

DOI: 10.13382/j.jemi.B2307072

融合 ECA 注意力层和轻量正则化的多视图三维重建*

刘韵婷 高宇 戴佳霖 谭明晓

(沈阳理工大学自动化与电气工程学院 沈阳 110159)

摘要: 为了有效解决多视图三维重建中的边缘缺失、网络内存消耗严重、重建精度低的问题,对基于深度学习的多视图三维重建网络的特征提取、正则化网络、损失函数、优化器等进行研究。首先,使用融合 ECA 注意力层的特征提取网络,提高网络对通道特征信息的关注;然后,在门控循环单元模块中加入卷积层,改进的门控循环单元组合成 GC 正则化网络,采用此网络对代价体进行正则化,降低网络的计算量;最后,使用 SmoothL1 损失函数和 Adam 优化器,提高模型训练后期的收敛精度,优化模型的损失和参数。在 DTU 公开数据集上训练和测试,提出的融合注意力机制和轻量正则化的多视图三维重建网络(EGF-MVSNet)相比于经典的 MVSNet 网络完整性上提高了 22.1,模型总体评分提高了 11.5%。能够实现物体的重建,显著提高重建结果的质量,降低网络对内存的消耗。

关键词: 三维重建;多视图;深度学习;ECA 注意力

中图分类号: TP391;TN911.73 **文献标识码:** A **国家标准学科分类代码:** 510

Multi-view 3D reconstruction combining ECA attention layer and lightweight regularization

Liu Yunting Gao Yu Dai Jialin Tan Mingxiao

(School of Automation and Electrical Engineering, Shenyang University of Technology, Shenyang 110159, China)

Abstract: In recent years, multi-view 3D reconstruction technology based on deep learning has become one of the research hotspots in the field of machine vision and is applied in many fields. However, the 3D reconstruction technology still has problems such as edge missing, serious network memory consumption and low reconstruction accuracy. In this paper, based on the existing problems of 3D reconstruction technology, EGF-MVSNet network is proposed based on deep learning. First, a feature extraction network incorporating the ECA attention layer is used to improve the network's attention to the channel feature information; then, an improved combination of GRU modules is used to obtain the GC regularization network for regularization and to reduce the computation of the network; finally, the SmoothL1 loss function and Adam optimizer are used to improve the convergence accuracy at the later stage of model training and to optimize the model's losses and parameters. Through testing and validation on the DTU public dataset, the EGF-MVSNet network proposed in this paper improves the completeness by 22.1% and the overall model score by 11.5% compared to the MVSNet network, which confirms that the EGF-MVSNet network can significantly improve the quality of the reconstruction results and reduce the network's consumption of memory.

Keywords: 3D reconstruction; multi-view; deep learning; ECA attention

0 引言

近年来,深度学习^[1-3](deep learning, DL)的发展彰

显了其独特的优越性,大批的学者也将深度学习应用于多视图三维重建领域,不仅提升了目标物体重建的效果,也提高了目标图像恢复的鲁棒性。基于深度学习的三维重建按相机数量可划分为单目、双目以及多目的重建^[4]。

收稿日期: 2023-11-23 Received Date: 2023-11-23

* 基金项目: 辽宁省自然科学基金项目(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图三维重建按重建方式又可分为基于点云重建、体素重建、网格重建和深度图重建。

其中,基于深度图的三维重建方法是多视图重建研究中应用较多的方法,该算法输入多张图像,依次通过特征提取、代价体构建、代价体正则化、深度图优化,最终进行目标图像的重建。Gallup 等^[5]将 Yang 等^[6]提出的 Plane-Sweeping 算法应用于三维重建的研究中,采用多个分割平面投影到摄像机上点的颜色进行一致性对比的方法,判断该点是否为物体表面上的点。此方法易受物体遮挡、光照以及物体表面纹理的影响,为了有效地解决此问题,Shu 等^[7]提出了一种深度区间估计的方法。虽然采用几何的计算法可以实现对深度图的计算,但实际使用中易受到物体遮挡的影响,对物体缺失纹理的区域具有较差的结果。

