 不碰到任何其他粒子的概率，该项的设计表明采样点离需要渲染的像素点越远，那该采样点对该像素颜色信息的贡献就越小。

通过最小化渲染得到的预测像素颜色和真实像素颜色之间的均方误差来优化神经辐射场：

， （3）

其中， 为目标视角P的所有相机射线的集合。

**3、以因果关系推理的方式推导出现有技术的缺点是什么？**

**针对这些缺点，说明本发明的目的。**（缺点可以有多个，但一定要分清主次。最主要的缺点原则上只能有一个。）

虽然神经辐射场和与其相关的方法在三维重建领域取得了显著的突破，但神经辐射场在质量和速度两方面依然有着局限性。首先一般的神经辐射场需要大量的训练数据来捕捉三维场景的多样性和复杂性，这意味着需要大量的二维图像和相应的姿态信息，而大量的姿态数据的获取过程比较繁琐，且并不一定准确；并且神经辐射场在处理视角变化和遮挡问题时仍然存在挑战，当某些区域无法被多个视角观察到时，模型可能无法准确地重建这些区域；神经辐射场模型在从训练数据中未见过的区域进行泛化的能力仍然是一个挑战，特别是在极端光照或环境条件下，这往往意味着我们需要大量的时间和计算成本去建模一个特定的场景。这些缺陷对于神经辐射场的落地使用都是挑战。总结一下，本发明主要针对神经辐射场的以下问题提出优化：

（1）大多数方法采用一个模型对应单一场景的范式，这意味着我们需要额外的时间和空间成本来训练一个新的场景，从而影响模型的应用和落地。本发明目的在于得到一个可应用于多个场景的统一模型，该模型具有较强的可泛化性，即使在面对未知场景时也能够得到高质量的重建结果。

（2）源视图数量不足情况下的信息稀疏问题：少视角输入图像通常仅提供有限的视角和位姿信息，输入的少视角图像无法涵盖完整场景，可能会出现部分信息缺失或被遮挡的情况，从而对重建和渲染产生负面影响，导致重建结果不完整或不精确。

（3）密集监督下的几何与外观失真问题：虽然原始的神经辐射场输入的是每个采样点的坐标和视角方向输出的也是每个采样点的辐射值，但是对于每个采样点的辐射值并没有可以直接用做监督的真值信号，我们用于监督的都是对应一条目标射线上所有采样点的渲染结果的一个真实目标像素颜色，这可能会导致由于局部区域采样点预测不准而造成的局部区域市镇与外观失真问题。

**4、本发明技术方案的详细阐述，应该结合流程图、原理框图、电路图、时序图进行说明**（所有英文缩写都应有中文注释；**所有附图都应该有详细的文字描述，以别人不看附图即可明白技术方案为准**；同时附图中的关键词或方框图中的注释都尽量用中文；方法专利都应该提供流程图，并提供相关的系统装置）。

技术问题

本发明涉及三维重建领域，具体涉及少视角输入下的可泛化新视角图像合成方法。

技术方案

本发明提出了一种基于少视角协作的可泛化神经辐射场方法，该方法不需要任何额外的监督信息，通过考虑输入源视图之间的协作机制实现了精确和广义的三维场景建模能力。具体来说，该方法包括使用预训练的编码器从源图像中提取特征体，然后应用跨视角的特征协作融合以自适应的融合不同源视图的特征。该协作融合方法通过注意力机制隐式引入三维场景的几何先验，并通过结合不同视点中相应部分的信息来校正被遮挡区域的特征。此外，协作机制还被应用于神经辐射场的输出端，从几何和外观的角度进行约束。其中，几何正则化采用自监督方式，旨在于最大限度地减少相邻目标射线预测深度之间的差异。外观正则化利用与每个目标射线在源视图中对应的极线作为颜色伪标签，从而引导目标射线的生成。本发明所训练的模型具有适用于任意场景的泛化能力。

