学习汇报

|  |  |  |  |
| --- | --- | --- | --- |
|  | 文章名称： | 大模型学习报告 |  |
| 研究领域： | 视图合成 |
| 汇 报 人： | 杨孟衡 |
| 联系电话： | 15398615549 |

|  |
| --- |
| 一、研究主要目标、研究内容、技术关键、技术路线 |
| 1. 研究目标   目前视图合成，神经渲染，立体视觉这一些相关领域的研究报告非常多，其中又以三维重建等技术高速发展为标杆，衍生出了针对不同细节要点的新技术，如下文即将介绍的基于残差颜色学习的新视图合成。  神经隐式表示（neural implicit representation）自2019年开始兴起，在2020年NeRF获得candidate best paper之后受到了广大的关注。仅2021年一年，关于NeRF或者神经隐式表示的文章就多达150多篇，其中有数十篇高质量的文章。  目前关于NeRF或者神经隐式表示的Survey文章非常多，在个人学习中我发现了一篇刊登在《IEEE TRANSACTIONS ON IMAGE PROCESSING》的学术论文《Learning\_Residual\_Color\_for\_Novel\_View\_Synthesis》，这篇论文是由我国高科技人才Lei Han等各位大佬共同编撰的，从中学习到了一种先进的视图合成方法，基于残差颜色学习的视图合成方法。  在这篇论文中，针对NeRF技术自身的不足，大佬们提出了一系列的优化方法，以及对该领域研究的展望。本次学习报告主要研究学习了视图合成，立体视觉等领域的前沿先进学术论文，包括NeRF技术，一种基于残差颜色学习的视图合成方法。   1. 研究内容   （1）首先来介绍第一部分研究内容，NeRF技术的背后原理。  NeRF全称名为 Representing Scenes as neural Radiance Fields for View Synthesis发表于candidate，凭借其震撼的视图合成效果获得了当年的Best Paper。在之后的两年里，NeRF的引用量多达937次，仅2021年一年就有上百篇相关的文章，其中不乏高质量的文章和令人惊艳的合成效果。其中，有学者再次凭借隐式表示中的物体编辑和组合获得CVPR2021年的best paper。这些都证明了NeRF以及隐式表示在近些年的高涨热度。  提到NeRF就不得不说NeRF技术诞生时的背景。视图合成方法通常使用一个中间3D场景表征作为中介来生成高质量的虚拟视角。根据表示形式，3D场景表征可以分为“显式”和“隐式”表示。  显式表示（explicit representation），包括Mesh，Point Cloud，Voxel，Volume等。显式表示的优点是能够对场景进行显式建模，从而合成照片级的虚拟视角。缺点是这种离散表示因为不够精细化会造成重叠等伪影，而且最重要的，它们对内存的消耗限制了高分辨率场景的应用。  隐式表示（implicit representation），通常用一个函数来描述场景几何。隐式表示使用一个MLP模拟该函数，输入3D空间坐标，输出对应的几何信息。隐式表示的好处是它一种连续的表示，能够适用于大分辨率场景，而且通常不需要3D信号进行监督。在NeRF之前，它的缺点在于无法生成照片级的虚拟视角，如occupancy field、signed distance function（SDF）。  需要理解的是，无论是显式表示还是隐式表示，都是对3D场景进行表征。这种表征并不是凭空臆测或者天马行空的，而是根据现实中数据格式进行发展。例如现实中的3D数据主要有面数据、点数据、体数据，所以对应催生了一些Mesh、Point Cloud、Volume等中间表示。隐式表示则是借鉴了图形学中的一些表示形式，例如signed distance function。  （2）在理解NeRF之前，还需要理清两个概念，神经场与体渲染。  神经场已经在很多高质量的学术论文中得到论述，简单来说，场（field）是为所有（连续）空间和/或时间坐标定义的量（标量），如电磁场，重力场等。因此当我们在讨论场时，我们在讨论一个连续的概念，而且他是将一个高维的向量映射到一个标量。而神经场表示用神经网络进行全部或者部分参数化的场。