雷有元,周杰,罗岩,等. 基于深度重提取的三维重建[J].湖南科技大学学报(自然科学版),2025,40(4):99-107. doi:10.13582/j.cnki.1672-9102.2025.04.011

LEI Y Y, ZHOU J, LUO Y, et al. 3D Reconstruction Based on Depth Re-extraction [J]. Journal of Hunan University of Science and Technology (Natural Science Edition), 2025, 40(4):99-107. doi:10.13582/j.cnki.1672-9102.2025.04.011

基于深度重提取的三维重建

雷有元[1],周杰[1,2*],罗岩[1],邵根富[3],朱军[1]

(1.南京信息工程大学 电子与信息工程学院,江苏 南京 210044;2.日本国立新潟大学 电气电子工学科, 日本 新潟 950-2181;
3.杭州电子科技大学 自动化学院,浙江 杭州 310000)

摘　要:基于深度学习的三维重建在生活中的许多领域都有应用,但当前绝大多数的研究在特征提取时采用普通卷积,普通卷积对弱纹理和无纹理地区特征提取作用有限,容易模糊,使细节不清晰,影响重建结果,因此,提出了一种基于深度重提取的方法.首先,为了解决普通卷积在低纹理区域的提取错误,提高重建精度,引进一种自适应特征聚合模块,利用可变卷积核的特点,使其在低纹理区域能够自适应的增大卷积核的感受野,在纹理丰富的区域减小感受野;其次,为了聚合不同尺度信息,丰富特征提取信息,使得最终的重建精度有所优化,引进了多空间空洞卷积模块;最后,经过与多组研究对比,所提方法对于低纹理区域的特征提取有较大优化,最终的重建精度也有所提升,整体性提升了 3.4%,可适用于大多数场景.

关键词:三维重建;可变卷积核;自适应特征;空洞卷积

中图分类号:TP391.41　　**文献标志码:**A　　**文章编号:**1672-9102(2025)04-0099-09

3D Reconstruction Based on Depth Re-extraction

LEI Youyuan[1], ZHOU Jie[1,2], LUO Yan[1], SHAO Genfu[3], ZHU Jun[1]

(1. School of Electronic and Information Engineering, Nanjing University of Information Science and Technology, Nanjin 210044, China;
2. School of Electronic and Electrical Engineering, Niigata University, Niigata 950-2181, Japan;
3. School of Communication, Hangzhou University of Electronic Science and Technology, Hangzhou 310000, China)

Abstract: Deep learning based 3D reconstruction has been applied into many fields of daily life, but the vast majority of current research uses ordinary convolutions for feature extraction. Ordinary convolutions have limited ability to extract features from weakly textured and non-textured areas, making them prone to blurring and blurring details, which affects the reconstruction results. Therefore, a depth re-extraction method is proposed. Firstly, in order to solve the extraction errors of ordinary convolution in low texture areas and improve reconstruction accuracy, an adaptive feature aggregation module is introduced, which utilizes the characteristics of variable convolution kernels to adaptively increase the receptive field of convolution kernels in low texture areas and reduce the receptive field in textured areas. Secondly, in order to aggregate information at different scales, enrich feature extraction information, and optimize the final reconstruction accuracy, a multi-space dilated convolution module is introduced. Finally, after comparing with multiple studies, the proposed method has significantly optimized feature extraction in low texture areas, and the final reconstruction accuracy has also been improved, with an overall improvement of 3.4%, making it suitable for most scenarios.

Keywords: 3D reconstruction; variable convolutional kernel; adaptive features; dilated convolution

收稿日期:2023-11-24

基金项目:国家自然科学基金资助项目(61971167)

