

基于神经表示的三维重建建模

刘烨斌¹ 章国锋² 周晓巍² 许 岚³

¹ 清华大学

² 浙江大学

³ 上海科技大学

关键词：三维重建 神经表示 三维视觉

神经表示的种类

传统的描述物体的三维空间信息的表示方法包括点云 (point cloud)、体素 (volume) 以及网格 (mesh) 等离散表示。近年来, 基于神经网络的表示, 尤其是基于隐函数的表示方式在高精度重建等问题上展现了极佳的效果, 得到了计算机图形学领域研究学者的广泛关注。

基于隐函数的神经表示 (也称为基于坐标的神经表示) 以空间点的坐标为输入, 输出物体在此空间点的信息。相较于之前基于体素的表示方式, 基于隐函数的表示方式允许在任意空间分辨率下采样表示三维物体, 这使得基于隐函数的神经表示在高精度重建领域取得了极佳的效果。具体而言, 基于隐函数的神经表示输出物体在某一个空间点的相关信息, 这种信息包括占用 (occupancy)、符号距离函数 (Signed Distance Function, SDF) 以及神经辐射场 (Neural Radiance Fields, NeRF) 等。

占用是指空间点在物体内外信息, 基于占用的隐函数神经表示以占用网络^[1] (occupancy networks) 为代表。具体而言, 占用网络使用网络从空间点坐标和表示物体的特征向量中估计占用, 并使用深度神经网络分类器将三维物体表面表示为连续的决策边界, 而且可以在较小的内存占用下, 实现以任意分辨率输出物体的三维结构信息。这一表示可以从包括图像、点云和离散体素等在内的多种不

同输入推断出物体的三维结构, 并在相应的三维重建任务中取得了很好的效果。

符号距离函数定义为一个空间点到物体表面的距离, 符号表示空间点在物体内/外, 基于符号距离函数的隐函数神经表示以 DeepSDF^[2] 为代表。具体而言, DeepSDF 使用一个全连接网络直接从 3D 坐标和潜在编码回归出符号距离函数。DeepSDF 隐式地将形状的边界编码为符号距离函数场的零水平集, 以稀疏的点云观测优化网络, 得到重建。之后, Yariv 等人提出了基于符号距离函数的可微分渲染器^[3], 这使得网络可以通过多视角图片进行监督训练。除了应用在基于隐函数的神经表示, 符号距离函数也经常应用在基于体素的神经表示中, 例如, Atlas^[4]、NeuralRecon^[5] 等工作通过使用卷积神经网络估计场景中的符号距离函数以完成重建, 这些方法均需要三维的符号距离函数地面真值的监督。

神经辐射场将每个空间点的物体信息建模为辐射 (radiance) 和体密度 (volume density)。给定场景的多个图像, 神经辐射场^[12] 推断出最能解释这些图像的几何 (以体密度场为载体) 和外观 (以辐射场为载体)。神经辐射场表明, 使用多层感知机能够表示复杂场景, 并且能借助体渲染技术渲染出照片般逼真的新视图, 其效果优于先前最先进的工作。神经辐射场通过可微分的体渲染技术渲染图片, 用多视角图片即可完成对网络的监督训练。

除了上述基于隐函数的神经表示, 也有一些

结合显式与隐式的神经表示方法。例如,神经稀疏体素场(Neural Sparse Voxel Fields, NSVF)^[9]、DeepLS^[27]以一个显式的稀疏体素场表示物体粗糙的几何,在每一个体素中储存了一个潜在编码,利用网络结合此潜在编码与局部的坐标估计此点的物体信息。这样的表示允许我们在渲染时能够直接估计场景内容范围内的信息,跳过不包含相关场景内容的体素,从而加快渲染速度。此外,显式的稀疏体素表示还可以方便地应用于场景的编辑与合成。

神经表示的研究趋势

基于神经表示重建的快速优化研究

基于隐函数表示的重建(包括即时解码刷新(Instantaneous Decoding Refresh, IDR)^[27]、可微分体积渲染(Differentiable Volumetric Rendering, DVR)^[34]、NeRF^[12]等)往往是在一个场景的多视图上训练得到此场景的重建,但无法泛化到新场景,导致基于此表示获取一个重建结果需要数小时乃至数天的网络训练时间。因此,许多研究者致力于加速基于神经表示重建的训练过程。

