DOI:10.19651/j.cnki.emt.2314535

基于门控循环深度范围预测网络的多视角重建*

高 宇 朱立忠 刘韵婷 刘晓玉
(沈阳理工大学自动化与电气工程学院 沈阳 110159)

摘 要:针对三维重建技术难以处理高分辨率图像、重建后的点云图精度低、边界模糊的问题,本文提出基于门控循环单元的多阶段多尺度动态深度范围预测网络模型。首先,利用曲率引导的动态尺度卷积网络作为特征提取模块,通过计算图像上多个尺度的表面法曲率,得到图像最优像素的特征信息;然后,将精细的特征信息与一种新的深度范围估计模块相结合,动态估计下阶段的深度范围假设,从而更好的合并邻域像素的信息,实现参考图像和源图像之间的精确匹配。本文网络与其他 10 多种方法进行了比较,在 DTU 数据集上,整体性能比第 2 的网络提高 2.2%。在 Tank&Temple 数据集上,Lighthouse、M60 和 Panther 等场景的重建表现都有大幅提升。同时,本文进行了对比和消融实验,实验结果证明本文提出的动态深度预测网络,减小内存消耗的同时,显著提高了重建后点云图的精度和完整度。

关键词:多视角;三维重建;深度估计;点云
中图分类号:TP391 **文献标识码**:A **国家标准学科分类代码**:510

Multi-view stereo reconstruction based on gated recurrent deep range prediction network

Gao Yu Zhu Lizhong Liu Yunting Liu Xiaoyu
(School of Automation and Electrical Engineering, Shenyang University of Technology, Shenyang 110159, China)

Abstract: Aiming at the problems that 3D reconstruction techniques are difficult to deal with high-resolution images, and the reconstructed point cloud maps have low accuracy and fuzzy boundaries, this paper proposes a multi-stage multi-scale dynamic depth range prediction network model based on gated recurrent units. First, a curvature-guided dynamic scale convolutional network is used as a feature extraction module to obtain the feature information of the optimal pixels of the image by calculating the surface normal curvature at multiple scales on the image; then, the fine feature information is combined with a new depth range estimation module to dynamically estimate the depth range assumptions of the next stage, so as to better merge the information of neighboring pixels, and to achieve an accurate matching between the reference image and the source image. The network in this paper is compared with more than 10 other methods, and on the DTU dataset, the overall performance is improved by 2.2% over the network in 2nd. On the Tank&Temple dataset, the reconstruction performance of the Lighthouse, M60 and Panther scenes are substantially improved. Meanwhile, comparison and ablation experiments are conducted in this paper, and the experimental results demonstrate that the dynamic depth prediction network proposed in this paper significantly improves the accuracy and completeness of the reconstructed point cloud maps while reducing the memory consumption.

Keywords: multi view; 3D reconstruction; depth inference; point cloud

0 引 言

传统的 MVS 方法[1]大多采用手工设计的特征描述符来确定不同图像中像素之间的对应关系,并且应用工程正则化来恢复三维点云,这使得在弱纹理和镜面反射等光滑区域的重建效果较差。随着深度学习方法[2-5]的日趋成熟,基于卷积神经网络的深度估计方法在各类 MVS 算法的基准测试上表现出了明显优势。

收稿日期:2023-09-05
* 基金项目:辽宁省自然科学基金(2022-KF-14-02)、国家重点研发计划(2017YFC0821001-2)、辽宁省教育厅面上项目(LJKMZ20220617)资助

基于深度学习的多视角三维重建[6-8]用三维卷积神经网络对 3D 代价体(Cost Volume)进行正则化,最终从概率体中回归深度图。虽然这种方法在基准测试上取得了令人印象深刻的表现,但它并不能很好地扩展到高分辨率场景,运行速度和内存分配同样难以满足大多数应用场景。为了提高效率,近期基于深度学习的方法主要可分为两类:递归方法 R-MVSNet[9]和多阶段方法 CasMVSNet[10]。