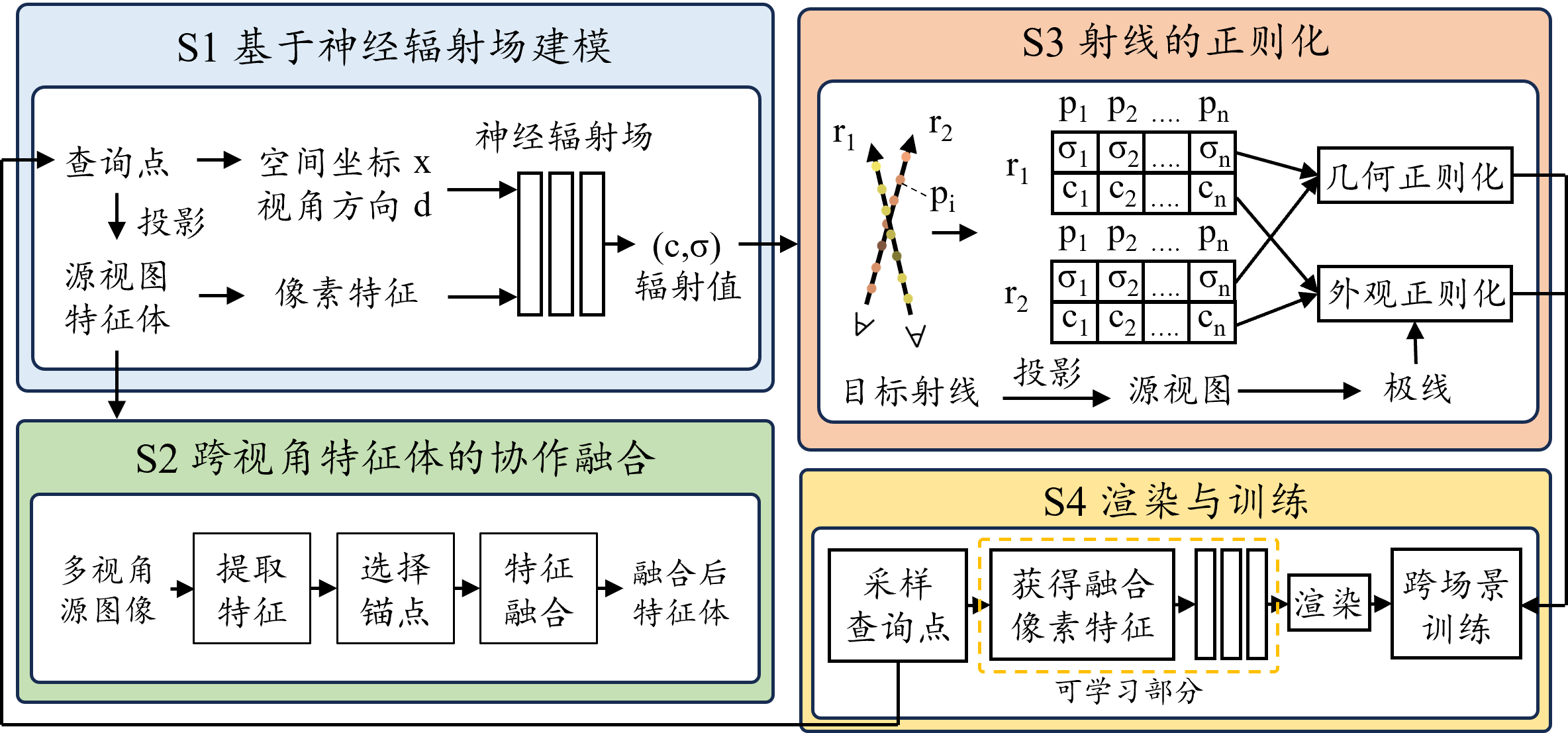


图 1 本发明的整体框架流程图。

进一步地，如图 1 所示，所述新视角图像生成过程包含以下步骤。

步骤 1 构建神经辐射场

本发明在神经辐射场的5D输入基础上引入另外一源图像像素特征，因此改进后的神经辐射场可以表示为：

， （4）

其中，输入包括查询点坐标，视角方向 以及对应源图像的像素特征 ，输出的是查询点对应位置的辐射值 。表示计算平均的操作，表示源视图的数量， 对应神经辐射场中的位置编码操作。

步骤 2 跨视角特征体的协作融合

首先在神经辐射场的输入端进行多视角源图像的协作融合，其目的在于获得个融合之后的图像特征体，这些特征体整合了其他源视图的信息，从而具有鲁棒的场景建模能力。

步骤 3 射线的正则化

接下来，将协作机制进一步拓展到神经辐射场输出的每个采样点的预测辐射值上。从几何和外观两个角度出发，使用显式的方式直接约束神经辐射场的输出。

步骤 4 体渲染与训练

在每次优化迭代中，随机采样一批相机射线进行样本点查询。然后，使用体渲染过程来渲染每条目标射线的颜色。本发明的训练损失包括预测像素与真实像素颜色之间的平均平方误差以及步骤3中所提出的射线正则化约束，联合应用于神经辐射场以及跨视角协作融合模块的优化。

技术效果

与现有技术相比，本发明具有以下有益效果：

（1）协作融合机制：本发明提出了一个新颖的跨视角协作模块，在输入和输出端将多视图补偿和一致性集成到神经辐射场中，以对齐来自不同视图的信息。

（2）射线的正则化：通过多方向目标射线的自监督约束实现相邻区域的几何正则化。此外，利用目标射线对应的原视图极线约束目标射线的颜色预测。

（3）优异的泛化性：本发明所提出的协作机制使模型具备隐式建模场景几何和外观的能力，从而可以广泛应用于未知新场景并实现高效的新视角重建。

附图说明

图 1 为本发明的整体框架流程图；

图 2 为本发明的细节框架流程图

图 3 为本发明步骤2的详细描述；

图 4 为本发明步骤3的详细描述；

图 5 为本发明与之前方法在DTU数据集上3个和6个输入视角的可视化比较结果；

图 6 为本发明与之前方法在LLFF数据集上3个输入视角的可视化比较结果；

具体实施方式

下面结合附图和具体实施例对本发明进行详细说明。本实施例以本发明技术方案为前提进行实施，给出了详细的实施方式和具体的操作过程，但本发明的保护范围不限于下述的实施例。