直观的想法是利用网络在多个场景学习重建,使网络学习到不同场景的先验,从而能够在新场景上实现快速训练或直接泛化。MetaSDF^[20]、MetaNeRF^[23]将元学习的方法应用在基于隐函数表示的重建问题上,展现了元学习的方法可以大大加快隐函数表示重建的收敛速度。另一类方法致力于利用潜在编码控制估计神经辐射场。生成辐射场(Generative Radiance Fields, GRAF)^[19]使用分别表示场景几何和外观的两个潜在编码控制估计神经辐射场,从而实现在新场景上的泛化。PixelNeRF^[28]、基于图像的渲染网络(Image-Based Rendering Network, IBRNet)^[25]利用卷积神经网络获取图像特征,然后在空间点上融合来自多视角的图片特征,利用融合后的特征结合空间点坐标估计此空间点的占用或神经辐射场信息。多视角三维重建神经辐射场(MultiView Stereo NeRF, MVSNeRF)^[13]进一步展

示了通过使用三维代价体(cost volume)中的特征可以更好地估计空间点的物体信息,利用图像特征的方法可以直接泛化至新场景,也可以通过进一步的细化快速获得更精细的重建。

基于神经表示重建的快速渲染研究

基于NeRF^[12]技术的渲染速度是非常慢的。为了渲染图像中的一个像素,NeRF从相机中心投射出一条穿过此像素的射线,在此射线上采样数百个点,计算每个点的神经辐射场,利用体渲染技术将此射线上的神经辐射场聚集为此像素的颜色。渲染一张分辨率为800×800的图片,NeRF往往需要花费半分钟到数分钟不等。

为了提高NeRF的渲染速度,NSVF^[9]提出使用稀疏的体素场组织场景,允许NeRF仅在场景中不为空的地方做采样并进行体渲染。NSVF的渲染速度是NeRF的10倍,但仍然不能达到实时的要求。PlenOctree^[29]、FastNeRF^[6]、BakingNeRF^[8]、NeX^[35]等工作提出对场景中的神经辐射场进行预计算并且缓存起来,从而实现渲染时的加速。由于辐射场在各个角度上是不一样的,缓存360度视角的辐射场需要极大的存储空间,因此这些工作使用球谐函数等技术解耦辐射场与依赖视角的效果,仅缓存不依赖视角的辐射场。但是这些方法针对每个场景需要大量额外的储存空间。DONeRF^[14]尝试使用一个多层感知机估计像素的深度,然后在深度附近采样少量的点,从而实现加速,但这种方法需要深度图像的监督。

基于神经表示的动态场景重建研究

近两年,许多工作尝试将针对静态场景的神经表示扩展到动态场景下。神经体(neural volumes)^[38]首次实现了基于神经表示的动态场景建模和渲染,其核心思想是将动态场景分解为两个特征体,一个表示参考姿态下的场景几何和纹理,另一个表示从参考姿态到每一帧的变形。占用流(occupancy flow)^[36]在隐式占有场的基础上,针对动态序列学习一个连续的速度场,根据每一时刻速度场的取值,参考帧的占有场可以变形到其他动态帧上,从

而实现动态序列的时空联合隐式建模。神经变形图(neural deformation graph)^[37]则利用嵌入式变形图学习序列中的动态变形,实现从点云序列中建立全局一致的动态隐式表示序列。