在视觉领域，我们可以认为神经场就是以空间坐标或者其他维度（时间、相机位姿等）作为输入，通过一个MLP网络模拟目标函数，生成一个目标标量（颜色、深度等）的过程。  体渲染呢简而言之是从体数据渲染得到2D图片的过程。现实生活中，有一些领域采集的3D数据是以体数据格式存储的，例如医疗中的CT和MRI、地质信息、气象信息等，这些数据需要渲染到2D图像才能够被人类理解。除此之外体数据建模（Volume）相比于传统的Mesh、Point，更加适合模拟光照、烟雾、火焰等非刚体，因此也在图形学中有很多应用。  体数据的渲染主要是指通过追踪光线进入场景并对光线长度进行某种积分来生成图像或视频，具体实现的方法包括：Ray Casting，Ray Marching，Ray Tracing。  基于体渲染的研究在NeRF之前有很多，因为体渲染是一种可微渲染，非常适合与基于统计的深度学习相结合。目前可微渲染领域也有了一些研究，是未来计算机视觉和计算图形学结合的一个重要方向。  （3）了解了NeRF有关的概念以后，还要研究学习NeRF本身存在的一些问题，以及NeRF的发展。一种技术的发展，不外乎就在于技术自身问题的优化，还有就是技术实际应用的发展。本小节主要阐述学习报告中研究学习的有关NeRF的发展方向内容。  针对NeRF自身问题的优化，主要包括：  ·速度慢  ·只针对静态场景  ·泛化性差  ·需要大量视角  ·其他  基于NeRF的应用场景发展方向，主要包括：  ·逆渲染  ·可控编辑  ·数字化人体  ·多模态  ·图像处理  ·视频处理  ·特殊领域  （4）NeRF方法生产生图像时，每个像素都需要近200次MLP深度模型的前向预测。尽管单次计算规模不大，但逐像素计算完成整幅图像渲染的计算量还是很可观的。其次，NeRF针对每个场景需要进行训练的时间也很慢。  针对速度慢这问题有非常多的论文研究讨论，例如AutoInt（Automatic integration for fast neural volume rendering），Depth-supervised NeRF（Fewer views and faster training for free）等论文中论述的。AutoInt在测试时直接预测积分值，能够比NeRF快10倍    图1 AutoInt测试  针对训练时间慢的问题，Depth-supervised NeRF使用SfM的稀疏输出监督NeRF，能够实现更少的视角输入和更快的训练速度。这种思路很有意思。    图2 Depth-supervised NeRF  （5）接着我们来阐述一下静态场景问题。NeRF方法只考虑了静态场景，无法拓展到动态场景。这一问题主要和单目视频做结合，从单目视频中学习场景的隐式表示。《Neural scene flow fields for space-time view synthesis of dynamic scenes》一文中将动态场景建模为外观、几何体和三维场景运动的时变连续函数。该方法只需要一个已知摄像机姿势的单目视频作为输入。    图3 Neural scene flow fields在时间维度的映射  （6）NeRF的泛化性差。NeRF方法针对一个新的场景需要重新训练，无法直接扩展到没有见过的场景，这显然与人们追求泛化性的目标相违背。因此一些文章开始对NeRF进行泛化性的改进。《GRF:Learning a general radiance field for 3d representation and rendering》一文中，学习2D图像中每个像素的局部特征，然后将这些特征投影到3D点，从而产生通用和丰富的点表示。与之类似的还有IBRnet、pixelNeRF等，比较核心的想法都是卷积与NeRF相结合。但是这种泛化都还比较初步，无法在复杂场景中取得理想效果。    图4 GRF能够渲染没见过的物体  （7）NeRF的视角数量问题。尽管NeRF方法能够实现出色的视角合成效果，但是它需要大量的（数百张）视角来进行训练，这限制了它在现实中的应用。pixelNeRF（《Neural radiance fields from one or few images》）和GRF想法类似，使用一个CNN Encoder提出图像特征，从而使得3D点具有泛化性，并且支持少量输入。pixelNeRF能够支持一张图像输入。从我个人的经验来看，针对NeRF泛化性和视角数量的改进，目前都局限在比较封闭的测试环境下，如合成物体或者单个物体，在真实开放世界上的效果并不好。