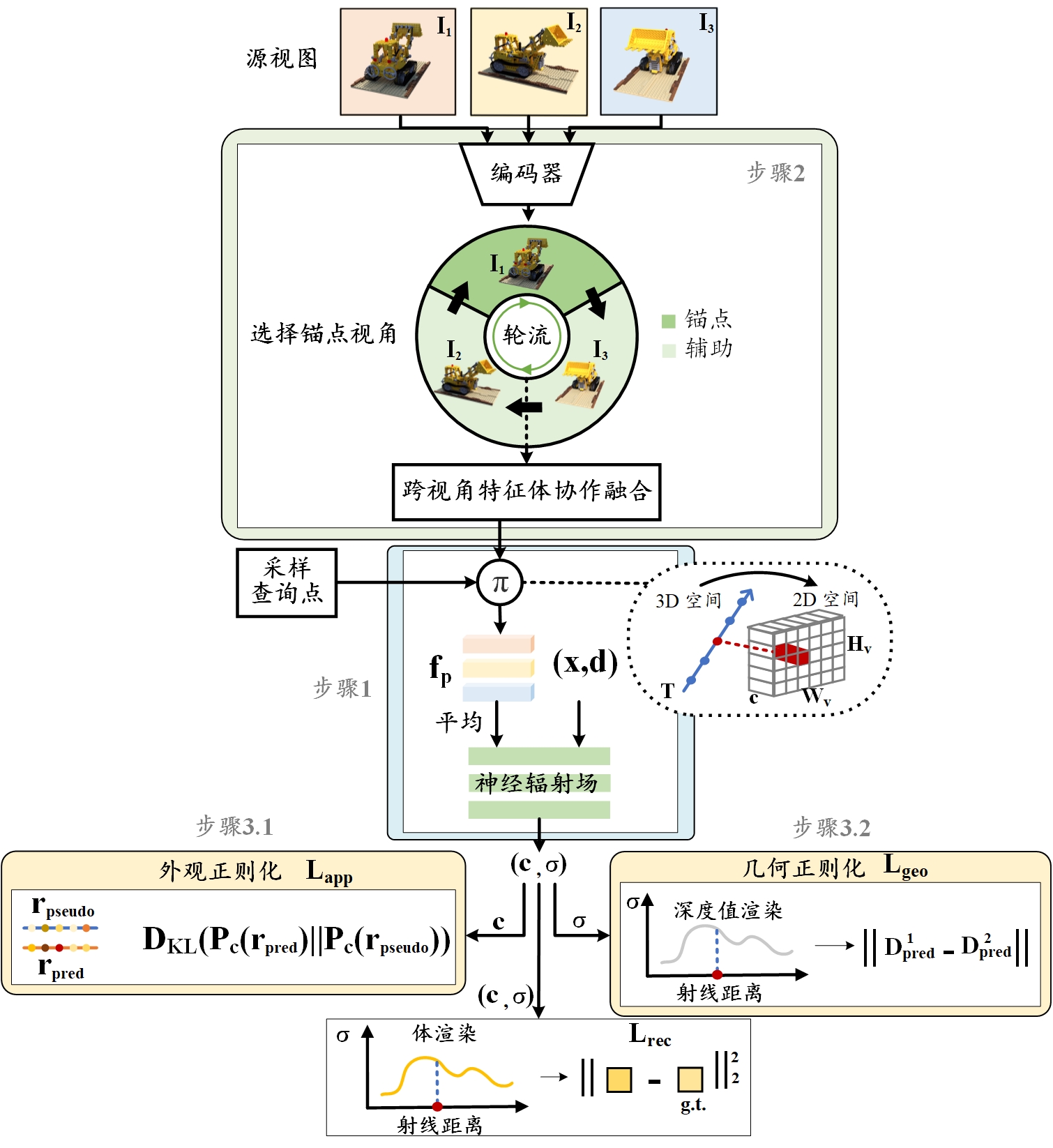


图 2 本发明的细节框架流程图。首先，步骤 1 给出基于神经辐射场的针对本发明目标的建模；为了获得更加鲁棒的场景建模能力以及更加精细的泛化性能，步骤 2 首先在神经辐射场的输入端进行多视角源图像的协作融合；接下来，步骤 3 将协作机制进一步拓展到神经辐射场输出的每个采样点的预测辐射值上。从几何和外观两个角度出发，使用显式的方式直接约束神经辐射场的输出。最后，使用体渲染过程来渲染每条目标射线的颜色。本发明的训练损失包括三个部分：(Ⅰ)重建损失 ：预测像素与真实像素颜色(Ground Truth，g.t.)之间的平均平方误差；以及步骤3中所提出的(Ⅱ)射线几何正则化损失 以及(Ⅲ)外观正则化损失 。

一种基于少视角协作的可泛化神经辐射场方法，如图 1 所示，包括步骤：

步骤 1 构建神经辐射场

该步骤为运用多层感知机隐式建模三维场景，本发明构建优化后的神经辐射场，在神经辐射场原始的五维输入基础上加入了另外一源图像像素特征 ，神经辐射场的表示更新为：

， （5）

其中输入包括 3D 位置 和视角方向 ，输出为颜色 和体积密度，表示平均操作，表示源视图的数量， 表示的是位置编码操作：

， （6）

像素特征 由训练采样点 在经过跨视角融合后的特征体 上投影插值得到：

， （7）

这里的 表示空间采样点 x 在第 个融合后的特征体 上插值得到的像素特征。

步骤 2 跨视角特征体的协作融合

该步骤为运用不同视角之间的协作融合机制，得到上述的携带跨视角信息的源图像特征，如图 2 所示，其具体包含如下 3 个子步骤：

步骤 2.1 提取源图像特征体

对于第个源图像 ，使用预训练的编码器 来提取视觉特征体：

， （8）

表示提取出的视觉特征体，，， 分别表示特征体的高度、宽度和特征通道维度，最终能够得到对应每个源视图的同样数量的 个视觉特征体， 表示编码器提取特征的操作。

步骤 2.2 锚点特征体的选择

我们依次轮流将每一个视觉特征体作为锚点特征体，并融合其他所有视觉特征体的信息。表示当前的锚点特征体，其他所有特征体的总和作为当前第个锚点特征的辅助特征体 ，该做法能够保证该融合机制能够接受任意数量视觉特征体的输入：