也有许多工作致力于研究将NeRF扩展到动态场景上。Nerfies^[17]、D-NeRF^[18]使用一个额外的多层感知机对视频的每一帧学习一个空间变形场。神经场景流场(Neural Scene Flow Fields, NSFF)^[31]在NeRF的输入基础上加入时间维度,通过额外预测深度和场景流进行正则化实现动态场景的重建。时空神经辐照度场(space-time neural irradiance fields)^[26]简单地使用时间作为附加输入,但需要仔细选择损失才能成功训练此方法来渲染动态场景,并且需要深度图像的监督。NR-NeRF^[24]也使用额外的建模可变形场景的多层感知机来模拟非刚性场景,除了相机参数之外,它不依赖于预先计算的场景信息,但是与Nerfies相比,产生的输出锐利度略低。自监督跟踪与重构(Self-supervised Tracking and Reconstruction, STaR)^[30]将多视图RGB视频作为输入,并将场景分解为静态和动态的体,但目前它只支持场景中一个物体的运动。此外,也有一些工作专注于人物的肖像。动态神经辐射场(Dynamic Neural Radiance Fields, DNRF)^[22]通过将可变形人脸模型加入到NeRF的方法中来增加归纳偏置,在动态人脸的重建上展现了比较好的效果。NeuralBody^[39]则通过引入参数化人体模板来关联稀疏视角下的观测信息,实现了针对人体运动的自由视点生成。

基于神经表示的重新光照研究

NeRF在静态场景渲染上取得了非常优秀的表现,但是不支持场景重新光照。近期有许多工作致力于研究神经渲染在重新光照方面的改进。NeRF-W^[10]通过改进NeRF在动态场景和光照变化数据集的表现,实现了可变光照建模和瞬时遮挡去除。具体而言,该方法通过引入可学习的隐式外观编码对不同拍摄条件下同一场景的光照变化进行建模,并通过学习同一场景不同图像的共享组分来分离出瞬时遮挡元素(例如游客或其他前景物的遮

挡)。神经反射场(neural reflectance fields)^[11]在NeRF的基础上加入了局部反射模型,同时预测并编码了场景的体密度、法向和反射率,并使用一个基于物理的可微光线追踪框架渲染图像,实现了镜面反射、阴影和遮挡等重新光照效果。神经反射和可见度场(NeRV Fields)^[21]提出了一种从环境和间接照明场景的图像中恢复可重新照明的神经体表示的方法。NeRV通过反射环境光照将NeRF的光学模型从场景辐射场修改场景反射场,实现了场景可视化。输入已知光照的场景图像集,NeRV可以学习出基于双向反射分布函数(Bidirectional Reflectance Distribution Function, BRDF)的场景材质信息,并实现在手动光照条件下的场景新视图渲染。值得一提的是,NeRV使用神经网络来学习一个可见度场的隐式查找表,以近似体渲染积分,从而加速了渲染时光照可见性的采样计算。

也有工作积极探索将场景分解为显式表示,以便于利用经典的渲染框架实时生成指定光照参数的新视图。神经反射分解(Neural Reflectance Decomposition, NeRD)^[16]使用形状、球形高斯光照、BRDF等基于物理的显式表示来分解表达物体在特定场景下的辐射场,并支持物体在旋转和改变背景时根据光照变化重新渲染,但是无法对阴影效果建模。

神经辐射因子分解(Neural Radiance Factorization, NeRFactor)^[32]解决了先前方法大多需要已知光照的问题,可以从未知光照条件的多视角图像中恢复出物体的形状和随空间变化的反射,能够在任意环境光下渲染物体的新视图和编辑物体的材质属性。NeRFactor将物体的神经辐射场表示的体几何提取为表面表示,并联合优化几何,同时解出随空间变化的反射和基于高动态范围图像(HDR)贴图的环境光照。具体而言,NeRFactor学习表面法线、光可见度、反照率和数据驱动的BRDF的三维神经场,通过对光的可见度进行显式建模,能够从反照率中分离阴影,从而支持在指定的环境光照贴图下合成逼真的阴影。

基于神经表示的场景合成研究

早期的场景编辑和新视图合成方法采用传统的

建模和渲染流程,虽然用传统的显式场景表示可以分解场景对象,但很难合成逼真的视图。在神经隐式表示取得优异效果后,将使用神经隐式表示的对象用于体渲染场景合成也成为了热门研究方向之一。