递归方法通常使用循环神经网络门控循环单元(gated recurrent unit, GRU)进行代价体正则化来降低内存消耗,但增加了运行时长。多阶段方法采用级联结构,由粗到细的估计深度图,这种方法可以平衡内存和运行时间,但会减少深度搜索范围。例如 CasMVSNet 和 CVP-MVSNet[11]网络以最优深度分辨率构造代价体,但在精细阶段的深度预测范围较窄,使重建完整性受到严重影响。为了控制运行成本的同时提高重建质量,UCSNet[12]和 DDR-Net[13]提出了修改精细阶段代价体的想法。UCSNet 利用深度分布的方差来确定下一阶段的深度范围假设,并相应地调整深度分辨率,但是 UCSNet 的有效性在很大程度上依赖于粗阶段深度预测的质量,这种静态深度假设方法不适合于高精度和大场景下的的深度估计。DDR-Net 解决了粗阶段深度预测质量差这一问题,提出了一种动态深度范围估计模块,能够从附近的像素中收集深度不确定性信息,从而扩大深度搜索范,但是 DDR-Net 对于图像特征的采集和对深度的预测方面缺少必要的联系,鲁棒性较差。另外,受内存消耗和运行时间的限制,MVS 网络经常只能在低分辨率图像上进行训练。因此,现有的固定尺度特征提取方法,在深度预测中无法有效推广到高分辨率场景。

针对当前 MVS 在多阶段方法上存在的上述一系列问题,本文提出 GRU-MVSNet 网络,使用深度范围估计模块(GREM)和上一阶段的概率分布生成细化的深度图,保证了下一阶段的深度假设范围内所覆盖的每个像素都对应图像正确的真实深度值,同时改善了 GPU 内存限制带来的低分辨率训练问题,从而实现高质量的三维重建任务。

1　相关工作

传统的 MVS 方法通常遵循由摄像机的内在和外在参数以及从 SFM 获得的稀疏点云作为输入,将从多视角图像中提取的特征点合并,并通过匹配、扩展和过滤生成密集的三维点云,但是内存消耗较大一直是传统 MVS 需要解决的问题。

基于深度学习的多视角三维重建方法在取代传统三维重建的每一个步骤方面都显示出了巨大的潜力。2018 年,Yao 等[14-17]提出了 MVSNet,是目前广泛使用的深度学习 MVS 框架之一。随着研究的深入,Yao 等又提出了 R-MVSNet 网络,采用 2D GRU 递归网络作为正则化模块,但是增加了运行时间。Fast-MVSNet[18]提出了一种新的稀疏到密集、粗到细的框架,用于在 MVS 中产生快速精确的深度估计。Point-MVSNet[19]提出了一种基于点的深度图细化网络。CasMVSNet 和 CVP-MVSNet 将从粗到细的策略应用到 MVS 重建中,构造了一个图像特征金字塔,以最粗的分辨率构建整个深度范围的代价体;然后,根据之前的深度预测方法计算出一个缩小的采样范围。由粗到细的体系结构降低了网络对内存的消耗,从而支持更深层次的主干网络和更高分辨率的输出。

本文的方法受上述工作的启发,整体网络结构采用从粗到细的策略,使用曲率卷积单元进行多尺度动态特征提取,同时采用 GREM 模块进行深度范围估计,以保证每个像素的真实值覆盖在下一阶段的范围假设中,提高网络对目标物体的重建质量。

2　MVS 算法

本文提出的 GRU-MVSNet 网络采用级联结构框架,由三个级联阶段组成,每个阶段通过 3 个步骤来估算深度,分别为特征提取、代价体正则化和动态深度范围估计。网络首先通过在特定的尺度上参考和源特征图之间的映射关系来计算三维代价体,然后用 3DCNN 进行正则化,最后逐步确定近似的深度范围,以更好地适应高分辨率场景的重建任务。

2.1　曲率特征提取网络

传统的 MVS 方法中常采用多个降采样卷积层或 2D U-Net 进行特征提取,但是增加网络的计算量,难以实现高分辨率场景的重建。为了避免该问题,并且在特征提取阶段获取更重要的全局信息,本文提出一个由曲率引导的动态特征网络,通过扩展搜索尺度空间,为每个像素选择最优尺度来减少匹配的模糊度,从而学习高鲁棒性的特征。为了降低网络的计算复杂度,在曲率卷积结构中使用较少数量的候选尺度,本文使用 3 个候选尺度。