， （9）

此处的 表示第 个锚点特征体， 表示的是第 个视觉特征体。

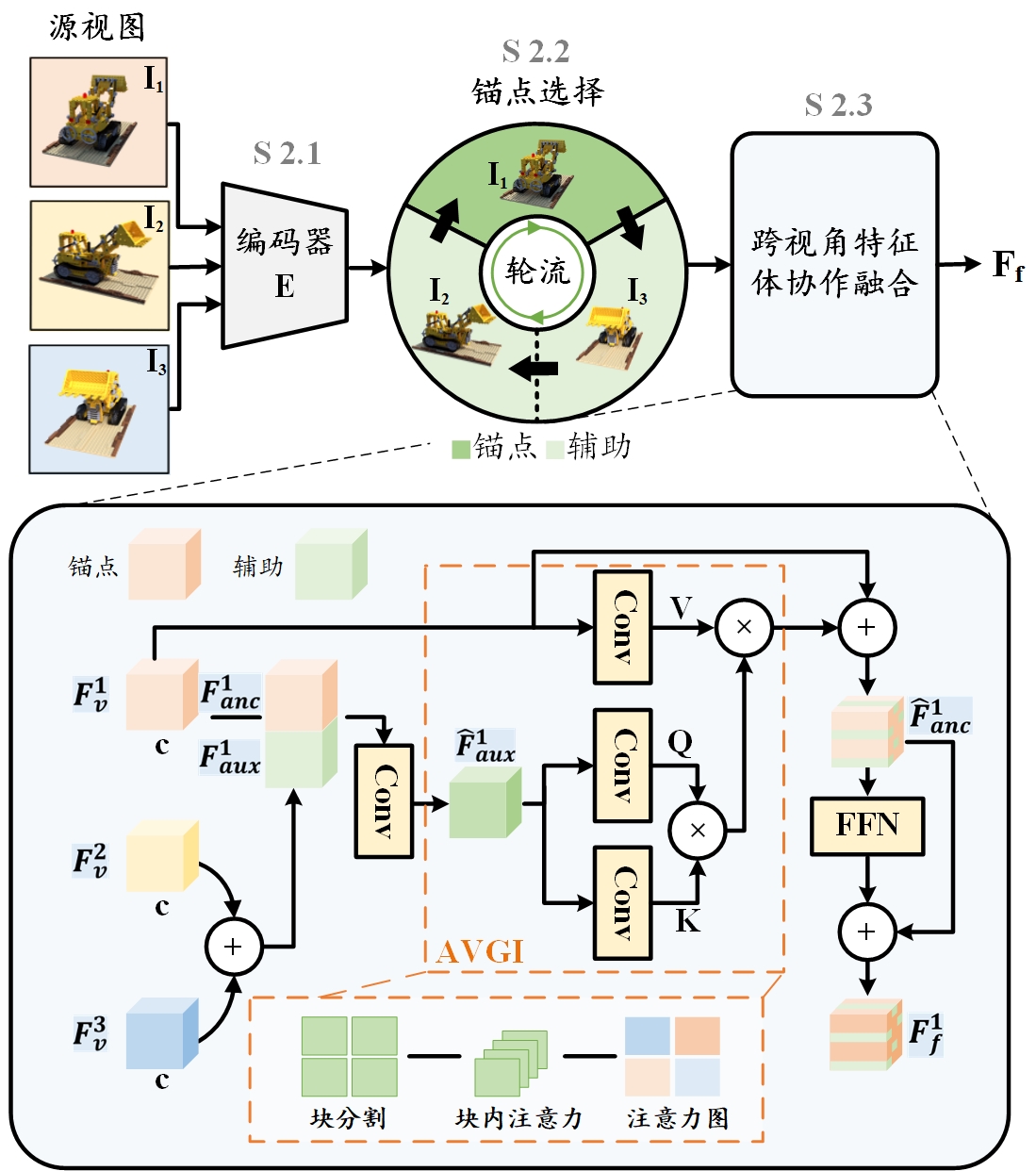


图 3 跨视角特征体的协作融合。首先步骤 2.1 使用编码器提取对应源视图的视觉特征体(公式(8))，其次步骤 2.2 选择当前的锚点特征体，图中演示的情况是源视图 1 对应的视觉特征体 作为锚点，源视图 2 和 3 对应的视觉特征体 和 作为辅助的情况，整个融合过程中每个视觉特征体会依次被作为锚点；蓝色圆角矩阵中展示的是步骤 2.3 对应的具体的融合过程，依旧以 作为锚点， 和 作为辅助的情况为例子，首先将 和 两个视觉特征体相加获得浅绿色的辅助特征体 (公式(9))，随后将 和 进行级联，并且一起输入一个卷积层(Conv，Convolutional Layer)以降低维度(公式(12))，所得到的 和 一起进行辅助特征体引导的融合(，Auxiliary Volume Guided Integration)， 经过两个不同的卷积变换之后作为 Query 和 Key， 同样经过一卷积变换之后作为 Value，该操作基于注意力机制进行跨视角的融合(公式(13))，所得结果再加上 作为残差得到 (公式(10))，最终 经过一个由两层卷积层组成的前馈网络(FFN，Feed Forward Network)，并加上 作为残差得到最终对应源视图 1 的融合特征体 (公式(11))。