神经场景图(neural scene graphs)^[33]提出了一种基于学习的场景图表示方法,使用神经网络隐式表示动态多对象场景,通过隐式编码对象变换和辐射,能够将动态场景表示和分解为单个场景对象,支持渲染重新排列对象的场景新视图。GIRAFFE^[15]能够分离单个物体,并支持在场景中平移和旋转它们以及改变相机姿势。GIRAFFE通过将场景表示为合成的神经特征场来支持对象分解,相比于其他体渲染方法直接渲染颜色,GIRAFFE输出特征向量,平均后以低分辨率体渲染到二维特征图,最后进行上采样(upsampling)。以对象为中心的神经散射函数(Object-Centric Neural Scattering Functions, OSF)^[7]自底向上地联合渲染可编辑场景,能够渲染出非常逼真的重新光照和自由排列物体的合成场景。OSF使用多层感知机对物体内部的光反射场进行建模,通过光线采样迭代计算物体之间的光反射和阴影,对物体内部和物体之间的光传输效果进行建模,包括镜面反射、间接照明、遮挡和阴影。但是OSF需要在只包含单独物体的图像上训练其辐射场模型,这意味着它只能使用训练好的物体库中的物体,限制了它在实际场景中的应用。

神经表示的人体三维重建与绘制研究

基于神经表示的人脸重建

传统的三维人脸重建研究往往基于通过主成分分析(Principal Components Analysis, PCA)算法生成的三维人脸形变模板(3D Morphable Model, 3DMM)^[40],存在表现力高度受限于数据质量、难以表现面部细节等缺点。近年来随着深度学习技术的发展,出现了一系列基于NeRF、生成式对抗网络(Generative Adversarial Networks, GAN)和

SDF的算法,在重建精细度上取得了进步。

NerFACE^[41]利用NeRF的基本结构,将原本的相机视角参数调整为基于3DMM拟合的表情、姿态参数,仅输入一段约2分钟的单视点头部视频,即可实现新视点、新表情下头部视频的生成,但不受3DMM控制的颈部区域仍存在一定程度的抖动。Huiwen Luo的工作^[42]利用可生成高真实感图片的StyleGAN改进传统3DMM算法,实现了具有更高表现力的三维人脸模板,其效果相对于传统PCA算法在纹理精细度和真实感上有明显提升。i3DMM^[43]利用SDF算法,创新性地提出了包含头发在内的3DMM算法,丰富变化的头发区域难以利用传统PCA算法进行表示,而隐式神经算法为其模板化表达提供了可能,尽管i3DMM的效果仍较为粗糙,但这一工作也展示出了隐式神经算法在三维模板化表达上的潜在能力。像素解码器虚拟化身(Pixel Codec Avatars, PiCA)^[44]结合卷积神经网络和类似NeRF结构的多层感知机渲染结构,实现了对单人头部的自由视点驱动与渲染,相对于传统三维重建方式,基于真实图片数据集的神经渲染方式取得了更高层次的真实感。

基于神经表示的人手重建

早期对于人手的三维重建工作大多基于简单的人手三维骨架表达。重建方法主要依赖于深度卷积神经网络,通过对二维热力图的估计和图片隐变量的回归来重建人手的三维骨架,代表工作有C.Zimmermann^[45]等人的工作和GANerated Hand^[46]等。由于单纯的三维骨架无法很好地表达人手的皮肤表面几何,人们提出一些人手模板的表达方式,其中最著名的就是MANO^[47]神经模型,该模型具有参数驱动、可微分的特性,可以很好地适应深度学习的网络框架,人们基于此提出了一些人手重建的方法。

HAMR(HAnd Mesh Recovery)^[48]提出了一种从单张RGB图片重建整个人手三维模型的方法。通过深度卷积网络提取图像特征估计人手关键点的二维热力图,最后得出MANO模型的姿态形状参数和相机的投影视角参数,从而重建出完整的人手

三维模型。Adnane Boukhayma^[49]等人同样实现了基于单张人手RGB图片的三维重建。他们通过深度残差网络提取图片的全局特征,从而估计MANO的驱动参数,该方法的特点在于对于图片运动模糊和自然光照条件具有较好的效果。Liu hao Ge^[50]等人提出了基于图卷积表达的三维人手重建方法,该方法通过二维卷积网络提取图片信息,通过人手表面的图卷积网络来传递表面顶点信息,最终直接重建出完整的三维人手模型。

除了基于传统的三角网格的表面重建,近些年人们提出了基于深度隐式距离场的三维表达方法,GraspField^[51]就是一种人手的隐式距离场表达方式,该方法通过多层线性感知器实现条件距离场,从而重建出与物体交互时的人手与物体的三维表面。