动态特征网络采用多个曲率引导的卷积结构组成,给定一组不同大小的卷积核 $\{C_1,C_2,\cdots,C_K\}$,对应 r 个候选尺度 $\{\sigma_1,\sigma_2,\cdots,\sigma_K\}$,曲率卷积结构如图 1 所示。

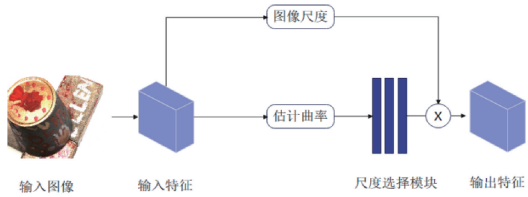


图 1　曲率卷积结构

首先,在 r 个候选尺度上估计近似正常的曲率;然后进行尺度选择步骤,从估计的曲率中输出最优尺度。网络中采用法曲率估计图形曲面沿特定方向上特定点的弯曲程度,法曲率 Δz_{pq} 计算如下:

$$v(x,\omega)=\frac{u^2I_{xx}(x,\sigma)+2uvI_{xy}(x,\sigma)+v^2I_{yy}(x,\sigma)}{\sqrt{1+I_x^2(x,\sigma)+I_y^2(x,\sigma)}\,(1+(uI_x(x,\sigma)+vI_y(x,\sigma))^2)} \tag{1}$$

$I(x,\sigma)=I_x*G(x,\sigma)$ 表示图像尺度 σ 中像素 x 的图像强度，由图像 I 与高斯核 $G(x,\sigma)$ 卷积决定。I_x、I_y、I_{xx}、I_{xy} 和 I_{yy} 的导数由原始图像 I 与高斯核 $G(x,\sigma)$ 的导数卷积得到，如下：

$$\frac{\partial^{i+j}}{\partial x^i\partial y^j}I(X,\sigma)=I(X)*\frac{\partial^{i+j}}{\partial x^i\partial y^j}G(X,\sigma) \tag{2}$$

由于 I_x、I_y、I_{xx}、I_{xy} 和 I_{yy} 导数计算需要进行 5 次卷积运算，为了降低计算成本，同时保持高维特征输入，本文使用可学习的核。因此，对于每个尺度的 σ，引入了 3 个可学习的卷积内核 K_σ^{xx}、K_σ^{xy} 和 K_σ^{yy} 来分别代替 G_{xx}、G_{xy} 和 G_{yy}。这些核适用于输入特征，用来计算图像表面的二阶导数。F^{in} 表示输入特征，正态曲率由式(3)得到：

$$curv_\sigma(x,\omega)=\omega\begin{bmatrix}F^{in}*K_\sigma^{xx} & F^{in}*K_\sigma^{xy}\\ F^{in}*K_\sigma^{xy} & F^{in}*K_\sigma^{yy}\end{bmatrix} \tag{3}$$

特征输出 F^{out} 由式(4)得到。其中，$\{C_1,C_2,\cdots,C_K\}$ 表示特征输入 F^{in} 中的 K 个候，$*$ 是卷积运算符。

$$F^{out}=\omega_1(F^{in}*C_1)+\omega_2(F^{in}*C_2)+\cdots\omega_K(F^{in}*C_K) \tag{4}$$

2.2　构造代价体

3D 代价体的构建是基于深度学习三维重建方法的关键步骤。给定一个深度采样假设 $d_j(j=1\cdots D)$ 及所有的相机参数 $\{K_i,R_i\}$，通过可微的单映性变换将提取的特征图 F^{out} 从源图像映射到参考图像上，在多个尺度上构建多个代价体。给定每个视图的相机内参和外参矩阵，参考视图深度 d_j 映射矩阵表示为：

$$H_i(d)=K_iR_i(I-\frac{(t_1-t_i)n_1^{\mathrm{T}}}{d})R_1^{\mathrm{T}}K_1^{-1} \tag{5}$$

其中，$H_i(d)$ 指第 i 个视图的特征图与深度 d 处的参考特征图之间的映射关系。此外，R_i 和 t_i 分别表示第 i 个视图的旋转和平移参数。n_1 为参考相机的主轴。