步骤 2.3 跨视角协作融合

我们的目标是通过跨视角的协作融合操作 得到 个融合后的特征体，CCVI的计算过程如下：

， （10）

， （11）

， （12）

其中， 表示前馈网络， 表示用于降维的卷积操作， (Auxiliary Volume Guided Integration) 表示辅助特征体引导的融合：

， （13）

其中，

，

， （14）

，

分别代表可学习变换， 为 Q、K 的特征通道维数。

值得注意的是，提出的 AVGI 在图像块的层面进行注意力计算。本发明使用局部注意力机制来强调每个图像块中最重要的区域。具体来说，我们将每个特征体分割为 P×P 图像块，并且在每个图像块内进行基于局部注意力机制的融合。然而，如果只是直接的简单地将每个特征图划分为图像块，并在小块中进行融合，会阻碍相邻块之间的信息传递。为了解决这个问题，本发明将每个图像块扩展一个额外的像素带，从而得到大小为 (s+2a)×(s+2a) 的图像块，其中 s 表示原本的图像块宽度，a 是额外的像素带带宽。如果存在相邻块，则从相邻块复制额外的带像素；否则，则直接使用零来填充额外的像素带。这种扩展允许相邻块之间进行信息交换。

步骤 3 射线的正则化

接下来，将协作机制进一步拓展到神经辐射场输出的每个采样点的预测辐射值上。如图 3 所示，本发明从几何和外观两个角度出发，使用显式的方式直接约束神经辐射场的输出。

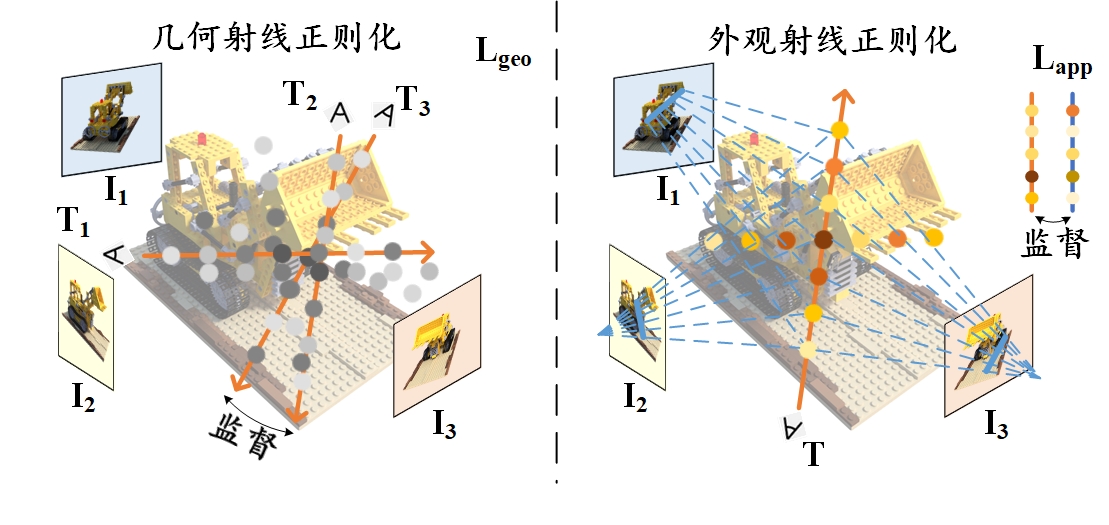


图 4 射线的正则化。图中 T 代表的是目标射线方向（Target），步骤 3.1 几何射线正则化中，使用射线方向相邻近的两条目标射线进行互相监督，具体做法是拉近两条射线的预测深度值；步骤 3.2 外观射线正则化中，橙色代表目标射线，蓝色的实线代表目标射线在源视图上投影得到的极线，具体做法是，使用极线颜色的平均值作为伪标签，拉近伪标签与目标射线的颜色分布。

步骤 3.1 几何正则化

该步骤的目标是约束神经辐射场对每个采样点预测的体密度并增强模型对于视角抖动的鲁棒性。由于每个点的体密度真值无法获得，故利用相邻射线进行协作式相互监督，具体表现为将最接近的组射线配对，并最小化每组两条射线预测深度值的损失:

， （15）

， （16）

其中， 和 表示一组相邻的目标射线， 表示预测的深度值， 表示的是逐元素点乘，本发明使用掩码 来忽略只有一条光线与物体相交的射线对，以减少可能会造成的边缘模糊伪影。 表示射线的累积密度，其中 表示射线上的第个采样点。本发明设置。

步骤 3.2 外观正则化

使用类似于提取采样点像素特征的投影插值操作，提取每条目标射线对应极线的颜色信息作为每条目标射线的伪标签。采用最小化两种光线颜色分布 和 之间的散度 来进行约束:

， （17）

其中，表示训练过程中发射的目标射线，表示对多个源视图中目标射线对应极线的颜色进行平均后得到的“伪标签”。 表示第条射线上的第 个采样点。使用另一个掩码 来排除被不正确投影的采样点：

， （18）

其中， 表示目标射线上的点进行单应变换得到的源图像中对应的像素坐标，是指源图像中的最大像素值。

步骤 4 渲染与训练

步骤 4.1 采样查询点

关于采样射线的设置。当采样的目标射线过于集中在某个局部区域时会导致过拟合现象，并且也不利于模型的泛化性能，本发明在发射目标射线时使用随机的原点和方向，使用的目标射线数量为 。首先随机的采样 112 条射线并将其中最后 16 条射线作为参考射线，然后在其附近采样 16 条邻近的射线作为配对射线用于几何正则化。

关于采样点的设置。遵循和原始的NeRF类似的做法，同时训练两个网络，一个粗糙网络一个精细网络，其中粗糙网络在训练时使用分层抽样对一组 位置进行抽样：

， （20）

给定粗糙网络的输出后，再在精细阶段沿着每条射线采样携带更多信息的 个采样点，这些采样点偏向于物体的相关部分，例如物体表面，而这些相关部分则是由粗网络的结果指示的。