基于神经表示的人体重建

传统的动态人体重建往往依赖复杂的多相机多灯光拍摄系统,而深度相机和GPU的流行引领了更轻便的容积重建,如经典的DynamicFusion^[52]和Fuison4D^[53]算法。近两年,许多工作进一步探索了基于不同数据类型的人体可微神经渲染,实现了基于简易输入的高真实感人体动态几何和纹理重建。

神经人体渲染(Neural Human Rendering, NHR)^[54]使用PointNet++提取动态点云特征,结合基于点云的可微渲染模块,实现从低质量的三维动态点云序列直接生成高质量的人体外观渲染结果。小样本神经人体渲染(Few-shot Neural Human Rendering, FNHR)^[55]进一步研究了基于点云的人体渲染在时域和空域上的数据冗余性,实现仅及于小样本的快速训练。

基于隐式神经表达的方法也在人体几何重建和纹理渲染上取得了很大的突破。像素对齐隐函数(Pixel-aligned Implicit Function, PIFu)^[56]结合像素对齐的隐式特征和多层感知机网络结构从单张RGB图片上学习完整和带细节的人体几何模型。参数化模型条件隐式表示(Parametric Model-Conditioned Implicit Representation, PaMIR)^[57]进一步引入参数化人体模板的语义特征,对深度隐函数进行正则

化,提高了在高难度人体姿势和各种服装拓扑下的人体几何重建泛化能力。NeuralHumanFVV^[58]使用6路RGB输入,进一步实现了实时的隐式人体几何重建,并且结合混合权重学习实现高分辨率的实时高真实感人体外观纹理生成。

此外,一些方法还依托网格的表达方式对传统基于三维骨骼的人体模型进行神经网络化驱动。DeepCap^[59]使用多视图弱监督的方式实现了高质量的单目人体姿态和非刚性形变参数估计;深度动态角色(Deep Dynamic Characters, DDC)^[60]进一步将这种弱监督方式拓展到人体动态外观纹理上,依赖图卷积网络,实现可控的实时人体动态纹理和几何变形生成。

基于NeRF的方法可以实现高真实感人体渲染。NeuralBody^[39]通过引入参数化人体模板来关联稀疏视角下的观测信息,实现了针对人体运动的辐射场生成和人体外观渲染。神经角色(neural actor)^[61]使用数以百计的视角信息进行训练,结合粗糙的人体模板作为神经辐射场约束,使用二维纹理图作为潜在变量来预测残余变形和动态外观,实现了任意视角和任意可控姿势的高质量人体外观渲染。ST-NeRF^[62]使用16路RGB视频输入,实现了较大动态场景的可编辑自由视点视频生成,并且场景中的每一个动态人体都被单独进行分层神经表示和建模,从而实现了人体对象的比例操控、位置复制或重新定时等功能,在动态场景下实现高质量、高真实感、可编辑的自由视点视频生成。

展望

三维重建的最终目标是服务于场景的理解和交互。现有的研究主要关注神经三维表示的设计和训练,尚未考虑如何将神经表示应用到三维场景的理解和交互(如识别、分割、跟踪、预测、编辑)之中。此外,现有的研究工作往往在单个物体或场景的过拟合上表现出色,却在泛化能力上存在瓶颈。展望未来,今后的研究将在现有神经表示的基础上,一方面提高泛化能力,另一方面着眼于应用性能更强的神经表示提高三维场景的理解能力和交互水平。■