第 1 阶段，从一个预定义的深度区间中均匀采样来构造体积模型，然后采用基于方差的度量方式生成单个代价体。在第 2 阶段和第 3 阶段，其深度假设根据先前深度预测的像素级进行再一次的不确定性估计，生成具有空间变化的深度值。本文将第 $K+1$ 个阶段的同源函数设为：

$$H_i(d_k^m+\Delta_{k+1}^m)=K_iR_i(I-\frac{(t_1-t_i)n_1^{\mathrm{T}}}{d_k^m+\Delta_{k+1}^m})R_1^{\mathrm{T}}K_1^{-1} \tag{6}$$

其中，d_K^m 为第 K 阶段第 m 个像素的预测深度，Δ_{K+1}^m 为 $K+1$ 阶段要学习的第 m 个像素的剩余深度。

单映性变换作为连接二维特征提取和 3D 正则化网络的核心步骤，实现了深度预测的端到端训练模式。在构造成本体积后，应用 3DCNN 对代价体进行正则化，3DCNN 的末端应用 SOFTMAX 环节来预测每个像素的深度概率。

每个代价体由多个平面组成，使用 D_K 表示 K 阶段的平面数。$P_{K,j}$ 表示像素深度的概率分布，映射的 $P_{K,j}$ 组成概率体，表示第 K 个阶段的第 j 个平面的深度假设。$L_{K,j}$ 表示像素 x 处的深度为 $L_{K,j}(x)$ 的概率。通过加权和重建了第 K 阶段的深度图 $\hat{L}_K(x)$：

$$\hat{L}_K(x)=\sum_{j=1}^{D_k}L_{K,j}(x)\cdot P_{K,j}(x) \tag{7}$$

本文的 3 个阶段使用相同的网络架构，如图 2 所示，通过不共享权重的方式，每个阶段都可以学习不同尺度的信息并各自处理。

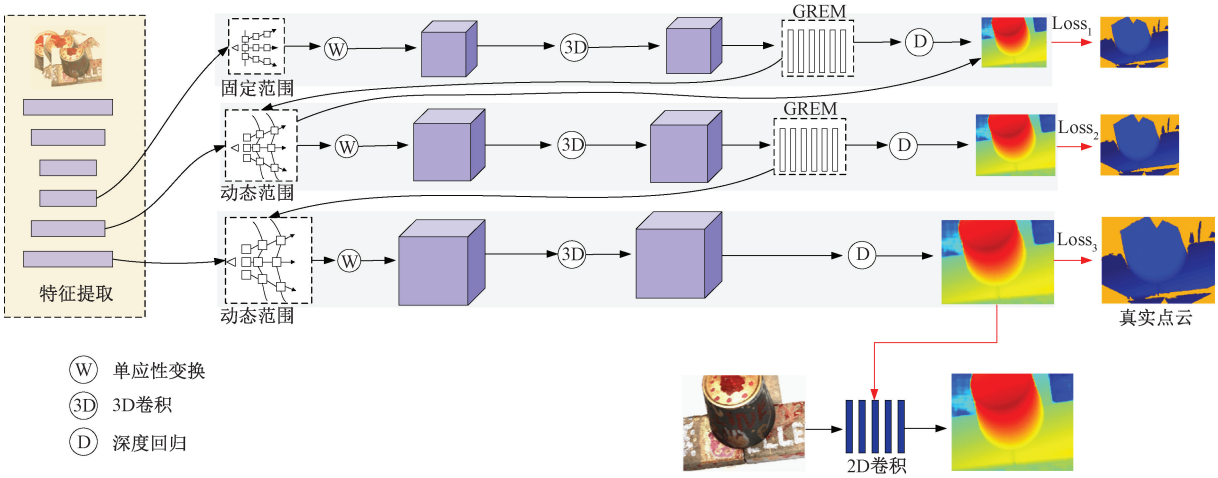


图 2　模型结构

2.3　动态深度范围估计

GRU-MVSNet 网络的关键是逐步细分局部空间，通过细化深度预测的方式来提高重建的精度。本文基于前一阶段预测深度的不确定性提出使用一种新的范围估计模块(GREM)。可以利用先前的概率体信息，自适应地估计动态深度范围，再利用动态特征提取模块捕获附近像素的特性和上下文信息，GREM 结构如图 3 所示。