为了有效解决传统 MVS 方法在重建过程中对物体弱纹理区域、表面反射等问题处理较差,基于深度学习的多视图三维重建算法应运而生。文献[8-15]中表明,MVSNet 重建网络为基于深度学习的多视图三维重建开辟了一个全新的研究领域。但此网络在进行正则化时,需要消耗大量的内存,这就使得该网络难以对高分辨率的场景进行重建。

Yao 等^[16]在 MVSNet 网络的基础上进行改进,提出了 R-MVSNet 网络,有效的提高了重建网络的整体性能。Yu 等^[17]提出了 AACVP-MVSNet 网络,引入自注意层,实现分层特征提取;使用相似度量的方法生成成本量,提高

了网络对大规模图像重建的能力。Wei 等^[18]提出了 AA-RMVSNet 网络,采用了视图内聚合模块,提高上下文间的感知力,减小遮挡问题对重建结果的影响。Gao 等^[19]提出了 MSCVP-MVSNet 网络,采用多策略的方法以及训练过程中使用单峰分布作为训练标签的方式,提高重建结果的精确度。Giang 等^[20]提出了动态尺度特征提取网络,提高网络的鲁棒性,减小运行的时间。

本文针对当前基于深度学习的多视图三维重建网络存在的重建精度低、完整度较差和内存消耗大的问题展开研究。首先,在基础的特征提取网络基础上添加 ECA (efficient channel attention) 注意力层模块,提高对通道特征信息的关注;然后,使用 GC 正则化网络,减少网络的计算量及网络对内存的消耗,同时使用优化的损失函数和 Adam(adaptive moment estimation) 优化器,提高网络训练后期学习率保持不变时的收敛精度并且优化网络模型中的参数,提升训练模型的鲁棒性。

1 EGF-MVSNet 网络模型

随着机器视觉的发展,深度学习在三维重建中的应用被广泛研究,MVS 系列算法在多视图的重建中表现更出色,能够更好的还原初始模型。本文选用 Fast-MVSNet 作为基础网络,网络整体划分为特征提取、代价体以及深度图优化 3 部分,融合注意力机制和轻量正则化的多视图三维重建网络 (EGF-MVSNet) 网络的整体框图,如图 1 所示。

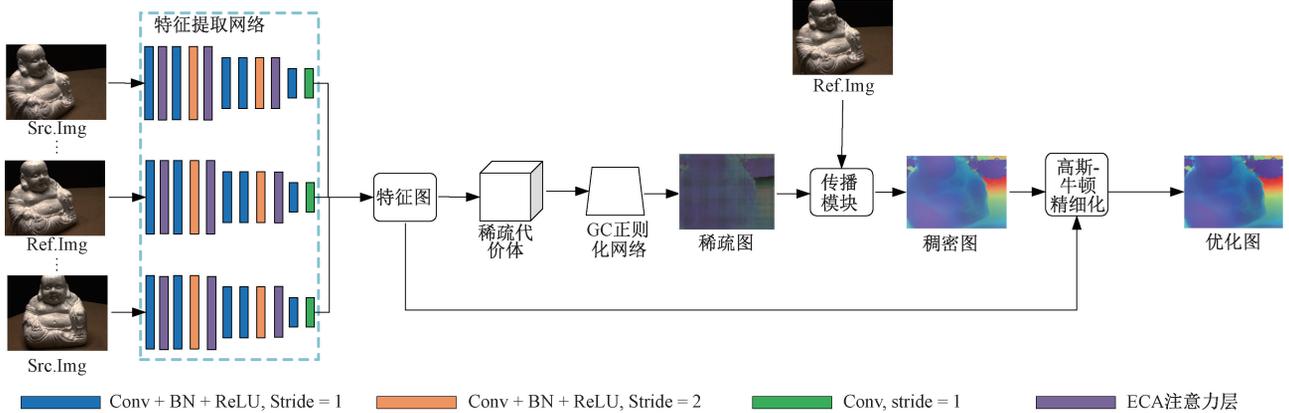


图 1 EGF-MVSNet 网络整体框图

Fig. 1 EGF-MVSNet overall network diagram

其中,在 2D 卷积神经网络 (convolutional neural networks, CNN) 组成的特征提取网络中添加 ECA 注意力层。代价体部分又分为代价体构建和代价体正则化,特征图经变换后得到代价体,代价体经 GC 正则化网络完成正则化;深度图优化过程中使用高斯—牛顿算法,实现对深度图由粗到细的精细化。同时那个优化的损失函数