CVPR2022出现了一些开放式的工作，例如Urban-NeRF、Block-NeRF等，尝试在复杂环境下用NeRF进行建模。这个方向也比较重要。    图5 pixelNeRF的网络结构图  （8）NeRF的其他问题。还有一些工作对NeRF框架进行了改进，其中比较具有实质性突破的是Mip-NeRF。Mip-NeRF[8]提出了一种基于视锥的采样策略，实现基于NeRF的抗锯齿。Mip-NeRF减少了令人不快的混叠伪影，并显著提高了NeRF表示精细细节的能力，同时比NeRF快7%，大小为NeRF的一半。这篇文章是NeRF发展的一个重要突破，非常值得关注。    图6 Mip-NeRF使用视锥的采样策略  （9）除了NeRF自身的优化外，对应用场景的优化也不可忽视。一种重要的NeRF应用就是反向渲染。  从真实数据中估计不同模型参数（相机、几何体、材质、灯光参数）的过程称为反向渲染，其目的是生成新视图、编辑材质或照明，或创建新动画。这里我简要介绍三种比较重要的反向渲染任务：  几何与代理几何：NerfingMVS（《Guided optimization of neural radiance fields for indoor multi-view stereo》）用SfM估计的稀疏深度来监督单目深度估计网络，调整其尺度，然后再输入NeRF网络中实现视角一致性。  照明：NeRV（《Neural reflectance and visibility fields for relighting and view synthesis》）以一组由无约束已知光照照亮的场景图像作为输入，并生成一个可以在任意光照条件下从新视点渲染的三维表示。  相机（位姿估计）：《Self-calibrating neural radiance fields》在没有任何校准对象的情况下，共同学习场景的几何结构和精确的相机参数，提出了一张适用于具有任意非线性畸变的普通摄像机的摄像机自标定算法。  （10）NeRF的应用-可控编辑。虽然NeRF提供了对场景合理的表示，但是它并不允许人们对形状、外观进行编辑。对场景表示进行可控的编辑，是NeRF发展的一个重要方向。编辑的方向主要包括：形状、外观、场景组合。可控编辑一直是计算机视觉的重中之重，NeRF在这方面的发展可以单独作为一个分支来阐述，主要发展从EidtNeRF（Editing conditional radiance fields），GRAF（Generative radiance fields for 3d-aware image synthesis）到GIRAFFE (CVPR2021 Best Paper)。这些方法主要通过GAN和NeRF结合，实现了可控的编辑。这里主要展示了GIRAFFE的结构：    图7 GIRAFFE的网络结构  （11）NeRF的应用-数字化人体。数字化人体是立体视觉中的一个重要领域。NeRF跟其他3D场景表征一样，也被应用于对人体进行建模。数字化人体主要包括：  ·脸部建模  ·人体建模  ·手部建模等  脸部建模：4D Facial Avatar（《Dynamic neural radiance fields for monocular 4d facial avatar reconstruction》）将3DMM和NeRF结合，实现了一个动态神经辐射场。输入一个单目视频，该方法能够实现人脸的位姿、表情编辑。    图8 NeRF与3D MM的结合  人体建模：Animatable（《Animatable neural radiance fields for modeling dynamic human bodies》）引入神经混合权重场来产生变形场，实现了人体建模。需要输入多视角视频。这个领域目前主要向SMPL靠近，就是给定一个规范空间，或者说template，然后从不同观测空间估计规范空间。这个领域也很有意思，有一些很好的demo。    图9 Animatable的流程图  （12）NeRF的应用-多模态。目前基于NeRF的扩展工作，大部分使用的是图像、单目视频作为输入。探索其他模态如文字、音频等与图像的结合，能够催生惊艳的应用效果。CLIP-NeRF（《Text-and-Image Driven Manipulation of Neural Radiance Fields》）将CLIP和NeRF结合，实现了通过文字和图像编辑场景。