* **通信作者,**E-mail:zhoujienuist@139.com

三维重建是计算机视觉领域长期研究发展的方向.目前,三维重建已经在自动驾驶、机器人、电影、定位、导航、医学等领域得到广泛应用.三维重建包括传统重建方法和基于深度学习重建方法.传统式三维重建方法按照目标物体深度信息获取方法分为主动式三维重建和被动式三维重建.主动式方法是利用激光、声波、电磁等光源或能量照射到目标物体反射回的光波获取深度信息,但其对设备要求相对严格;被动式方法则是利用自然光反射,通过相机获取图像,通过特定方法计算得到目标物体的空间信息,根据需要的相机数量又分为单目视觉、双目视觉以及多目视觉三维重建方法.基于单目视觉的三维重建是对每张图片进行特征点检测并提取描述子,且根据特征点建立运动轨迹,去除轨迹外的几何约束点[1],从而获取相机的内参以及外参,根据三角测量原理计算图像像素间的位置偏差,来获取景物三维信息,并对稀疏重构进行优化.基于双目视觉的三维重建,主要利用 2 部相机得到 2 组不同的角度图片,然后依靠三角测量原理计算像素间位置偏差恢复出三维点,但其匹配过程时间较久,实时性很差.Code SLAM1 提出一个稠密的场景集合表达,该研究利用神经网络架构,并结合图像几何信息实现单目稠密 SLAM 系统.该研究使用深度学习算法,从单张图像中用神经网络提取出图像深度,相比传统三维重建,获得极大优化.基于深度学习三维重建以实现端到端方式隐式估计物体或者场景 3D 构造,不需要例如关键点检测与匹配等阶段,极大地优化重建过程.三维重建最主要任务是深度估计,在传统方法中利用三角测量原理结合相机内外参数以及相机旋转和平移量获得.而基于深度学习算法是从一组甚至单张照片获得,后者在传统算法中无法实现[3].随着计算机与云计算等技术的发展,基于深度学习三维重建方法开始逐渐成为主流趋势.传统 MVS 算法使用人为计算相似度量与工程正则化(比如归一化互相关和半全局匹配)计算并生成密集点云,但当生成低纹理点云时,纹理复杂区域不完整.其在遮挡、照明变化、无纹理区域仍然有未解决问题.YAO 等[4]提出 MVSNet 网络,首次将深度学习和 MVS 结合,实现端到端训练方式.MVSNet 网络将输入视图分为参考视图和源视图,其中任选一张为参考视图,其余视图为源视图,输入数量不定.先通过一个特征提取网络(权重共享)得到相应特征图,再将源视图特征图的单应性变化投射到参考视图方向上,并根据预存深度范围确定每一张特征图在参考视图主轴上对应深度位置[5];最后计算参考特征图与源视图特征图的相似度匹配值并构造代价体.由于代价体会受到噪声污染,例如被遮挡点、光照影响等,因此,需要进一步对代价体进行正则化,在 MVSNet 网络中利用 3DCNN 处理,通过编码解码集合相邻信息[6],进而得到一个单通道深度图.用 SoftMax 函数进行归一化,得到最终待优化深度图.利用真值深度图与待优化深度图去构成残差网络继续优化深度图,但是 3DCNN 耗费时间和内存,因此通常先进行下采样,再构建成本体[7].为减少内存,R-MVSNet 使用 GRU 顺序正则化 2D 成本映射,但是增加时间复杂度.Fast-MVSNet 构建一个稀疏代价体来学习稀疏深度图,然后使用高分辨率 RGB 图像和 2D CNN 对其进行稠密化[8-10].

上述方法大多采用普通卷积方式,但普通卷积对于弱纹理区域以及薄结构的特征提取效果有限.低纹理区域像素值变化较小,普通卷积再进行加权平均,特征计算会导致信息丢失与特征信息丢失,使卷积结果不准确.另一方面低纹理区域缺乏明显的特征,普通卷积无法有效提取特征,反而产生模糊效果,使细节不清晰.针对这一问题,本文对特征提取模块进行改进.先将一组照片送到特征提取网络,得到不同尺度的特征图,然后分别送入深度图生成网络中,经过多次迭代直接输出粗深度图.最后对粗深度图进行优化生成细深度图.论文主要创新点如下:

1)现有研究大多数方法采用普通卷积,对弱纹理区域的特征提取有限,容易提取错误,产生模糊背景,严重影响深度预测阶段的结果,导致最终重建精度下降.由此,引入自适应聚合模块,该模块主要通过 1 个可变卷积核处理低纹理区域出现的提取错误.利用该卷积核可根据图像纹理复杂程度而自适应地改变感受野大小的特征,使其在纹理丰富的区域感受野小,在纹理不明显区域感受野小一点,减少了特征提取错误,优化了重建精度.

2)通常在提取图像特征时,需要保证特征具有较大的感受野和高分辨率,以免丢失图像边缘的信息.传统的处理方式是采用图像金字塔,将原图调整到不同尺度,再输入到相同的网络中进行特征提取和融合,这种方式能够提高精确度,但是会大幅降低速度.为解决这一问题,采用多空间空洞卷积模块,通过对不同尺度特征图进行特征提取,聚合不同尺度信息,融合更多的特征信息,进而提高模型重建精度.

1　LQ-Net 算法框架

系统框架(LQ-Net 理论)如图 1 所示,基于特征提取深度图生成网络,将输入图像最终转化为粗糙深度图,通过优化模块生成最终深度图.引进重提取网络模块,目标就是优化低纹理区域或无纹理区域.在进入深度图生成网络前,经过特征网络,得到 3 种尺度不同特征图,分别送入 3 个阶段深度图生成网络中,在每个阶段最终会得到该尺度下深度图,再将此深度图通过上采样,输入到下一阶段深度图生成网络中.经过 3 个阶段深度图生成网络最终输出 $(H/2)\times(W/2)$ 深度图.最后一个阶段不引用深度图生成,经过上采样与参考图像输入到优化网络,输出优化后深度图.