GREM 结构由不同尺寸的卷积层和 GRU 层组成。其中，GRU 层作为预处理阶段，合并收集到的特征，在特征达到最低分辨率后，通过两个不同尺寸反卷积层进行解码操

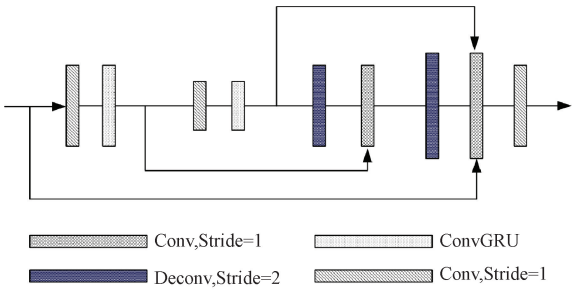


图 3　GREM 结构

作。GREM 模块可以自适应地估计动态深度范围,最终利用 SOFTMAX 层将不确定性映射的输出值限制为[0,1]。

网络的输出是一个不确定性的特征图 $C=\{C(x)\}_{x\in I}$,$C(x)$ 指的是图像像素的不确定性深度分布。给定由 GREM 获得的不确定度值 $C(x)$ 和像素 x 处之前的深度预测 $L(x)$,本文由式(8)确定下一阶段的深度范围 $D(x)$:

$$D(x)=[L(x)-\lambda C(x),L(x)+\lambda C(x)] \tag{8}$$

其中,λ 是一个决定置信区间有多大的超参数。利用 GREM 模块对上一阶段概率体积的学习,考虑了附近的像素点,以更高的置信度学习深度范围假设,并对每个阶段的概率预测进行调整,以获得更好的优化间隔,因此网络具有更好的空间分区能力。

本文网络使用 L1 损失函数。最终的损失函数表示为:

$$Loss=\sum_{k-1}^{3}a_k Loss_k+\sum_{k=1}^{2}b_k Loss_k^{refined} \tag{9}$$

式中:$Loss_K$、$Loss_K^{refined}$、α_K 和 β_K 分别表示第 K 阶段的损失和细化损失及其对应的权重。

3　实验与分析

3.1　实验数据集介绍

本文对改进的网络分别在室内 DTU[20] 数据集和室外 Tank&Temple[21] 数据集上进行评估。DTU 数据集包含 124 个不同的室内场景,每个场景包含 49 或 64 张图像,其视点和照明条件都是精心设计的。该数据集提供了摄像机位姿参数和 Ground Truth 点云。图像分辨率为 1 600×1 184,每个场景的深度范围在 425～935 mm 之间。通过计算点云的平均精度,平均完整度和总体得分评估网络模型。

Tank&Temple 数据集包含两个场景集,即中间和高级,本文使用中间场景集来进行评估,使用 f1 分数作为评估标准。有 8 个不同的户外场景,即 Family、Francis、Horse、Lighthouse、M60、Panther、Playground 和 Train。Tank&Temple 数据集建场景非常大,在物体的表面有很多反射和遮挡,非常具有挑战性。

3.2　实验设置

本文的网络在 DTU 数据集上进行训练,使用降采样图像和 Ground Truth 来优化训练过程。输入图像的分辨率设置为 640×512,3 个阶段的深度平面数分别设置为 48、32 和 8 个,损失权重分别设置为 $\alpha_1=0.5$、$\alpha_2=1.5$ 和 $\alpha_3=2.5$。细化损失的权重分别设置为 $\beta_1=2.0$ 和 $\beta_2=0.5$。

本文使用 PyTorch 搭建框架,并使用 Adam 优化器来训练模型。整个网络在 1 个 NVIDIA Telsa P100 显卡上进行了 16 次的训练,批次大小为 16。初始学习率设置为 0.001,并在之后的阶段迭代减半。

3.3　实验结果分析

首先在 DTU 测试数据集上与其他 MVS 评估了本文提出的方法。定量结果如表 1 所示,与原方法 DDR-Net 相比,本文提出的 GRU-MVSNet 可以显著提高点云的准确性和完整性。

表 1　DTU 数据集评估表(越小越好)