目前还局限在椅子、汽车等简单模型中。进一步探索鸟、花等简单场景可能是一个方向。    图10 CLIP-NeRF实现了CLIP和NeRF的结合  （13）NeRF与神经场结合的应用-图像处理。NeRF作为一种隐式表示，为传统的图像处理方法提供了一种新思路，即从隐式神经表示，或者神经场的角度来处理图像。这里的图像处理方法包括：压缩、去噪、超分、inpainting等。目前已经有一些文章在视频压缩、去噪，图像超分等领域进行了尝试，但是对其他领域比如增强、恢复、去模糊等还没有比较完备的讨论，这一个领域可能是后面的一个方向。Neural Knitworks（《Neural Knitworks: Patched Neural Implicit Representation Networks》）提出了一种用于自然图像神经隐式表示学习的体系结构，它通过以对抗的方式优化图像补丁的分布，并通过增强补丁预测之间的一致性来实现图像合成。    图11 Neural Knitwork的网络  （14）NeRF与神经场结合的应用-视频处理。在之前的小节中，我们已经见过一些NeRF与视频的互动，但是他们主要关注Video for NeRF，即主要重点在于基于视频输入的NeRF。还有一些方法关注NeRF for Video，比如使用神经场的方法来进行视频压缩、视频编辑。这些方法证明了单目视频与NeRF或者神经场方法结合会是一个重要方向。Layered Neural Atlases（《Layered neural atlases for consistent video editing》）提出了一种将输入视频分解并“展开”为一组分层2D地图集的方法，每个地图集都提供了视频上对象（或背景）外观的统一表示。该文章能够用一组参数来存储视频，能够实现令人惊艳的编辑效果。    图12 Layered Neural Atlases的网络结构图  （15）NeRF与神经场结合的应用-特殊领域。除了之前列举的一些视觉和图形学领域的工作外，NeRF和神经场在一些特殊领域也有用武之地。这些特殊领域包括：  ·机器人  ·医疗成像  ·偏微分方程求解  这里主要展示了医疗成像的一些应用。在医疗成像中，如CT和MRI，传感器探测的数据是人不可读的，需要经过离散采样并重建成体数据(3D)或者切片(2D)供人类观看。如果能够减少采样率，则可以减少CT和MRI的时间。NeRP（《Implicit Neural Representation Learning with Prior Embedding for Sparsely Sampled Image Reconstruction》）提出一种在稀疏采样下进行神经场重建的框架，并证明可以推广到医疗图像中。    图13 NeRP的网络结构图  （16）场景表示网络(scene Representation Networks, SRN)在近年来的研究中已被证明是一种强大的新视图合成工具。 它们使用全连接网络学习从空间点的世界坐标到辐射颜色和场景密度的映射函数。 然而，场景纹理在实践中包含复杂的高频细节，这些细节很难被参数有限的网络记忆，导致在渲染新视图时出现令人不安的模糊效果。《Learning Residual Color for Novel View Synthesis》一文中通过 “残差颜色” 而不是“辐射颜色” 来进行新视图合成， 即表面颜色和参考颜色之间的残差。 这里的参考颜色是基于空间颜色先验计算的，这些先验是从输入视图观察中提取的。这样一种策略的美妙之处在于，对于大多数空间点来说，辐射颜色和参考颜色之间的残差接近于零，因此更容易学习。提出了一种利用SRN学习残差颜色的视图合成系统。在公开数据集上的实验表明，所提出的方法在保留高分辨率细节方面取得了有竞争力的性能，导致了比目前最先进的技术在视觉上更令人愉快的结果。  新视图合成作为虚拟现实应用的基本技术，旨在从场景的给定观察样本中创建新视图。通过使用同步结构化相机阵列作为捕获设备，已经显示出重大进展。然而，从稀疏视图输入进行高质量新视图合成仍然是一项具有挑战性的任务。文中认为当前隐含场景表示网络简单的编码空间坐标表示的每一个点而忽视点可能自己不同特点投影到输入时的观点。具体来说，不同视角下的反投影 观测值(记为空间颜色先验)对于朗伯曲面上的点是一致的，而对于非曲面点则有显著变化。 