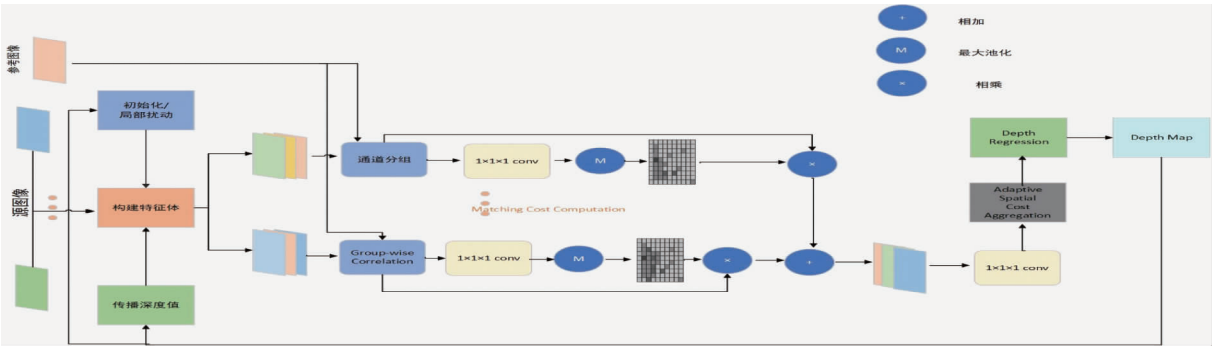


图 1　LQ-Net 系统框图

1.1　特征提取网络

特征提取是计算机视觉中是必不可少的一步,普通特征提取只能提取到局部、低层次特征,此方法重建结果细节粗糙,没有纹理,尤其对高分辨率图片[11]影响更为严重.对此,研究者研究出金字塔结构特征提取网络,特征金字塔是用于检测不同尺度对象,识别系统中基本网络构造.它主要通过在不同图像尺寸上提取特征,来提取不同分辨率物体信息,再融合不同尺度特征,提高网络对物体抗噪干扰能力和对物体形状适应性[12].本文特征提取引进自适应聚合模块,见图 1 虚线框特征提取模块,为本文特征提取网络(Adaptive Extraction Network,AEN)有效解决低纹理或无纹理区域可能会出现的问题,还引进一个多空间空洞卷积,其类似于特征金字塔网络[13],能有效提高特征提取精度,聚合不同尺度特征.先将输入经过 3 种普通卷积,将三通道 RGB 像变为 32 通道特征图,通过 AFA 模块,自适应改变感受野,大大提高特征提取效果,接下来经过一层卷积,变为 64 通道,经过 MSC 模块,金字塔聚合不同尺度特征信息,最后输出 3 种尺度特征图,分别送入 3 个阶段深度图生成网络中.

1.2　自适应特征聚合模块

大多数方法在特征提取模块采取普通卷积,而普通卷积在薄结构或者无纹理以及低纹理区域容易出现错误,影响最终重建精度.为解决普通卷积在反射、低纹理或无纹理区域出现的问题,引进可变卷积核,利用可变卷积核根据纹理复杂度自适应改变感受野大小的特点,即在低纹理区域感受野大一些,在纹理丰富区域感受野小一点,从而设计自适应特征聚合模块(Adaptive Feature Aggregation Module,AFA),如图 2 所示.

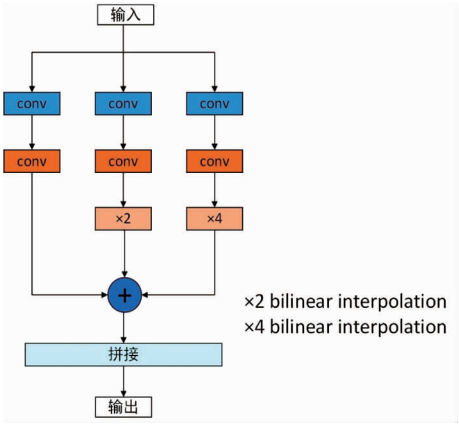


图 2　自适应特征聚合模块

输入图片先通过第一层卷积核特征提取尺度大小不同 3 张特征图,大小分别为 $H\times W$,$(H/2)\times(W/2)$ 和 $(H/4)\times(W/4)$;将 3 张特征图分别送入第 2 层可变性卷积核(3×3 且参数不共享)进行特征聚合;将得到的 3 种尺度进行双线性插值保证尺度一致;通过 C(维度)拼接起来.

可变形卷积定义如式(1)所示.

$$f'(p)=\sum_k w_k f(p+p_k+\Delta p_k)\Delta m_k. \tag{1}$$

式中:$f(p)$ 为像素 p 特征;w_k,p_k 为卷积核权重和偏移量;Δm_k,Δp_k 为可变形卷积层网络自适应产生的偏移量和调整权重;$f'(p)$ 为该点最终的特征.

1.3 多空间空洞卷积模块

在特征提取方面,有时既需要增大感受野,又不想改变图像分辨率,常用方法是采用特征金字塔,通过不同分辨率提取特征最后拼接到一起,但是该种方法效率低[14].而空洞卷积在增大感受野同时也保证提取效率.相较于普通特征金字塔,该模块设置一个采样率参数,在卷积核相邻值之间填充 0 来扩充感受野,一方面增大感受野以便检测分割大目标,另一方面,相较于通过下采样增大感受野,设置采样率可以精确定位目标.设置不同采样率,感受野随之变化,以此获取多尺度信息.本文设计多空间空洞卷积模块(Multi-space cavity convolutionnetwork,MSC),利用不同采样率来提取特征进行拼接,融合不同层次的特征信息,丰富和提高图像特征提取精度.具体流程如见图 3 所示.