和适配的优化器,提高网络的整体性能和增强网络的泛化能力。

1.1 融合 ECA 注意力层的特征提取网络

特征提取网络旨在将输入的二维图像特征提取出来,为后续的重建提供可靠的物体信息。Fast-MVSNet 多视图三维重建网络在特征提取过程中采用了 8 层的

2DCNN 网络,将 1 个输入的参考图像与 $N - 1$ 个原图像经特征提取网络变换后,输出 N 个 32 通道的特征图。由于输入图像为 RGB 三通道图像,在特征提取过程中会忽略掉一些重要的通道信息,导致重建结果的精确性和完整性下降。

为避免上述问题对重建结果的影响,提高重建结果对物体细节等重要特征信息的还原,本文提出了融合 ECA 注意力层的特征提取网络。ECA 注意力模块是通道注意力机制类别中的一种,采用较少的参数,提高特征提取过程中对物体特征通道信息的关注,使获得的特征图更大限度的接近于原图像特征,重建结果中展示出更全面的物体信息。

1.2 轻量化的正则化网络

将特征提取网络输出的 N 个特征图变换得到的 N 个特征体,对特征体进行计算得到代价体。正则化过程就是对代价体进行深度处理,得到一个概率体。MVSNet 网络正则化使用了传统的三维卷积神经网络,对大小为 $1/4H \times 1/4W \times N \times F$ (N 为深度平面数, F 为特征通道数)的三维成本量进行正则化处理,该过程中将会进行大量的计算,也是导致网络内存消耗严重的原因之一。通过对目前常用的神经网络以及正则化方法研究后,本文提出了一种可以改善上述问题的正则化网络。

1) GRU 模块概述

随着研究的深入,神经网络的多种变形网络应运而生,不仅改进了原本网络的弊端,也提升了网络的性能。

神经网络可划分为卷积神经网络和循环神经网络 (recurrent neural network, RNN) 两大类,本文对传统的门控循环单元 (gate recurrent unit, GRU) 作出改进,组合后得到 GC 正则化网络,这样不仅能够解决噪声对重建的影响,也会大大降低计算量,达到减小内存消耗的目的。GRU 是长短期记忆网络 (long short term memory, LSTM)^[21-22] 的一种变体,属于循环神经网络,主要部分为更新门和重置门。每个状态的更新如式(1)所示。

$$\begin{aligned} R_t &= \sigma(W_R \cdot [H_{t-1}, X_t]) \\ Z_t &= \sigma(W_Z \cdot [H_{t-1}, X_t]) \\ \tilde{H}_t &= \tanh(W_{\tilde{H}} \cdot [R_t \times H_{t-1}, X_t]) \\ H_t &= (1 - Z_t) \times H_{t-1} + Z_t \times \tilde{H}_t \end{aligned} \quad (1)$$

式中: R_t 代表重置门, Z_t 代表更新门, H_t 代表传递到下一时刻的隐藏状态, \tilde{H}_t 代表候选隐藏状态。

2) GC 正则化网络

由于代价体正则化过程中需要把上下文信息间的关系以及通道间的关联性考虑进来,本文提出在传统的 GRU 模块中加入卷积门 (卷积层),这样不仅能够实现对成本量的映射,也使代价体的特征尽可能的保留下来。在使用过程中,仅仅一个 GRU 单元并不能很好的实现网络的正则化,所以采用了 3 个改进后的 GRU 单元作为一个正则化模块,将上层的输出作为下层的输入,并把多个正则化模块组合到一起,形成了 GC 正则化网络,如图 2 所示。