步骤 4.2 体渲染和深度渲染

穿过场景的任何光线的颜色可以用经典的体渲染原理来表示，体密度可以解释为射线在x处终止于无穷小粒子的微分概率。想要渲染一个特定视角图像的某个特定像素时，可以假定是从这个像素沿着该视角方向向场景发射一条射线，而对这条射线上所有点的颜色和体密度信息的乘积积分就可得到该像素点最终的 信息，其中相机射线可表示为 ，其中的 表示与射线原点之间的距离，在已知远近边界 和 的情况下，一条射线的期望颜色可以由公式（1）表示，类似的，一条射线的期望深度可以表示为:

， （19）

步骤 4.3 损失函数

在每次优化迭代中，遵循上述方法随机采样一批相机射线并查询样本点。然后，使用体渲染过程预测每条目标射线的颜色，所采用的重建损失如公式（3）所示，再结合之前部分提到的射线正则化，在训练过程中使用的损失函数由三个部分组成：

其中， 和 是控制各项损失函数权重的超参数，在实验中设置 。

**5、本发明的关键点和欲保护点是什么？**（对于上一部分给出的详细的完整技术方案，在本部分是提炼出技术方案的关键创新点，列出1、2、3...，以提醒代理人注意，便于专利代理人撰写权利要求书）

1. 有助于模型泛化性的采样和训练方式

为了突破一个模型对应一个场景的范式限制，训练得到具有泛化能力的模型，本发明应用优化后的辐射场模型，其输入不再仅限于原始的五维输入数据，而是加入了与采样点相关的图像像素特征，并且在跨场景数据集上对神经辐射场进行预训练。为了更好的与本发明提出的射线正则化部分配合，并且避免因局部集中采样训练射线导致的对某个场景或局部区域的过拟合现象，保证模型的能力，本发明提出一个可控制的稀疏射线采样方式：首先随机的采样112条射线并将其中最后16条射线作为参考射线，然后在其附近采样16条邻近的射线作为配对射线用于几何正则化。该做法既保证模型的泛化性，又保证了本发明的几何正则化部分的可控性。

2. 跨视角源图像特征的协作融合机制

为解决源视图数量不足情况下的信息稀疏问题，本发明采用一种跨视角的特征协同融合方法，以自适应地整合来自不同源视图的特征。该协同融合方法通过引入注意力机制，隐含地考虑了三维场景的几何先验，同时通过整合来自不同视点的相关信息来校正被遮挡区域的特征。具体而言，与以往方法侧重于像素级特征不同，本发明关注更全局的图像块，旨在生成多个融合后的特征体，这些特征体包含了来自其他源视图的信息。本发明选择了注意力机制作为可训练的聚合函数，其具有选择性地强调源视图中的重要特征的能力，通过整合不同视点中相应部分的信息并突出重要信息以校正被遮挡区域的特征。同时，为了确保适应任意数量的输入源视图，在每次融合时选择一个参考源视图特征体作为锚点，将其他所有源视图的特征体合并，作为锚点源视图的辅助特征体，并逐步的融合锚点和辅助特征体。这个过程可以看作是不同视角源图像之间的相互协同作用，这种跨视角的融合过程有助于隐式地建模源图像的每个局部区域在其他视图中的对应区域，以实现不同视图信息的对齐，并利用来自多个视角的融合信息来纠正源图像中可能存在的潜在偏差。此外，这个过程还有助于增强同一场景的多视角一致性。

3. 基于相邻目标射线的自监督几何外观正则化

为了减缓缺失密集监督下的几何与外观失真问题，尽可能保证相邻区域在不同视角中的几何外观一致性，本发明研究在不加入任何额外监督的情况下对模型进行几何与外观两个方面的正则化约束，该约束直接针对重建场景的每个采样点，实现更加局部更加细节的优化。具体而言，为了实现几何正则化，本发明使用相邻采样点的体密度预测结果相互监督的方法，以约束神经辐射场在局部区域的几何稳定性，最大程度地减少相邻目标射线预测深度之间的差异。该操作相当于平滑了相邻区域的几何形状，可以减轻空间中某个点体密度值剧烈变化的情况，并进一步防止重建结果中出现雾状漂浮物缺陷的可能性。在外观正则化方面，本发明研究了使用目标射线对应的源视图极线来约束目标射线的颜色生成。在源视图中，与每个目标射线最相关的区域应该对应于其投影所得的源视图极线。因此，使用极线的颜色作为目标射线颜色的伪标签。需要注意的是，在计算伪标签时，需排除不正确投影的点。由于仍然存在部分遮挡区域，直接将极线上每个对应点的颜色调整到伪标签可能会引入一些负面偏差，因此本发明选择了拉近伪标签和预测值之间的KL散度。

**6、用推理方式推导出本发明的优点**（务必与第3部分的现有技术的缺点相对应），可以对应3部分所要解决的技术问题或发明目的来描述。

**（1）提高模型泛化性的建模方式。**为了突破一个模型对应一个场景的范式限制，训练得到具有泛化能力的模型，本发明构建了一个优化后的辐射场模型，在神经辐射场的 5D 输入基础上引入另外一源图像像素特征，并且选择在跨场景数据集上对模型进行预训练，该做法使得预训练的神经辐射场在面对不同像素特征的输入时能够产生不同的预测结果。即使在面对未知的场景时，只要提供与该场景相关的源图像像素特征，神经辐射场也能够准确预测新场景的几何和外观，从而实现模型的泛化能力。