刘焯斌

CCF 专业会员。清华大学自动化系长聘教授。主要研究方向为三维视觉、计算机图形学、计算摄影学。

liuyebin@mail.tsinghua.edu.cn



章国锋

CCF 专业会员。浙江大学计算机科学与技术学院教授。主要研究方向为三维视觉和混合现实。

zhangguofeng@zju.edu.cn



周晓巍

CCF 专业会员。浙江大学“百人计划”研究员。主要研究方向为三维计算视觉和图形学。

xzhou@cad.zju.edu.cn

其他作者：许 岚

参考文献

- [1] Mescheder L, Oechsle M, Niemeyer M, et al. Occupancy networks: Learning 3d reconstruction in function space[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019:44604470.
- [2] Park J J, Florence P, Straub J, et al. Deepsdf: Learning continuous signed distance functions for shape representation[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2019:165174.
- [3] Yariv L, Kasten Y, Moran D, et al. Multiview neural surface reconstruction by disentangling geometry and appearance[OL]. ArXiv preprint arXiv:2003.09852, 2020.
- [4] Murez Z, van As T, Bartolozzi J, et al. Atlas: End-to-end 3d scene reconstruction from posed images[C]// *Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII* 16. 2020:414431.
- [5] Sun J, Xie Y, Chen L, et al. NeuralRecon: RealTime Coherent 3D Reconstruction from Monocular Video[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021:1559815607.
- [6] Garbin S J, Kowalski M, Johnson M, et al. Fastnerf: High-fidelity neural rendering at 200fps[OL]. ArXiv preprint arXiv:2103.10380, 2021.
- [7] Guo M, Fathi A, Jiajun Wu, and Funkhouser T. Object-centric Neural Scene Rendering[OL]. arXiv preprint arXiv:2012.08503, 2020.
- [8] Hedman P, Srinivasan P P, Mildenhall B, et al. Baking Neural Radiance Fields for RealTime View Synthesis[OL]. ArXiv preprint arXiv:2103.14645, 2021.
- [9] Liu L, Gu J, Lin K Z, et al. Neural sparse voxel fields[OL]. ArXiv preprint arXiv:2007.11571, 2020.
- [10] Martin-Brualla R, Radwan N, Sajjadi M S M, et al. NeRF in the Wild: Neural Radiance Fields for Unconstrained Photo Collections[OL]. arXiv preprint arXiv:2008.02268, 2020.
- [11] Bi S, Xu Z, Srinivasan P, et al. Neural Reflectance Fields for Appearance Acquisition[OL]. arXiv preprint arXiv:2008.03824, 2020.
- [12] Mildenhall B, Srinivasan P P, Tancik M, et al. Nerf: Representing scenes as neural radiance fields for view synthesis[C]// *European Conference on Computer Vision*. 2020:405421.
- [13] Chen A, Xu Z, Zhao F, et al. MVSNerf: Fast Generalizable Radiance Field Reconstruction from Multi-View Stereo[OL]. ArXiv preprint arXiv:2103.15595, 2021.
- [14] Neff T, Stadlbauer P, Parger M, et al. DONerf: Towards realtime rendering of neural radiance fields using depth oracle networks[OL]. ArXiv eprints, 2021:arXiv2103.
- [15] Niemeyer M, Geiger A. GIRAFFE: Representing Scenes as Compositional Generative Neural Feature Fields[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021: 11453-11464.
- [16] Boss M, Braun R, Jampani V, et al. NeRD: Neural Reflectance Decomposition from Image Collections[OL]. arXiv preprint arXiv:2012.03918, 2020.
- [17] Park K, Sinha U, Barron J T, et al. Deformable neural radiance fields[OL]. ArXiv preprint arXiv:2011.12948, 2020.
- [18] Pumarola A, Corona E, Pons Moll G, et al. Dnerf: Neural radiance fields for dynamic scenes[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021:1031810327.
- [19] Schwarz K, Liao Y, Niemeyer M, et al. Graf: Generative radiance fields for 3d aware image synthesis[OL]. ArXiv preprint arXiv:2007.02442, 2020.
- [20] Sitzmann V, Chan E R, Tucker R, et al. MetaSDF: Meta-learning signed distance functions[OL]. ArXiv preprint arXiv:2006.09662, 2020.
- [21] Srinivasan P P, Deng B, Zhang X, et al. NeRV: Neural Reflectance and Visibility Fields for Relighting and View Synthesis[C]// *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2021:74957504.
- [22] Gafni G, Thies J, Zollhofer M, et al. Dynamic neural radiance fields for monocular 4d facial