方法	准确性/mm	完整性/mm	总体评分/mm
COLMAP[22]	0.400	0.664	0.532
MVSNet	0.396	0.527	0.462
CasMVSNet	0.346	0.351	0.348
CVP-MVSNet	0.296	0.406	0.351
Point-MVSNet	0.342	0.411	0.376
Vis-MVSNet[23]	0.369	0.361	0.365
UniMVSNet[24]	0.352	0.278	0.315
UCSNet	0.339	0.349	0.344
DDR-Net	0.339	0.320	0.329
本文	0.352	**0.272**	**0.312**

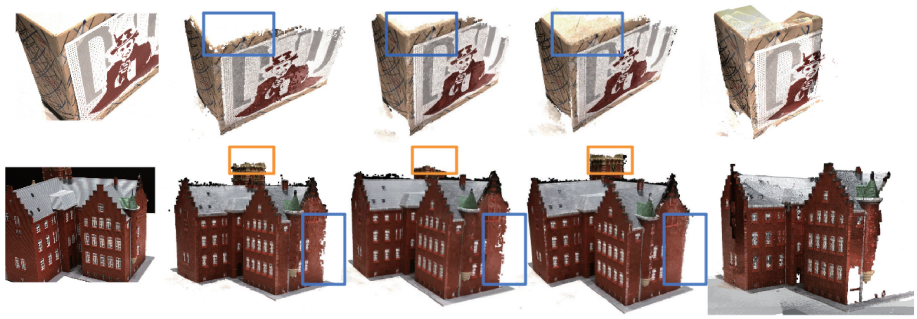
网络输出的深度图如图 4 所示,本文选择可视化 DTU 数据集中的 scan15 和 scan25 场景在不同的网络上进行了比较,可以看出 GRU-MVSNet 生成了更精细的深度图。

为了更直观的体现重建后的效果,本文还将重建的点云结果与同样使用深度范围假设的网络 UCSNet 和 DDR-Net 进行了比较。对比结果如图 5 所示,本文的方法不论是在整体点云图的精度上还是在黄色框内部分的完整性上都展现出了优于其他方法的表现,本文的方法可以重建出更完整、更准确的点云图。

本文使用 DTU 训练集生成的模型测试 Tank&Temple 数据集中的 intermediate 数据集来评估 GRU-MVSNet 网络的泛化能力。实验使用视角数 $N=5$,图像分辨率 $W\times H=960\times 560$,在没有任何微调的 Tank&Temple 数据集上进行评估,如表 2 所示。由于 Tank&Temple 数据集包含了许多大型的户外场景,需要消耗大量的 GPU 内存,所以实验降低了 Family、Francis 和 Horse 的输入图像分辨率,因此会导致整体评估结果有所降低。

(a) 原始图像　(b) CasMVSNet　(c) UniMVSNet　(d) DDR-Net　(e) 本文网络

图 4　DTU 数据集深度图可视化结果



(a) 原始图像　(b) CasMVSNet　(c) DDR-Net　(d) 本文网络　(e) Ground Truth

图 5　DTU 数据集点云图可视化结果

表 2　Tank&Temple 数据集评估表(越大越好)

方法	Mean	Family	Francis	Horse	Lighthouse	M60	Panther	Playground	Train
COLMAP	42.14	50.41	22.25	25.63	44.83	44.83	46.97	48.53	42.04
MVSNet	43.48	55.99	28.55	25.07	50.79	53.96	50.86	47.90	34.69
R-MVSNet	48.40	69.96	46.65	32.59	42.95	51.88	48.80	57.43	47.54
CVP-MVSNet	54.03	76.50	47.74	36.34	55.12	57.28	54.28	57.43	47.54
Point-MVSNet	48.27	61.79	41.15	34.20	50.79	51.97	50.85	52.38	43.06
UniMVSNet	**64.36**	**81.20**	**66.43**	**53.11**	63.46	66.09	64.84	**62.23**	**57.53**
CasMVSNet	56.42	76.36	58.45	46.20	55.53	56.11	54.02	58.17	46.56
UCSNet	54.83	76.09	53.16	43.03	54.00	55.60	51.49	57.38	47.89
DDR-Net	54.91	76.18	53.36	43.43	55.20	55.57	52.28	56.04	47.17
本文	61.66	80.66	63.15	50.60	**64.03**	**65.55**	64.80	58.50	46.00