**（2）研究不同视角之间的协作融合机制**。本发明通过跨视角的协作融合机制将多视图补偿和一致性集成到神经辐射场中，在无需加入任何额外监督信息的前提下，对齐来自不同视图的信息。由于源视图数量有限，难免会有部分的场景信息是未被覆盖的，或者是存在负面偏差的，例如某局部区域在某个源视图中可见，然而在目标视图中处于被遮挡区域这种情况。本发明通过跨视角的协作融合机制将多视图补偿和一致性集成到神经辐射场中，在无需加入任何额外监督信息的前提下，对齐来自不同视图的信息。通过融合跨视图的源图像特征体，我们可以增强有限的稀疏信息。根据像素空间到相机空间的坐标变换，某个图像块的确切空间位置可以通过使用相机参数将它们重投影到 3D 空间中来确定。而通过跨视图的特征融合，我们可以利用注意力机制将相关的图像块互相匹配并且隐式实现相关区域的空间位置定位。注意力机制作为可训练的聚合函数，可以选择性地强调源视图中的重要特征，通过结合不同视点中相对应部分的信息，并且强调重要的信息来校正被遮挡区域的特征。该过程可以被视作不同视角源图像之间的互相协作，这种跨视角的融合过程有助于隐式建模源图像的每个局部区域在其他视图中的对应区域，从而对齐来自不同视图的信息，并且可以利用来自多个视角的融合信息来校正源图像中可能存在的潜在偏差，此外这个过程也有助于增强同一场景的多视角一致性。

**（3）设计目标射线的正则化。**通过多方向目标射线的自监督约束实现相邻区域的几何正则化，同时利用目标射线对应的源视图极线约束目标射线的颜色预测，以实现更加直接的几何和外观约束。本发明进一步探索如何将协作机制拓展到神经辐射场的输出端，虽然原始的神经辐射场输入的是每个采样点的坐标和视角方向输出的也是每个采样点的辐射值，但是对于每个采样点的辐射值并没有可以直接用做监督的真值信号，我们用于监督的都是对应一条目标射线上所有采样点的渲染结果的一个真实目标像素颜色。我们研究使用现有的信息对模型进行几何与外观两个方面的正则化约束。其中，几何正则化采用自监督方法，旨在于最大限度地减少相邻目标射线预测深度之间的差异。外观正则化利用与每个目标射线的源视图中对应的极线作为颜色伪标签，从而引导目标射线的生成。

总结起来，与现有技术相比，本发明具有以下有益效果：

（1）优异的泛化性：本发明所提出的协作机制使模型具备隐式建模场景几何和外观的能力，从而可以广泛应用于未知新场景并实现高效的新视角重建。

（2）协作融合机制：本发明提出了一个新颖的跨视角协作模块，在输入和输出端将多视图补偿和一致性集成到神经辐射场中，以对齐来自不同视图的信息。

（3）射线的正则化：通过多方向目标射线的自监督约束实现相邻区域的几何正则化。此外，利用目标射线对应的原视图极线约束目标射线的颜色预测。

**7、针对4中的技术方案，是否还有别的替代方案同样能完成发明目的？**

无

**8、其他有助于专利代理人理解本技术的资料**（给代理人提供更多的信息，可以有助于代理人更好更快的完成申请文件）

**9、案例（实施例）**

数据集

实验在 DTU、LLFF 数据集上进行测试。为了证明模型的泛化能力，本发明额外引导了一个在DTU上进行训练，并在 LLFF 上测试的实验。为了评估本发明在少视角输入下的新视角合成质量，分别进行了 3 视图、6 视图和 9 视图输入的实验。

方法应用

本发明在发射目标射线时使用随机的原点和方向，使用的目标射线数量为 。首先随机的采样 112 条射线并将其中最后 16 条射线作为参考射线，然后在其附近采样16条邻近的射线作为配对射线用于几何正则化。这些新采样的射线与参考射线共享相同的相机参数和原点，但在像素平面上显示最多 7 个像素的偏移。对于 3 视图和 6 视图的训练，我们将训练批次大小（Batch Size, BS）设置为 3，对于 9 视图的训练，BS 设置为 2。整个训练过程中保持固定 1e-4 的学习率。

实施效果对比

表1 DTU数据集上本发明与其他模型的对比实验结果

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 实验设定 | PSNR | | | SSIM | | |
| 3视角 | 6视角 | 9视角 | 3视角 | 6视角 | 9视角 |
| DietNeRF (ICCV 2021) | 在DTU上  逐场景训练 | 10.01 | 18.70 | 22.16 | 0.354 | 0.668 | 0.740 |
| DS-NeRF(CVPR 2022) | 16.50 | 20.50 | - | 0.540 | 0.730 | - |
| InfoNeRF(CVPR 2022) | 11.23 | - | - | 0.445 | - | - |
| RegNeRF(CVPR 2022) | 15.33 | 19.10 | 22.30 | 0.621 | 0.757 | 0.823 |
| FreeNeRF(CVPR 2023) | 18.02 | 22.39 | **24.20** | 0.680 | 0.779 | **0.833** |
| SRF(CVPR 2021) | 在DTU上  跨场景训练 | 15.84 | 17.77 | 18.56 | 0.532 | 0.616 | 0.652 |
| MVSNeRF(ICCV 2021) | 16.33 | 18.26 | 20.32 | 0.602 | 0.695 | 0.735 |
| PixelNeRF(CVPR 2021) | 18.74 | 21.02 | 22.23 | 0.618 | 0.684 | 0.714 |
| ColNeRF (本发明) | **19.55** | **22.94** | 23.93 | **0.716** | **0.797** | 0.824 |