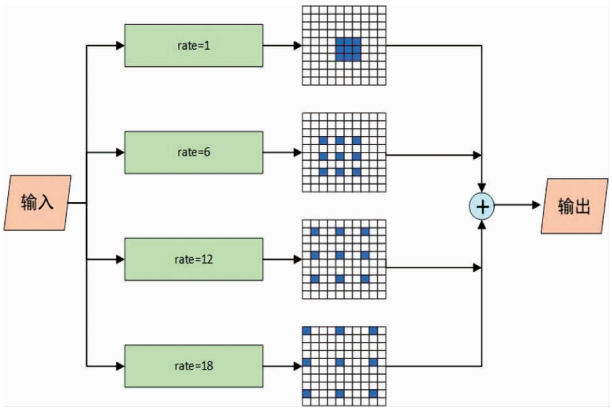


图 3 多空间空洞卷积

1.4 深度图生成模块

深度图生成模块主要引用 Patchmatch 算法.原始 MVSNet 网络形成 3D 代价体后利用 3D 正则化,要想构建精细三维景物,则需要假设更多深度平面,消耗更多时间和内存,为此,将 Patchmatch 算法代替 3D 代价体正则化,消耗内存更小,运行时间更短.Patchmatch 算法在 2 幅图像中搜索最近邻域内相似度最高 Patch,主要是依据随机采样思想,根据图 4 中领域内相似性,在整个区域内快速搜索并匹配,在双目匹配中应用较广.基于深度学习 Patchmatch 包括 3 个步骤:

a)初始化:生成随机假设;

b)传播:向邻域传播假设;

c)评估:计算相应 Patch 匹配代价,选择最优.

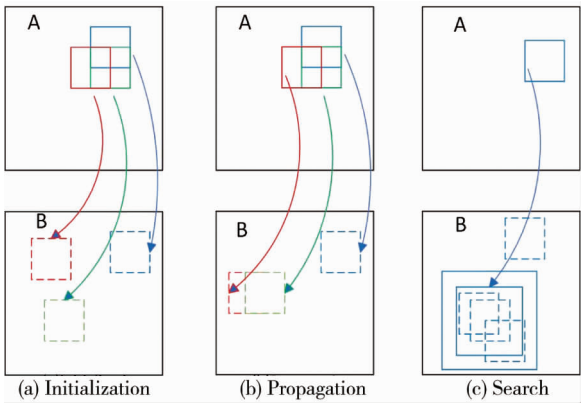


图 4 Patchmatch 算法传播

具体操作步骤:对于图像 A 中每个像素,在 B 中随机找到 1 个像素与之对应,两者之间依靠偏移量联系.之后传播,计算 B 中该像素点周围像素点与 A 中对应像素点的偏移量,选择最小一个.最后,计算在 B 中偏移量最小像素点周围邻域像素点与 A 中像素点偏移量,选择最小一个[15].

该模块在基于传统 Patchmatch 算法基础上结合深度学习,将经过处理特征图最终转化为深度图.具体流程见图 3.在首次阶段中,将 $(H/8)\times(W/8)$ 特征图输入到 Patchmatch 网络中.随机初始化 D_f 个深度样本(为保证均匀采样,在深度区间$[d_{\min}-d_{\max}]$划分 D_f 个区间),在反深度范围内随机采样,这样有利于适应大规模场景和复杂区域.

根据单应性变化矩阵,将源视图特征图像投射在参考视图角度下不同深度平面上,形成特征体.

$$p_{i,j} = K_i[\boldsymbol{R}_{0,i}(\boldsymbol{K}_0^{-1}pd_j) + t_{0,j}]. \tag{2}$$

式中:$p_{i,j}$ 为变化后像素位置;$\boldsymbol{R}_{0,i}$ 为旋转矩阵;$\boldsymbol{K}_0^{-1}$ 为内参矩阵;$t_{0,j}$ 为位移变化.

将每个特征图深度下通道划分为 G 组,分别与参考视图做求相似度操作,形成相似体.如式(2)所示.

$$\boldsymbol{S}_i(p,j)^g = \frac{G}{C}\langle F_0(p)^g, F_i(p_{i,j})^g\rangle. \tag{3}$$

式中:$F_0(p)^g, F_i(p_{i,j})^g$ 为第 g 组参考视图和源视图特征;$\boldsymbol{S}_i(p,j)^g$ 为组内相似性向量.

在经过 1×1×1 卷积与最大池化,输出各像素权重图.经过相似体与重图相乘,得到加权相似体,如式(4)所示.

$$\bar{S}(p,j) = \frac{\sum_{i=1}^{N-1} w_i(p)S_i(p,j)}{\sum_{i=1}^{N-1} w_i(p)}. \tag{4}$$

经过一个 1×1×1 卷积得到代价体 C.如式(5)所示,对于成本聚合,采取基于 Patchmatch 和 AANet 网络自适应空间聚合策略[16].

$$\tilde{C}(p,j) = \frac{1}{\sum_{k=1}^{K_e} w_k d_k}\sum_{k=1}^{K_e} w_k d_k C(p + p_k + \Delta p_{k,j}). \tag{5}$$

最后利用 SoftMax,将成本 C 转换为概率 P 得到最后深度图.此过程为一次迭代,然后再在自适应传播和评估之间迭代,直至收敛,然后进入下一阶段.下一阶段深度样本来源于局部扰动和自适应传播.

局部扰动:以上一阶段生成深度值为中心,逆深度范围区间 $R_k[d_{\max} - d_{\min}]$ 随机采样 N_k 个点.