本文提出的 GRU-MVSNet 在 Tank&Temple 数据集上部分场景重建的点云结果如图 6 所示，表明本文的方法无论是在室内环境还是室外环境都有良好的表现，显示出了更强的泛化性和鲁棒性。

本文对 DTU 数据集中测试集的所有场景进行深度估计进而融合成点云图，如图 7 所示。

本文将该方法的一些综合性能与其他几种基于深度学习的多视角三维重建方法在 DTU 数据集上进行了比较，包括运行时间和内存消耗，输入图像分辨率为 $W\times H=1\ 600\times 1\ 184$，如表 3 显示了性能比较。

对于相同大小的输入图像，本文的方法在输入高分辨率图像的情况下，消耗了较小的内存占用量和合理的运行

图 6　Tank&Temple 数据集可视化结果

图 7　DTU 测试集点云图

表 3　DTU 数据集内存和时间重建质量

方法	输入尺寸	深度图尺寸	总体评分/mm	GPU 内存/MB	运行时间/s
COLMAP	1 600×1 200	400×288	0.532	18 600	8.50
MVSNet	1 600×1 184	400×288	0.462	22 511	2.76
R-MVSNet	1 600×1 184	400×288	0.417	6 915	5.09
Point-MVSNet	1 600×1 184	640×480	0.391	8 731	3.35
CVP-MVSNet	1 600×1 184	800×576	0.351	8 795	1.72
CasMVSNet	1 600×1 184	1 600×1 184	0.348	10 153	0.89
UCSNet	1 600×1 184	1 600×1 184	0.344	7 252	0.87
DDR-Net	1 600×1 184	1 600×1 184	0.329	7 345	0.88
本文	1 600×1 184	1 600×1 184	0.312	7 061	0.79

时间,良好的重建效果,充分体现了该深度估计方法的实用性。

由于多视图图像将为深度推断任务提供更多的信息,本文选择 $N=3$、5 和 7 的视图数进行测试,评估结果如表 4 所示。

表 4　不同视角数评估表

视图数	准确性/mm	完整性/mm	总体评分/mm
3	0.353	0.291	0.322
5	**0.342**	**0.272**	**0.312**
7	0.353	0.280	0.316

当 $N=5$ 时,重建结果最好,说明视图数量增加可以有效解决视角遮挡带来的问题。实验证明随着输入视图的增加,重构质量的准确性(Acc.)和完整性(Comp.)都有所提高。

3.4　消融实验

为了验证本文所提方法的优越性,本文消融实验在 DTU 数据集上对重建后的点云图进行评估,评价指标为点云图的精度和完整度求和均值,实验结果如表 5 所示。在特征提取阶段使用曲率卷积网络和普通卷积网络的前提条件下,使用本文提出的深度预测模块 GREM 比采用 REM 模块可以提高模型的性能。当不使用 REM 和 GREM 时,网络为 MVSNET 的[425,935]mm 均匀取样深度采样。实验结果证明,GREM 提升重建效果显著,并且当网络同时使用曲率卷积和 GREM 时,重建效果最佳。

表 5　DTU 数据集上深度预测模块的消融实验

Conv	Curconv	REM	GREM	总体评分/mm
√				0.462
	√	√		0.336
√			√	0.329
	√		√	**0.312**

4　结　　论

本文提出了一种使用多尺度特征提取网络的动态多视角深度估计网络框 GRU-MVSNet,该框架的核心思想是利用由曲率引导的特征提取网络学习最优像素的尺度来学习多尺度特征,该特征结合 REM 模块动态推断深度假设,可以进行更大范围的深度预测,达到了较好的重建质量,在 DTU 数据集和 Tank&Temple 数据集上效果显著。

参考文献

[1]　GALLIAI S, LASINGER K, SCHINDLER K. Massively parallel multiview stereopsis by suface normal diffusion [C]. Proceedings of the IEEE International Conference on Computer Vision, Boston USA, 2015: 873-881.

[2]　MI Q, GAO T. 3D reconstruction based on the depth image: A review [C]. International Conference on Innovative Mobile and Internet Services in Ubiquitous Computing. Cham: Springer International Publishing, 2022: 172-183.