因此，空间颜色先验和每个点的实际辐射亮度颜色之间存在着很强的联系。基于这一观察，论文中提出了一种用于新视图合成的残差颜色学习框架。具体而言，文中将每个点的空间颜色先验作为参考颜色，并使用场景表示网络(如NeRF)来回归表面颜色与参考颜色之间的残差。 图14显示了渲染结果的分解。注意，对于大多数空间点，残差都是小值或接近于零。因此， 它们比之前直接强制网络记忆错综复杂的纹理细节的方法更容易学习。所提出方案为新视图合成保留了更清晰的细节，比最先进的方法带来了更令人满意的视觉 结果。值得注意的是，对于复杂场景，以前的方法如NeRF 存在模糊的伪影，而论文中的方法由于残差学习方案而取得了显著的改进。    图14 渲染结果对比  针对文中提出的学习方法，接下来阐述都学习了哪些相关工作。  1.基于纹理的渲染呈现  基于纹理的绘制旨在重建环境的精确彩色三维模型，以实现新视点绘制。目前已经有很多学术论文探讨了利用多视角观测和极线几何的密集匹配来重建3D模型。Elastic-Fusion使用帧到模型配准和基于窗口的surf融合。还有的使用基于空间哈希的体积融合和TSDF融合实现实时重建。除此之外随着机器学习的发展，神经网络也被用于预测显式3D模型。还有一些学者将2D特征投影到3D体素网格，并使用3D卷积得到体素模型。使用可微的基于点的渲染器来获得3D模型。点的坐标和颜色是学习参数。使用多层感知器将点云完成为网格模型。训练一个基于块的条件判别器来指导纹理优化，使其能够容忍错误对齐。其性能受限于现有3D模型的质量。在显式3D模型的帮助下，基于纹理的渲染具有良好的效率和可编辑性。然而，在重建的模型中很难避免失真、空洞和模糊的部分，尤其是对于凌乱的场景。生成模型的不足会在渲染图像时带来伪影和模糊的细节。  2.基于图像的渲染呈现  与基于纹理的渲染不同，基于图像的渲染无需明确的3D模型即可生成新颖的视图。通过转换采样图像生成新视图。采样图像扭曲成一个新颖的观点基于相机姿态估计的估计。有的学者使用贝叶斯估计来获得新视图中每个像素的颜色值。  神经网络被广泛应用于隐式场景表示。它在记忆场景方面显示出巨大的潜力，包括几何和纹理。几何由于其低频性质，可以很容易地用神经网络来表示，而高频纹理细节则很难被神经网络记忆。在不同层生成不同透明度的多平面图像，通过对多平面图像的整合可以得到新的视图。通过变换的相邻多平面图像的加权组合生成新视图，这些图像由相应的透明度调制。推导出用于TSDF值预测的可微体绘制。将世界坐标映射到局部场景属性的特征表示，并使用场景表示网络来预测不同类型场景的特殊网络。使用编码器根据多视图图像产生一个潜代码z，然后将其解码为一个volume，该volume为每个体素提供颜色和透明度值。NPBG将一组RGB视图和一个点云作为输入。对每个点拟合一个神经 描述子，之后可以渲染一个场景的新视图。FVS通过多视图立体计算输入图像的3D代理几何。给定一个目标视图，根据投影深度将附近的源图像映射到目标视图中，然后使用循环卷积网络将映射后的图像进行混合。这两种方法都需要高质量的3D几何图形作为输入。渲染性能很大程度上受点云或重建的3D几何的质量影响。如果用于映射的 3D模型遗漏了场景的大部分或有粗糙的异常值，管道将产生可见的伪影。NeRF通过多层感知器表示一个场景，并通过体绘制对其进行训练。使用位置编码和分层采样来提高渲染性能。  隐式场景表示在真实感绘制中显示出巨大的潜力，但仍然是一项具有挑战性的任务。NeRF使用多层感知器和体渲染进行隐式表示。它实现了显着的渲染性能改进，并且有许多方法可以提高NeRF。NSVF和我们的方法都旨在提高不同角度的新视图合成的渲染质量。NSVF利用三维空间中表面稀疏的先验性，只需要处理经过表面的体素，并使用局部参数来提高场景表示网络的能力。不同的是，我们的方法提出了空间颜色先验，通过从投影像素计算参考颜色来降低高频纹理细节的学习难度。Nerfies引入变形代码来处理动态场景，并使用外观代码来处理光线变化。KiloNeRF利用数千个微小MLP来取代原来单一的大型 MLP进行加速。我们的方法是对此类方法的补充。为了提高保持高频纹理的能力，论文基于多视角观察提出了新的基于残差的多视角先验。利用所提出的空间颜色先验，为高质量的隐式场景表示引入了残差学习方案。其中该方法将一组稀疏的视图作为输入，旨在给定的视点上渲染新视图。整体框架如图15所示。