自适应传播:根据自适应算法选取该像素点周围 K_p 个点深度作为深度样本.

后几次迭代中,在每个像素假设值附近添加扰动,扰动值为归一化逆深度区间值乘以扰动因子,随着迭代次数增加,扰动越来越小.自适应传播是在同一物理表面像素上传播,这对寻找有纹理和无纹理区域深度值更精准[17].

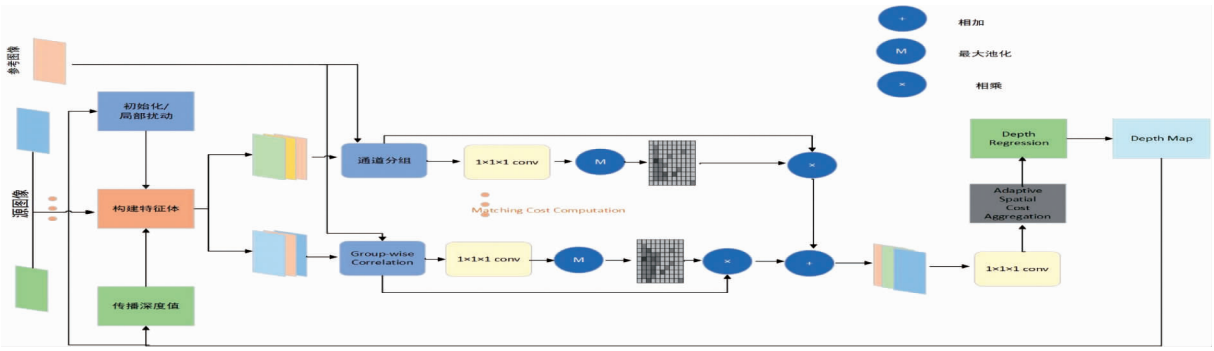


图 5　深度图生成模块流程图

1.5　损失函数

使用损失函数将 3 个阶段产生所有深度图在相同分辨率下损失求和,已经最后一个阶段上采样优化后深度图损失相加.如式(6)所示,其中 L_i^k 为第 k 阶段第 i 次迭代损失,利用 Smooth L1 损失函数,L_{ref}^0 为最后优化后损失.

$$L_{\text{total}} = \sum_{k=1}^{3}\sum_{i=1}^{n_k} L_i^k + L_{\text{ref}}^0. \tag{6}$$

2　试验结果与分析

2.1　试验配置与评价指标

本次实验基于 Pytorch 深度学习框架,Python 编程语言,软件环境为 Python3.10 版本,硬件环境为 TeslaA40 处理器,48 G 内存.开发工具为 PyCharm 专业版 2023.2.1,批处理(batch size)为 4,初始学习率为

0.001,网络训练优化器为 Adam,本试验采样 DTU 数据集.DTU 数据集是一个大规模多视图立体视觉数据集,是在受控实验室环境中采集,具有精确摄像机运动轨迹[18].它包含 7 种不同光照下 49 个场景,共 128 组图片.每组图片都有对应 RGB 图像和相机参数,是近年来 MVSNet 系列文章最常用数据集.取 79 组图片作为训练集,22 组作为测试集,18 组作为评估集合.每张处理图片大小为 1 600×1 200.

本试验使用 Acc(准确性)、Comp(完整性)、overall 作为三维点云的评价指标,数值越低表示重建效果越低.在基于深度学习三维重建任务中最常用数据集有 DTU 数据集、Tanks and Temples 数据集和 ETH3D 数据集.

2.2 对比试验

将所提出的方法与 MVSNet[19], PatchmatchNet[20] 和 AA-RMVSNet[21] 在 Acc, Comp, overall 评价指标下进行对比.在 DTU 数据集上进行测试试验结果显示:批处理为 4,学习率为 0.001,待模型收敛后,根据最终生成的深度图以及点云文件,利用 MATLAB 软件,计算出三维重建的评价指标 Acc 以及 Comp,其结果根据原作者提供的代码和参数进行测试得出.最终得到结果呈柱状图,如图(6)所示.

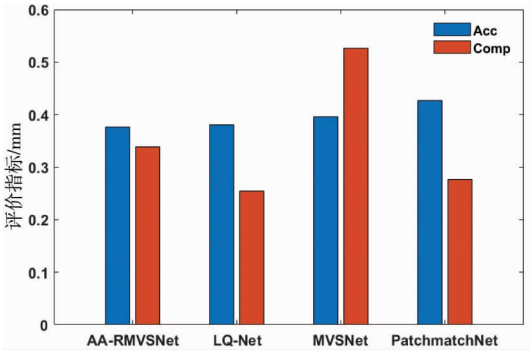


图 6 实验结果对比

基于 Acc 及 Comp,由式(7)可得 overall,如表 1 所示.