[3] 霍智勇,乔璐.基于结构化损失的单目深度估计算法研究[J]. 电子科技大学学报, 2021, 48(6): 1001-0548.
[4] 李爱军.基于卷积神经网络和 NBV 的三维重建方法[J].电子测量技术,2021,44(8):70-75.
[5] LIU Y, YAN X, WANG N, et al. A 3D reconstruction method of image sequence based on deep learning[C]. Journal of Physics: Conference Series. IOP Publishing, 2020, 1550(3): 032051.
[6] YU J, YIN W, HU Z, et al. 3D reconstruction for multi-view objects [J]. Computers and Electrical Engineering, 2023, 106: 108567.
[7] YAN X, HU S, MAO Y, et al. Deep multi-view learning methods: A review[J]. Neurocomputing, 2021, 448: 106-129.
[8] CHEN H, CHEN W, GAO T. Ground 3D object reconstruction based on multi-view 3D occupancy network using satellite remote sensing image[C]. 2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS, IEEE, 2021: 4826-4829.
[9] YAO Y, LUO Z X, LI S W. Recurrent mvsnet for high-resolution multi-view stereo depth inference[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, USA, 2019: 5525-5534.
[10] GU X D, FAN Z W, ZHU S Y. Cascade cost volume for high-resolution multi-view stereo and stereo matching[C]. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle USA, 2020:2495-2504.
[11] YANG J Y, MAO W, ALVAREZ J M. Cost volume pyramid based depth inference for multi-view stereo[C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle USA, 2020: 4877-4886.
[12] CHENG S, XU Z X, ZHU S L. Deep stereo using adaptive thin volume representation with uncertainty awareness[C]. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle USA, 2020:2524-2534.
[13] YI P Y, TANG S K, YAO J. Learning multi-stage multi-view stereo with dynamic depth range[EB/OL]. https://doi.org/10.48550/arXiv.2103.14275, 2021.
[14] 薛俊诗,易辉,吴止锾,等.一种基于场景图分割的混合式多视图三维重建方法[J].自动化学报,2020,46(4):782-795.
[15] CHEN R, HAN S, XU J, et al. Point-based multi-view stereo network[C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 1538-1547.
[16] 刘万军,王俊恺,曲海成.多尺度代价体信息共享的多视角立体重建网络[J].中国图象图形学报,2022,27(11):3331-3342.
[17] 王思启,张家强,李丽圆,等.MVSNet 在空间目标三维重建中的应用[J].中国激光,2022,49(23):176-185.
[18] YU Z H, GAO S H. Fast-mvsnet: Sparse-to dense multi-view stereo with learned propagation and gauss newton refinement[C]. Proceedings of the IEEE/CVF Con ference on Computer Vision and Pattern Recognition, 2020:1949-1958.
[19] JCHEN R, HAN S F, XU J, et al. Point-based multi-view stereo network [C]. International Conference on Computer Vision. Seoul, Korea, 2019: 1538-1547.
[20] AANÆS H, RAMSBRØL A J, VOGIATZIS G, et al. Large-scale data for multiple-view stereopsis[C]. International Journal of Computer Vision, Boston, USA, 2016:153-168.
[21] ARONG K, JAESIK P, ZHOU Q Y. Tanks and temples: Benchmarking large-scale scene reconstruction[J]. ACM Transactions on Graphics(ToG), 2017,36(4):1-13.
[22] JOHANNES L S, ZHENG E L, FRAHM J, et al. Pixelwise view selection for unstructured multi-view stereo[C]. European Conference on Computer Vision. Amsterdam, Netherlands, 2016:501-518.
[23] ZHANG J Y, YAO Y, LI S W, et al. Visibility-aware multi-view stereonetwork[C]. British Machine Vision Conference(BMVC), 2020.
[24] PENG R, WANG R J, WANG Z Y, et al. Rethinking depth estimation for multi-view stereo: A unified representation [C]. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, New Orleans Louisiana, 2022: 2201-01501.

作者简介

刘韵婷,博士,副教授,主要研究方向为人工智能、传感器网络、数据分析等。
E-mail:liuyunting0224@163.com