- avatar reconstruction[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:86498658.
- [23]Tancik M, Mildenhall B, Wang T, et al. Learned initializations for optimizing coordinatebased neural representations[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:28462855.
- [24]Tretschk E, Tewari A, Golyanik V, et al. NonRigid Neural Radiance Fields: Reconstruction and Novel View Synthesis of a Dynamic Scene From Monocular Video[J]. ArXiv preprint arXiv:2012.12247, 2020.
- [25]Wang Q, Wang Z, Genova K, et al. Ibrnet: Learning multiview imagebased rendering[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:46904699.
- [26]Xian W, Huang J B, Kopf J, et al. Spacetime neural irradiance fields for freeviewpoint video[C]// ProceedingsoftheIEEE/CVFFconferenceonComputerVisi onandPatternRecognition. 2021:94219431.
- [27]Chabra R, Lenssen J E, Ilg E, et al. Deep local shapes: Learning local sdf priors for detailed 3d reconstruction[C]//European Conference on Computer Vision. 2020:608625.
- [28]Yu A, Ye V, Tancik M, et al. Pixelnerf: Neural radiance fields from one or few images[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:45784587.
- [29]Yu A, Li R, Tancik M, et al. Plenotrees for realtime rendering of neural radiance fields[J]. ArXiv preprint arXiv:2103.14024, 2021.
- [30]Yuan W, Lv Z, Schmidt T, et al. STaR: Selfsupervised Tracking and Reconstruction of Rigid Objects in Motion with Neural Rendering[C]//Proceedings of the IEEE/ CVF Conference on Computer Vision and Pattern Recognition. 2021:1314413152.
- [31]Li Z, Niklaus S, Snavely N, et al. Neural scene flow fields for space time view synthesis of dynamic scenes[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021:6498 6508.
- [32]Zhang X, Srinivasan P P, Deng B, et al. NeRFactor: Neural Factorization of Shape and Reflectance Under an Unknown Illumination[J]. arXiv preprint arXiv:2106.01970, 2021.
- [33]Ost J, Mannan F, Thurey N, et al. Neural Scene Graphs for Dynamic Scenes[J]. arXiv preprint arXiv:2011.10379, 2020.
- [34]Niemeyer M, Mescheder L, et al. Differentiable Volumetric Rendering: Learning Implicit 3D Representations without 3D Supervision [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020.
- [35]Wizadwongsa S, Phongthawee P, et al. NeX: Real-Time View Synthesis With Neural Basis Expansion [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2021.
- [36]Niemeyer M, Mescheder L, et al. NeX: Occupancy Flow: 4D Reconstruction by Learning Particle Dynamics[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019.
- [37]Bo i A, Palafox P, et al. Neural Deformation Graphs for Globally-consistent Non-rigid Reconstruction[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
- [38]Lombardi S, Simon T, et al. Neural Volumes: Learning Dynamic Renderable Volumes from Images[J]. ACM Transactions on Graphics, Volume 38, Issue 4. 2019.
- [39]Peng S, Zhang Y, et al. Neural Body: Implicit Neural Representations with Structured Latent Codes for Novel View Synthesis of Dynamic Humans. [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021.
- [40]Blanz V, Vetter T. A morphable model for the synthesis of 3D faces[C]. international conference on computer graphics and interactive techniques, 1999: 187-194.
- [41]Guy Gafni, Justus Thies, Michael Zollhöfer, Matthias Nießner. Dynamic Neural Radiance Fields for Monocular 4D Facial Avatar Reconstruction. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021. 8649-8658.
- [42]Huiwen Luo, Koki Nagano, Han-Wei Kung, Qingguo Xu, Zejian Wang, Lingyu Wei, Liwen Hu, Hao Li. Normalized Avatar Synthesis Using StyleGAN and Perceptual Refinement. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [43]Tarun Yenamandra, Ayush Tewari, Florian Bernard, Hans-Peter Seidel, Mohamed Elgharib, Daniel Cremers, and Christian Theobalt. i3DMM: Deep Implicit 3D Morphable Model of Human Heads. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [44]Shugao Ma, Tomas Simon, Jason Saragih, Dawei Wang, Yuecheng Li, Fernando De La Torre, Yaser Sheikh. Pixel Codec Avatars. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [45]C. Zimmermann and T. Brox, Learning to Estimate 3D Hand Pose from Single RGB Images. ICCV 2017.
- [46]Mueller Franziska, Bernard Florian, Sotnychenko Oleksandr et al. GANerated Hands for Real-Time 3D