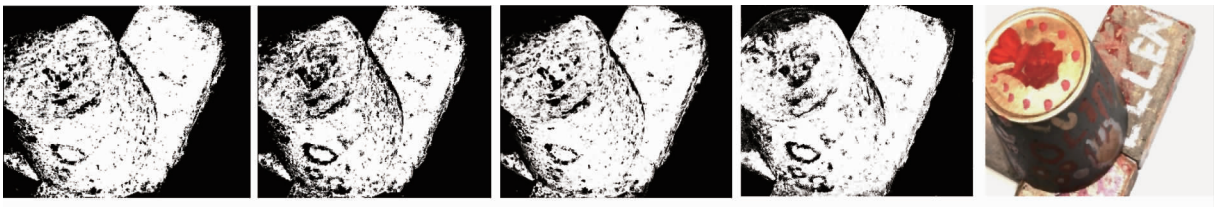
$$\text{overall} = \frac{\text{Acc} + \text{Comp}}{2}. \tag{7}$$

通过表 1 数据可以证明:本文模型在 DTU 数据集下完整性(Comp)较之 PatchmatchNet 提升 2.2%,且精度性(Acc)提升 4.6%,测试结果与 AA-RMVSNet 相比整体(Overall)提升 3.9%.实验表明:自适应特征聚合模块(AFA)和多空间空洞卷积模块(MSC)在精度提升方面有着一定优越性,对于弱纹理区域可能存在问题有着良好抑制性.而与 PatchmatchNet 相比,LQ-Net 特征提取框架 AFA 与 MSC 在特征聚合方面更精细.

表 1 各网络实验数据对比 单位:mm

Methods	Acc.	Comp.	Overall.
MVSNet[19]	0.396	0.527	0.462
PatchmatchNet[20]	0.427	0.277	0.352
AA-RMVSNet[21]	0.376	0.339	0.357
LQ-Net	0.381	0.255	0.318

图 7 为 MVSNet, PatchmatchNet, AA-RMVSNet 和 LQ-Net 最终结果图.通过观察,本文方法在图中红色部分的重建结果明显优于其他研究方法,重建精度更高.



(a) MVSNet (b) PatchmatchNet (c)AA-RMVSNet (d) LQ-Net (e) 原图

图 7 实验结果图对比

其中 MVSNet 在红色部分细节模糊,边缘不清晰,重建结果差;PatchmatchNet 在此区域较 MVSNet 稍有好转,但仍有局限性,细节不够突出,同时背景过于平滑;相比之下,AA-RMVSNet 在红色区域的背景细节处理上比前者清晰,但仍不够完整,在红色区域有缺失.而本文实验结果在红色区域细节突出,背景平滑,重建精度高.同时,由表 1 可知:3 种对比试验的最终重建精度均略差于 LQ-Net,主要是因为普通卷积在特征提取时,由于卷积核感受野固定的原因,会在低纹理区域出现提取错误的同时模糊图像,导致最终的重建结果偏低,而本文在特征提取模块引进自适应聚合模块,可以自适应改变感受野大小,做到纹理复

杂的区域感受野小一点,低纹理区域感受野大一点,有效避免在低纹理区域可能出现的提取错误.从而提高重建精度.同时,本文引进多空间空洞卷积模块,卷积核引进采样率的概念,一方面可以增大感受野,另一方面效率高.本文方法通过聚合多种尺度特征信息,提高最终重建精度.

2.3　消融实验

2.3.1　不同模块的组合对比

在本实验中,在原生 PatchmatchNet 的基础上引进自适应特征聚合模块和多空间空洞卷积模块.为研究各个模块对最后重建结果的具体影响,设计一个消融实验进行比对.分别将 PatchmatchNet,PatchmatchNet 与 AFA,PatchmatchNet 与 MSC 以及 LQ-Net 进行结果分析.在消融试验中,依次加入 MSC 模块,AFA 模块,最终将二者均加入 LQ-Net,其余参数保持不变,仍选择 DTU 数据集,批处理为 4,学习率为 0.001.分别根据实验结果得到的深度图和点云文件,利用 MATLAB 软件计算出最终的评价指标 Acc 和 Comp.

最终得到结果呈现为折线图,如图 8 所示.由表 2 可知:MSC 模块和 AFA 模块对重建结果均有提升,但 AFA 模块提升较之 MSC 模块效果较好.

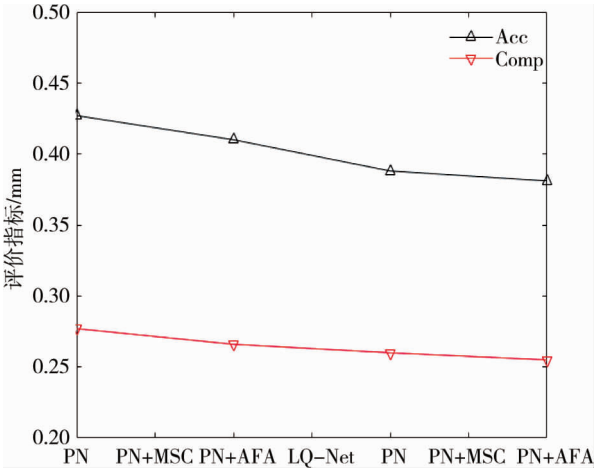


图 8　消融实验结果对比

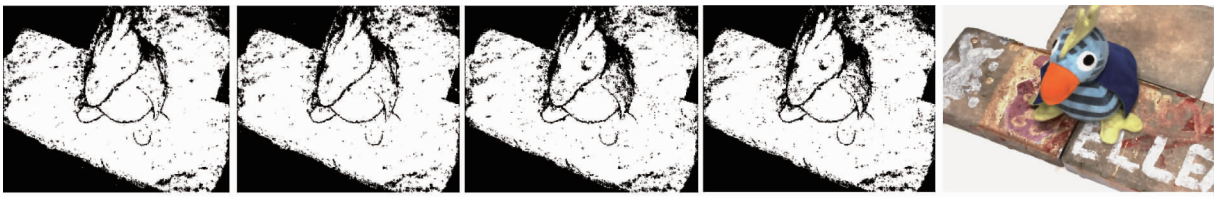
再由式(7),可得 Overall,结果如表(2)所示.