我们根据输入的多视角观察提出“空间颜色先验”，而被遮挡的像素则通过提出的patch特征滤波器被去除。参考颜色通过投票策略从空间颜色先验中获得。在此基础上，在隐场景表示网络中引入残差颜色学习方案，以降低对高频信息的网络容量要求。    图15 残差颜色学习框架  （17）在接下来的内容中，首先阐述该小节中的隐式场景表征，然后详细阐述小节中的空间颜色先验和小节中的残差颜色学习方案，其中具体的实现细节留到技术路线小节阐述。  1.隐式场景表征  采用全连接网络来隐式描述场景。它学习一个函数，该函数将连续的3D坐标映射到这些特征坐标处场景的特征表示。对于不同的目标，特征表示可能被转换为诸如密度或符号距离函数(图16)等属性。代表性的SRN方法NeRF将场景建模为神经辐射场，并应用体渲染进行新的视图合成。每个空间点由其3D坐标p=(x, y, z)和视角方向dr= (θ，φ)表示，使用全连接网络将其映射到密度(不透明度)σ和辐射度颜色c。摄像机光线r的期望颜色C(r)可以是由经典的体绘制技术绘制，如式1所示。其中tn和tf分别是r的近界和远界，dt是相机光线之间的距离。dr表示r的视点方向，“exp”为指数函数。基于体积绘制，式1的连续积分可以用数值正交代替:σ i、ci用一个全连接的网络Fθ(pi、dr)表示，分别表示第i个采样点的颜色和密度。δi表示两个采样点之间的距离。C(r)¯ 是根据权重wi将射线中的所有采样点相加计算出来的。 这里Fθ(pi， dr)可以通过最小化渲染视图C(r)¯和观察视图C(r)之间的差异来从给定的稀疏输入视图中学习，其中R是所有相机射线的集合，它的个数等于所有图像像素的个数:    图16 特征表示    式1    式2    式3  2.空间颜色先验  回想一下,现场几何和纹理信息隐含在颜色的一致性的多视图的观察,在此基础上,我们提出“空间颜色先知先觉”和剩余颜色学习计划来减少高频信息的网络容量需求。首先将空间点投影到观测图像上，得到其投影直方图。训练图像记为I ={Ii， I∈N}，对应的相机位姿记为H ={Hi， I∈N}。我们计算当前相机姿态Hc与H之间的距离，并从训练图像I中选择M个最近的图像，局部图像为Ilocal = {Ilocali, I∈M}。然后基于多视图几何计算反投影像素:其中K是相机的内蕴。u ={ui， i∈M}是点p在局部图像 Ilocal中的投影像素，点p的投影直方图被定义为u的统计直方图。采样点离物体表面的远近会导致投影直方图的特征不同。 图17（a）（b）说明了非表面点和表面点两种情况下的投影直方图。对于非表面点，不同视角的观察结果是不相关的，从其散乱投影直方图可以看出。    图17（a） 一个非平面点的投影像素直方图    图17（b） 一个平面点的投影像素直方图  从上面两幅图可以注意到非平面的直方图是分散的，而平面点的直方图是集中的，因此可以根据我们提出的空间颜色先验稳健地估计参考颜色。其他非平面点以及平面点也有类似情况。对于物体表面上的点，不同视角的观测结果是一致的，其投影直方图是集中的。由于投影直方图的颜色一致性隐含了场景几何和纹理信息，因此对于每个空间点，我们基于其投影直方图中的信息提出了“空间颜色先验”。如果该点在朗伯曲面上，除了被遮挡的像素外，投影像素是相似的。由于被遮挡的像素与其他投影像素无关，因此对于空间颜色先验来说，它们是无意义的噪声。为了处理它，我们采用了一个patch特征过滤器来从投影直方图中去除被遮挡的像素。同一3D点在不同视角下的局部图像块， 除了遮挡外，期望是相似的，这适合于遮挡去除。使用半尺寸图片的3 × 3块作为像素特征 因为在相同的局部块大小下，下采样图像具有更大的感受。 将投影像素的patch特征与当前视图进行比较。我们计算l2范数，删除与当前视图差异大于阈值的像素。提出的patch filter是一种简单但有效的方法来处理周围场景中的多个遮挡。它不需要非常精确，因为残差颜色预测将补偿小偏差。对于训练，当前视图的patch特征是从训练图像中提取的。对于推理，当前视图的patch特征是从预测的辐射亮度颜色C中提取的。有了patch特征过滤器，被遮挡的像素就可以被移除。之后，我们通过基于剩余投影像素u的投票策略来计算参考颜色cref。虽然我们通过特征滤波去除被遮挡的像素，但一些具有强反射率的投影像素仍然可能影响参考颜色的计算。因此我们计算u的均值，然后从均值中删除大于阈值的值。具有强反射率的像素通过投票策略被移除。