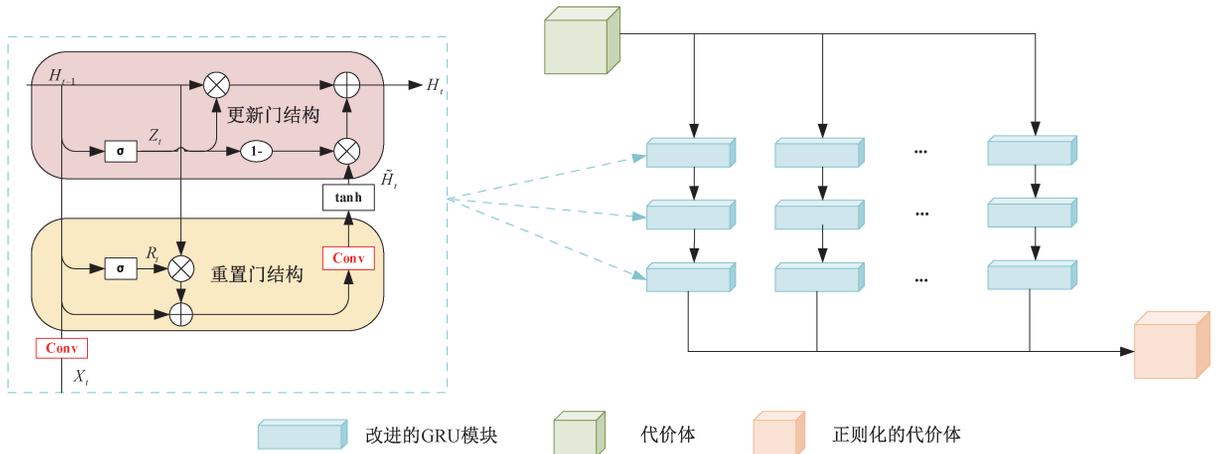


图2 GC 正则化网络

Fig. 2 GC regularization network

采用 GC 正则化网络能够在纵向上实现上下文信息较好的提取,在横向上可实现实时更新与重置。这不仅能够达到对代价体正则化的效果,也减少了正则化过程中的计算量,有效提升了网络的记忆能力,提高重建的完整度。

1.3 基于 Gauss-Newton 算法的深度图优化

本文使用的基础网络采用了由粗到细推断策略,提高了深度图的优化质量,更好的还原了物体的图像特征。正则化后得到一个概率体,经处理可得到一个稀疏的高分辨率深度图。若使用稀疏的高分辨率深度图进行优化

和重建,难以高度还原出原图像,所以需获得更稠密的深度图,再进行下一步操作。获得稠密深度图最简单的方式是采用最近邻的方法,但是此方法没有考虑到原始图像对深度图的影响,可能会获取不到边界信息。因此,本文采用变体的联合双边上采样对稀疏深度图进行稠密化。图像深度通过式(2)得到。

$$\tilde{D}(p) = \frac{1}{\sum_{z_p, q \in N(p)} D(q)} \cdot \omega_{p,q} \quad (2)$$

其中, $\tilde{D}(p)$ 代表的是稠密深度图, $D(q)$ 代表的是稀疏深度图, $\omega_{p,q}$ 代表网络的权重,此权重通过网络的学习即可得到。

虽然这种获得深度图的方式能够大幅度减少内存和成本的消耗,但是采用稠密深度图进行重建,精度远远不够。所以对深度图还需进行再优化,出于对重建结果质量的考量,本文采用了高斯—牛顿算法^[23]进行深度图优化,优化过程如下。