为了便于比较，本发明选择了几种最先进的方法，这些方法都在针对少视角输入的新视角重建任务提出改进。这些方法包括 PixelNeRF，SRF，MVSNeRF，DietNeRF，DS-NeRF，InfoNeRF，RegNeRF 和 FreeNeRF。前三个方法类似于本发明，在跨场景数据集中进行预训练，而其余五个方法则针对特定场景进行优化。下面主要报告了本发明与其他几个少视角输入方法在 PSNR、SSIM 上的性能比较。

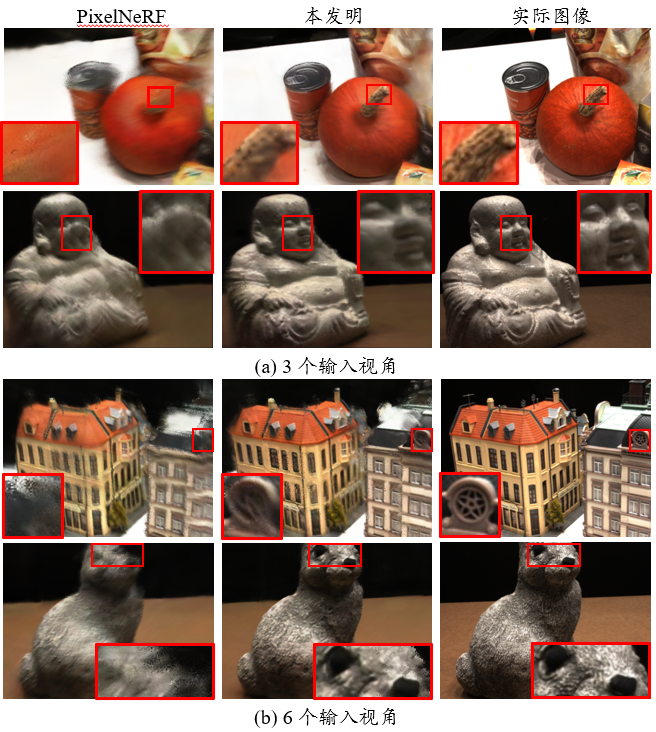
表 1 给出了 DTU 数据集上的定量结果。本发明在大多数实验设置中都表现出色。为了区分的明显，分别用加粗和加下划线的方式表明性能第一和第二的两种方法。

表 2 给出了本发明在 DTU 数据集上训练并且在 LLFF 数据集上测试的定量结果，将预训练模型推广到LLFF数据集，并在3视角、6视角、9视角的设定下分别对每个场景进行15K、10K和5K的微调迭代。虽然看上去本发明在性能上比不上像 FreeNeRF 这样的逐场景优化方法，但值得注意的是，它们在每个场景上训练单独的模型，并且每一次都要进行250K次的迭代。相比之下，本发明为所有场景训练单个模型，并以更少的微调成本获得可相媲美的结果。

表2 LLFF数据集上本发明与其他模型的对比实验结果

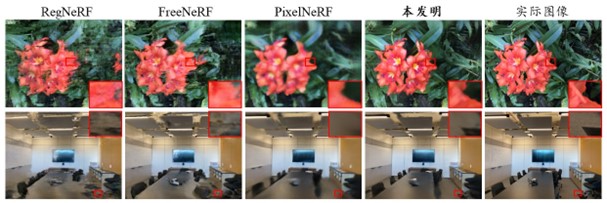
|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| 方法 | 实验设定 | PSNR | | | SSIM | | |
| 3视角 | 6视角 | 9视角 | 3视角 | 6视角 | 9视角 |
| DietNeRF (ICCV 2021) | 在LLFF上  逐场景训练 | 14.94 | 21.75 | 24.28 | 0.370 | 0.717 | 0.801 |
| RegNeRF(CVPR 2022) | 19.08 | 23.10 | 24.86 | 0.587 | 0.760 | 0.820 |
| FreeNeRF(CVPR 2023) | 19.63 | **23.73** | **25.13** | **0.612** | **0.779** | **0.827** |
| SRF ft(CVPR 2021) | 在DTU上  跨场景训练 | 17.07 | 16.75 | 17.39 | 0.436 | 0.438 | 0.465 |
| MVSNeRF ft(ICCV 2021) | 17.88 | 19.99 | 20.47 | 0.584 | 0.660 | 0.695 |
| PixelNeRF ft(CVPR 2021) | 16.17 | 17.03 | 18.92 | 0.438 | 0.473 | 0.535 |
| ColNeRF ft (本发明) | **20.97** | 23.32 | 23.52 | 0.587 | 0.747 | 0.762 |

图 4 给出了本发明和之前的方法 PixelNeRF 在 DTU 数据集上的可视化比较结果。由于 PixelNeRF 直接将源视图的像素特征的平均值输入到神经辐射场中，导致可能引入了负面偏差。特别是当投影点处于被遮挡区域时，PixelNeRF 表现出了模糊和形状扭曲等失真。相比之下，本发明渲染的结果校正了这些错误，并隐式地重建了物体的几何形状，从而实现了高质量的新视角重建。



**图5** 在 DTU 数据集上的可视化比较结果

图 5 给出了本发明和之前的方法在 LLFF 数据集上的可视化比较结果。FreeNeRF 和 RegNeRF 的渲染结果呈现出明显的噪声问题，而 PixelNeRF 的渲染结果则显示出严重的模糊问题。相比之下，本发明在提供高保真的渲染结果的同时保证了对物体精确和平滑的几何绘制。



**图6** 在 LLFF 数据集上的可视化比较结果