然后我们通过剩余像素的均值来计算参考颜色。注意，没有特征滤波器、残差学习的空间颜色先验已经在大部分区域有了明显的性能提升，但在遮挡时引入了较小的伪影。我们引入特征过滤器来处理遮挡。但是，直接使用特征滤镜会带来更差的结果，这是因为虽然特征滤镜提供了更准确的参考颜色。 它还会使一些非表面点的投影直方图更加集中，从而导致密度预测的准确性降低。特征滤波器必须与Eq.8的联合训练相结合，以增强密度预测的鲁棒性。然后遮挡中的伪影就会被成功丢弃。  3.残差颜色学习方案  通过对空间点的参考颜色计算，论文中提出了一种残差颜色学习方案，将空间颜色先验应用于新视图合成。对于每个空间点，我们根据之前小节所示的空间颜色先验计算其参考颜色cref，并通过SRN F（θ）预测其残差颜色cr。 参考颜色和残差颜色被组合为预测颜色c，用于光线R处颜色c¯R的体绘制， 如下所示:    式5  对于朗伯表面上的点来说，不同视图的像素颜色是相似的。通过稳健的参考颜色计算，不同视点的残差颜色预测比原始的辐亮度颜色预测小得多。将复杂高频纹理细节的学习任务简化为学习大多数空间点接近0的残差颜色，显著减轻了网络的负担。    式6  然而，我们也观察到，仅根据Eq.6所示的残差颜色来学习网络可能会导致过拟合，因为如果非表面点的参考颜色与目标颜色相似，则可能会将其分配给非零密度。辐射颜色和残差颜色可以有各自的密度预测。然而，为了增强 引入特征滤波器后密度预测的鲁棒性，我们提出了一种联 合训练方案，即通过学习相同密度的残差颜色和辐亮度颜色来利用辐射亮度颜色损失进行密度预测:    式7  δ i、 ci和cri是全连接网络的输出Fθ(pi、dr)。δi是密度预测。ci和 c r i分别是辐射色和残差色输出。网络由辐射度图像C¯W(r)和残差图像C¯R(r)的渲染损失共同训练:    式8  所提出的残差颜色学习方案大大减轻了网络的负担。因此， 我们提出的方法比NeRF获得了更好的性能，并且迭代次数更少(图18)。   |  |  |  | | --- | --- | --- | |  |  |  | | 图18（a）Attic | 图18（b）Library | 图18（c）Family |   至此，研究学习内容基本阐述完毕，之后实验细节，关键技术再进一步阐述学习内容。   1. 关键技术   在上一节中我们阐述了研究分析内容，接下来将以NeRF技术以及基于残差颜色学习的新视图合成技术顺序进行阐述关键技术。  （1）NeRF的关键技术。NeRF首次利用隐式表征表示实现了照片级的视角合成效果，与之前方法不同的是，它选择了Volume作为中间表示，尝试重建一个隐式的Volume。NeRF的主要关键技术如下：  ·提出了一种5D neural radiance field 的方法来实现复杂场景的隐式表示  ·基于经典的Volume rendering提出了一种可微渲染的流程，包括一个层级的采样策略  ·提出了一种位置编码（positional encoding）将5D坐标映射到高维空间  （2）基于残差颜色学习的新视图合成技术主要是两方面的技术创新。  ·空间颜色先验：鉴于多视角观测传达了辐射亮度颜色的先验信息所提出的学习框架配备了基于输入视图观察的空间颜色先验，这是隐式场景表示网络的补充信息，该网络只是将点的世界坐标映射到局部场景属性。  ·残差颜色学习：通过将提出的空间颜色先验作为参考，提出了一种残差颜色学习框架，以回归表面颜色和参考之间的残差。对于大多数空间点来说，残差接近于零，因此比之前直接回归表面颜色的工作更容易学习。在神经渲染中提出的残差学习框架简单而有效，可以很容易地与其他隐式场景表示方法相结合。   1. 技术路线   在前面小节的论述下，接着详细介绍基于残差颜色学习的新视图合成方法是如何进行实验的。  我们为输入视图监督的每个场景训练一个SRN。网络架构如图19所示。 在训练步骤中，从训练视图中随机采样像素射线。采用NeRF中提出的分层抽样策略对体积空间进行更有效的采样。 它优化了两个网络:一个粗一个细。粗网络使用分层抽样，细网络根据粗网络的输出使用更有信息的抽样。这个过程将更多的样本分配到我们期望包含可见内容的区域。在训练阶段为所有采样点计算空间颜色先验，而为了提高效率，只计算Eq.2)中权重(wi大于10−3的点。    