表 2　消融实验分析　　单位:mm

Methods	Acc.	Comp.	Overall.
PatchmatchNet	0.427	0.277	0.352
PatchmatchNet+MSC	0.410	0.266	0.338
PatchmatchNet+AFA	0.388	0.260	0.324
LQ-Net	0.381	0.255	0.318

通过本次消融实验,分别添加 MSC 模块、AFA 模块以及两者均添加,展现各个模块对重建效果提升,图 9 为各种情况下实验结果图.通过实验结果图对该动物眼睛观察,与干净背景图相比,PatchmatchNet 在图像重建过程中边缘清晰度下降,细节模糊的问题较为突出;PatchmatchNet+MSC 重建图像细节情况略有好转但细节恢复情况仍然不足;PatchmatchNet+AFA 在构建动物眼睛周围细节特征的效果表现上相对较好,但仍有局限性.相较之下,LQ-Net 通过单独引入多空间空洞卷积模块(MSC),突出多空间空洞卷积在特征提取的优越性.相比普通金字塔模型,多空间空洞卷积可以最大程度融合特征信息,保证特征信息的丰富性,提高重建精度,在扩大感受野的同时,效率也会有所提升.同时,添加自适应特征提取模块(AFA),可以使低纹理区域的细节特征提取更加精细.普通卷积在特征提取时,不会区分纹理是否复杂,卷积核采取相同的感受野,在低纹理区域容易出现错误.一方面普通卷积在进行卷积操作会将相邻像素信息加权平均,再进行特征计算,在低纹理区域像素值变化较小,加权会导致信息丢失,使卷积结果不准确.另一方面低纹理区域缺乏明显的特征,普通卷积无法有效提取特征,甚至产生模糊效果,使细节不清晰.由此,需要

使用感受野大的卷积核.改进后的特征提取网络可根据纹理复杂度自适应地改变卷积核感受野大小,在纹理复杂区域采取小感受野,低纹理区域采取大感受野,大大降低特征提取错误,提高最终重建精度.本文采取的自适应特征聚合模块较之普通卷积,大大降低特征提取在低纹理区域时出现的错误.其在低纹理区域自动降低感受野,有效避免低纹理区域特征提取时会出现的模糊背景问题.最终结果显示,本文方法最终重建效果优于 PatchmatchNet,其评价指标提升将近 3.4%.



(a) PatchmatchNet (b) PatchmatchNet+MSC (c) PatchmatchNet+AFA (d) LQ-Net (e) 原图

图 9 消融实验结果

2.3.2 实时性测试

根据对试验结果表 3 的详细分析,本文提出的方法在处理单帧图像时与 PatchmatchNet 相比表现出了相近的平均处理时间.尽管在算法中引入了自适应特征聚合模块和多空间空洞卷积模块,这 2 个模块稍微增加了处理时间,但增加的时间并不明显.这意味着,虽然在提高重建精度和质量方面付出了一些额外的计算代价,但这种代价是可以接受的,因为最终的结果明显优于 PatchmatchNet.值得注意的是,尽管处理时间有所增加,但本文方法的最终重建结果却明显优于 PatchmatchNet,展现出更高的重建精度和质量.这一结果表明,通过略微增加处理时间来提高重建精度的方法在实践中具有明显的优势.在现实应用中,重建精度和质量往往是首要考虑的因素,而略微增加处理时间是可以接受的.因此,本研究的方法为图像处理领域提供了一种有益的探索方向,即在追求更高质量的图像重建时,可以适度增加处理时间以换取更好的结果.这种方法不仅为图像处理技术的进步提供了新的思路,也为实际应用中的图像重建任务提供了一种可行的策略.未来,可以进一步研究如何在保持处理效率的同时进一步提高重建精度,以满足不断增长的图像处理需求.因此,本研究的发现对于促进图像处理技术的发展具有重要意义,并为未来的研究和应用提供了新的思路和方法.

表 3 单帧图像平均处理时间

Methods	单帧运行时间/s
PatchmatchNet[20]	0.38
LQ-Net	0.43

3 结论

1)LQ-Net 算法模型介绍:LQ-Net 是一种专为基于深度学习的三维重建特征提取算法进行优化的算法模型.在三维重建中,无纹理或弱纹理的地区往往难以获得准确的重建效果.为了解决这一问题,LQ-Net 算法模型针对这些特定区域进行了改进,有效提高了重建的精度和完整性.

2)自适应聚合模块的引入:在深度学习的三维重建中,普通卷积在处理弱纹理区域时存在提取错误的问题,这会严重影响深度预测阶段的结果.为了解决这一问题,LQ-Net 算法模型引入了自适应聚合模块.这个模块能够根据图像的纹理复杂程度自适应地改变卷积核的大小.在纹理丰富的区域,感受野被设置得较小;而在纹理不明显的区域,感受野稍大,这样能够有效地减少特征提取错误,从而优化重建的精度.