- Hand Tracking from Monocular RGB, CVPR 2018.
- [47]Javier Romero and Dimitrios Tzionas and Michael J. Black. Embodied Hands: Modeling and Capturing Hands and Bodies Together. ACM Transactions on Graphics, (Proc. SIGGRAPH Asia) 2017
- [48]Xiong Zhang, Qiang Li, Hong Mo et al. End-to-end Hand Mesh Recovery from a Monocular RGB Image. ICCV 2019
- [49]Adnane Boukhayma, Rodrigo de Bem , Philip H.S. Torr. 3D Hand Shape and Pose from Images in the Wild. CVPR 2019
- [50]Liuhaog Ge, Zhou Ren, Yuncheng Li. 3D Hand Shape and Pose Estimation from a Single RGB Image. CVPR 2019
- [51]Korrawe Karunratanakul, Jinlong Yang, Yan Zhang. Grasping Field: Learning Implicit Representations for Human Grasps. 3DV 2020
- [52]Newcombe Richard A., Dieter Fox, and Steven M. Seitz. DynamicFusion: Reconstruction and tracking of non-rigid scenes in real-time. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2015.
- [53]Mingsong Dou, Sameh Khamis, Yury Degtyarev, Philip Davidson, Sean Ryan Fanello, Adarsh Kowdle, Sergio Orts Escolano, Christoph Rhemann, David Kim, Jonathan Taylor, Pushmeet Kohli, Vladimir Tankovich, and Shahram Izadi. 2016. Fusion4D: real-time performance capture of challenging scenes. ACM Trans. Graph. 35, 4, Article 114.
- [54]Minye Wu, Yuehao Wang, Qiang Hu and Jingyi Yu. Multi-View Neural Human Rendering. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
- [55]Anqi Pang, Xin, Chen, Haimin, Luo, Minye, Wu, Jingyi, Yu and Lan, Xu. Few-shot Neural Human Performance Rendering from Sparse RGBD Videos. Proceedings of the 30th International Joint Conference on Artificial Intelligence (IJCAI-21).
- [56]Shunsuke Saito, Zeng Huang, Ryota Natsume, Shigeo Morishima, Hao Li, Angjoo Kanazawa. PIFu: Pixel-Aligned Implicit Function for High-Resolution Clothed Human Digitization. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.
- [57]Zerong Zheng, Tao Yu, Yebin Liu, Qionghai Dai. PaMIR: Parametric Model-Conditioned Implicit Representation for Image-based Human Reconstruction. IEEE Transactions on Pattern Analysis and Machine Intelligence, doi: 10.1109/TPAMI.2021.3050505.
- [58]Xin Suo, Yuheng Jiang, Pei Lin, Yingliang Zhang, Minye Wu, Kaiwen Guo and Lan Xu. NeuralHumanFVV: Real-Time Neural Volumetric Human Performance Rendering using RGB Cameras. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [59]Marc Habermann, Weipeng Xu, Michael Zollhoefer, Gerard Pons-Moll and Christian Theobalt. NeuralHumanFVV: Real-Time Neural Volumetric Human Performance Rendering using RGB Cameras. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
- [60]Marc Habermann, Lingjie Liu, Weipeng Xu, Michael Zollhoefer, Gerard Pons-Moll and Christian Theobalt. Real-time Deep Dynamic Characters. ACM Transactions on Graphics, (Proc. SIGGRAPH) 2021.
- [61]Lingjie Liu, Marc Habermann, Viktor Rudnev, Kripasindhu Sarkar, Jiatao Gu, Christian Theobalt. Neural Actor: Neural Free-view Synthesis of Human Actors with Pose Control. ACM Transactions on Graphics, (Proc. SIGGRAPH Asia) 2021.
- [62]Jiakai Zhang, Xinhang Liu, Xinyi Ye, Fuqiang Zhao, Yanshun Zhang, Minye Wu, Yingliang Zhang, Lan Xu and Jingyi Yu. Editable free-viewpoint video using a layered neural representation. ACM Transactions on Graphics, (Proc. SIGGRAPH) 2021.