图19 网络架构  为了与之前的方法进行公平的比较，我们在各种数据集上评估我们的方法:来自LLFF的正面数据，来自NeRF的合成数据，来自休闲3D数据集的室内周围数据，自收集的室外大规模数据(表1中的“礼堂”和“剧院”)，以及来自另一篇论文中的坦克和寺庙的包围数据数据集。图20显示了不同种类的不同射击轨迹的数据。    表1   |  |  | | --- | --- | |  |  | | 图20（a） | 图20（b） | |  |  | | 图20（c） | 图20（d） |   接下来，通过定量评估来验证所提方法的性能。  1.定量评估  定量评估采用PSNR、SSIM和LPIPS进行评估。PSNR和SSIM值越小，精度越高，LPIPS值越高，视觉质量越好。我们将我们的方法与包括SRN、NV、LLFF和NeRF在内的先前的技术水平进行了比较，如表1所示。对于视野范围较小的简单场景，如LLFF数据集中的“房间”和“堡垒”，NeRF在足够的内存容量下取得了良好的性能。空间颜色先验有助于揭示高频细节，改进相对较小。对于具有大规模周围视图的复杂场景，例如来自休闲3D数据集的“Library”和“Attic”，由于网络大小的限制，NeRF表现不佳。所提出的方法取得了更好的性能，因为所提出的空间颜色先验有助于降低大规模场景的网络容量要求。如图21所示，随着网络规模的增长，NeRF的性能会越来越好，这说明NeRF的渲染质量受到其网络容量的限制。然而，更大的内存大小需要更多的复杂度，这就限制了尺寸从增加太多。网络增长带来的改善也很小。另一方面，在提出的空间颜色先验的帮助下，对网络容量的要求大大降低，我们提出的残差学习方案即使在较小的网络上也实现了更好的质量。我们还比较了NeRF和我们提出的在不同分辨率下呈现新视图的方法的性能，如表2所示。对于更高的分辨率，我们的方法和NeRF之间的差距更大，证明了我们提出的方法在高分辨率下生成逼真渲染结果的能力。    表2    图21 NeRF和新合成方法在随机3D数据集“Library”中的比较  报告旨在阐述一种改善自由移动摄像机合成新视图的沉浸式体验，其中研读的论文认为试图使用全连接网络记忆环境的纹理细节和几何形状的传统场景表示网络在实践中无法保留高频细节，并提出了一个新的框架，通过使用所提出的空间颜色先验作为辐射度颜色预测的参考来学习残差颜色。实验表明，所提出的方法取得了比以前的技术更令人愉快的视觉结果，特别是对于包含复杂纹理和大表面积的环境。所提出的方法对朗伯曲面效果最好，仅对非兰伯量曲面达到与先前方法相当的性能。 |

|  |
| --- |
| 二、项目的创新性(理论创新、技术创新等方面，根据学习材料的具体内容描述)600字及以上 |
| 本项目/文章针对NeRF视图合成技术以及该项技术自身发展，阐述了具有良好效益的优化方案，研发了新的技术路线及产品，实现了基于残差颜色学习的新视图合成模型，有效解决大模型进行视图合成难题。具体研究意义与创新点如下：   1. 场景表示网络在近年来的研究中已被证明是一种强大的新视图合成工具。 它们使用全连接网络学习从空间点的世界坐标到辐射颜色和场景密度的映函数。 然而，场景纹理在实践中包含复杂的高频细节，这些细节很难被参数有限的网络记忆，导致在渲染新视图时出现令人不安的模糊效果。 2. 报告中介绍的新视图合成方法是建议学习“残差颜色”而不是“辐射颜色” 来进行新视图合成，即表面颜色和参考颜色之间的残差，其中参考颜色是基于空间颜色先验计算的，这些先验是从输入视图观察中提取的。 3. 这样一种利用残差色训练模型的策略的美妙之处在于，对于大多数空间点来说， 辐射颜色和参考颜色之间的残差接近于零， 因此更容易学习。 4. 文中作出的主要技术创新就是两点：   ·空间颜色先验：鉴于多视角观测传达了辐射亮度颜色的先验信息所提出的学习框架配备了基于输入视图观察的空间颜色先验，这是隐式场景表示网络的补充信息，该网络只是将点的世界坐标映射到局部场景属性。  ·残差颜色学习：通过将提出的空间颜色先验作为参考，提出了一种残差颜色学习框架，以回归表面颜色和参考之间的残差。对于大多数空间点来说，残差接近于零，因此比之前直接回归表面颜色的工作更容易学习。在神经渲染中提出的残差学习框架简单而有效，可以很容易地与其他隐式场景表示方法相结合。   1. 通过真实感渲染方式，其中包括基于纹理的渲染以及基于图像的渲染方式来对图像进行渲染合成，能有效的提升图像清晰度。 |

附件：知识树

|  |
| --- |
|  |