3)多空间空洞卷积模块的应用:在传统的图像特征提取中,为了保证特征具有较大的感受野和高分辨率,通常会采用图像金字塔的方式调整图像尺度.然而,这种方法会大大降低处理速度.为了在不牺牲速度的前提下提高重建精度,LQ-Net 算法模型采用了多空间空洞卷积模块.这个模块可以对不同尺度的特征图进行特征提取,有效地聚合不同尺度的信息,并融合更多的特征信息,从而提高了模型的重建精度.

4)未来研究方向:虽然 LQ-Net 算法模型已经在重建精度和完整性方面取得了显著的改进,但还有一

些潜在的改进空间.未来,研究将主要集中在对算法的时间成本进行优化,以提高模型的实用性和效率.此外,为了进一步优化模型的完整性和鲁棒性,计划将注意力机制引入 LQ-Net 算法模型中,以提高模型对关键信息的关注度,从而进一步提高重建的精度和准确性.这些改进将使 LQ-Net 算法模型在三维重建领域具有更广泛的应用前景.

参考文献:

[1] KIM M, KIM H S, KIM H J, et al. Thin-slice pituitary MRI with deep learning-based reconstruction: diagnostic performance in a postoperative setting[J]. Radiology, 2021, 298(1): 114-122.

[2] 佟帅, 徐晓刚, 易成涛, 等. 基于视觉的三维重建技术综述[J]. 计算机应用研究, 2011, 28(7): 2411-2417.

[3] XUE B, HE Y, JING F, et al. Robot target recognition using deep federated learning[J]. International Journal of Intelligent Systems, 2021, 36(12): 7754-7769.

[4] YAO Y, LUO Z X, LI S W, et al. MVSNet: depth inference for unstructured multi-view stereo[M]//Computer Vision-ECCV 2018. Cham: Springer International Publishing, 2018: 785-801.

[5] HE P P, HU D L, HU Y. Deployment of a deep-learning based multi-view stereo approach for measurement of ship shell plates [J]. Ocean Engineering, 2022, 260: 111968.

[6] LI Y, ZHAO Z J, FAN J H, et al. ADR-MVSNet: a cascade network for 3D point cloud reconstruction with pixel occlusion [J]. Pattern Recognition, 2022, 125: 108516.

[7] 陈凤东, 洪炳镕. 基于特征地图的移动机器人全局定位与自主泊位方法[J]. 电子学报, 2010, 38(6): 1256-1261.

[8] XUE B, YI W J, JING F, et al. Complex ISAR target recognition using deep adaptive learning[J]. Engineering Applications of Artificial Intelligence, 2021, 97: 104025.

[9] XUE B, TONG N N. Real-world ISAR object recognition using deep multimodal relation learning[J]. IEEE Transactions on Cybernetics, 2020, 50(10): 4256-4267.

[10] XUE B, TONG N N. DIOD: fast and efficient weakly semi-supervised deep complex ISAR object detection[J]. IEEE Transactions on Cybernetics, 2019, 49(11): 3991-4003.

[11] 姜良美, 王芳, 肖志坤, 等. 基于纹理特征的微山湖湿地信息提取研究[J]. 湖南科技大学学报(自然科学版), 2011, 26(4): 68-72.

[12] HAN J W, CHEN X M, ZHANG Y T, et al. DEMVSNet: Denoising and depth inference for unstructured multi-view stereo on noised images[J]. IET Computer Vision, 2022, 16(7): 570-580.

[13] XUE B, HE Y, JING F, et al. Dynamic coarse-to-fine ISAR image blind denoising using active joint prior learning[J]. International Journal of Intelligent Systems, 2021, 36(8): 4143-4166.

[14] GONG C Y, SONG Y C, HUANG G Y, et al. BubDepth: a neural network approach to three-dimensional reconstruction of bubble geometry from single-view images[J]. International Journal of Multiphase Flow, 2022, 152: 104100.

[15] YANG Y F, LI Q, YU Y Y, et al. Continuous digital zooming using generative adversarial networks for dual camera system [J]. IET Image Processing, 2021, 15(12): 2880-2890.

[16] CHEN B, ZHAO J, PAN B. Mirror-assisted multi-view digital image correlation with improved spatial resolution[J]. Experimental Mechanics, 2020, 60(3): 283-293.

[17] LI S, CAO J, YAO J F, et al. Adaptive aggregation with self-attention network for gastrointestinal image classification[J]. IET Image Processing, 2022, 16(9): 2384-2397.

[18] TANG Y Z, YANG X, WANG N N, et al. Person re-identification with feature pyramid optimization and gradual background suppression[J]. Neural Networks, 2020, 124: 223-232.

[19] WANG F, GALLIANI S, VOGEL C, et al. PatchmatchNet: learned multi-view patchmatch stereo[C]//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2021: 14194-14203.

[20] WEI Z Z, ZHU Q T, MIN C, et al. AA-RMVSNet: adaptive aggregation recurrent multi-view stereo network[C]//2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, 2021: 6167-6176.