参考图像上的点为 p , p 在第 i 张原图像上对应的投影点为 p'_i , p'_i 由式(3)得到:

$$p'_i = K_i(R_i R_0(\tilde{D}(p)K_0^{-1}p - t_0) + t_1) \quad (3)$$

式中: $\{K_i, R_i, t_i\}_{i=0}^N$ 分别表示对应图像的相机内参、旋转矩阵和平移向量。稠密深度图预测出的深度为 $D(p)$, 优化过程主要是求误差(4)最小化的过程。

$$E(p) = \sum_{i=1}^N \|F_i(p'_i) - F_0(p)\|_2 \quad (4)$$

式中: F_i 为原图像的深度特征图, F_0 为参考图像的深度特征图。像素点在每个原图像中的残差为 $r_i(p)$, 即原图像与参考图像的深度特征图的差值,由式(5)得到:

$$r_i(p) = F_i(p'_i) - F_0(p) \quad (5)$$

$J_i(p)$ 表示每个残差对 $\tilde{D}(p)$ 求一阶偏导数的结果,由式(6)得到:

$$J_i(p) = \frac{\partial F_i(p'_i)}{\partial p'_i} \cdot \frac{\partial p'_i}{\partial \tilde{D}(p)} \quad (6)$$

多个 J 堆叠为雅可比矩阵 $\{J_i(p)\}_{i=1}^N$, 多个 r 组合成了残差向量 $\{r_i(p)\}_{i=1}^N$, 通过式(7)可得到当前深度值的增量值 δ 。

$$\delta = -(J^T J)^{-1} J^T r \quad (7)$$

最后,经优化后的深度图表达如式(8)所示。

$$\tilde{D}'(p) = \tilde{D}(p) + \delta \quad (8)$$

1.4 优化损失函数

经典的 Fast-MVSNet 网络利用估计深度图和真实深度图的均值的绝对差(L1)作为损失函数,L1 损失函数使用简单,但会造成模型震荡,收敛困难等问题。重建过程中需进行平面、空间以及地面真实坐标间转换,因此,本文参考目标检测领域,提出使用 SmoothL1 损失函数作为重建过程中的损失函数,如式(9)所示。

$$\text{SmoothL1} = \begin{cases} 0.5x^2, & |x| < 1 \\ |x| - 0.5, & \text{其他} \end{cases} \quad (9)$$

其中, $x = |\tilde{D}'(p) - \hat{D}(p)|$ 表示预测值与真实值的差值。当预测值和真实值差值的绝对值小于 1 时,采用平方误差的方法计算损失函数,反之采用线性误差的方式计算。这就会提升网络的鲁棒性,能够对异常的值反应更加稳定。SmoothL1 损失函数不仅能够衡量位置损失,也能够优化原网络训练后期学习率不变时,收敛精度低的问题,提高重建质量。

2 实验结果与分析

2.1 DTU 数据集

本文实验使用了 DTU 公开数据集^[24-25],DTU 数据集中包括了 128 个不同的场景,每个场景中捕获了 49 个或者 64 个不同相机位置的图片,每个位置的图片由 7 种不同光照亮度组成^[26],数据集提供了由精确的结构光扫描仪扫描的物体模型和高分辨率的 RGB 图像,以及包括由 MATLAB 标定工具得到的高精度相机内参和外参。由于本网络是由粗到细的网络,在对网络进行测试过程中采取修正后的 DTU 测试集。

本文使用 3 个评估指标对测试结果进行评估,分别为精确度 (accuracy, Acc.)、完整度 (completeness, Comp.) 和总体评分 (overall, OA.)。其中,精确度用来评估重建点云相对真值点云在空间中的位置准确性,完整度反映的是重建点云相对于真值点云的完整性,总体评分是对点云重建的总体评估。评估过程中,数值越小代表网络性能越高。式(10)为精确度计算公式, R 代表重构点云集合, G 代表真值点云集合:

$$\text{Accuracy} = \frac{1}{|R|} \sum_{r \in R} e_{r \rightarrow G} \quad (10)$$

其中, $e_{r \rightarrow G}$ 表示重构点云上一点 $r(r \in R)$ 到真值点云的欧式距离,由式(11)得到:

$$e_{r \rightarrow G} = \min_{g \in G} \|r - g\| \quad (11)$$

完整度的计算由式(12)得到:

$$\text{Completeness} = \frac{1}{|G|} \sum_{g \in G} e_{g \rightarrow R} \quad (12)$$

其中, $e_{g \rightarrow R}$ 表示 Ground Truth 点云上一点 $g(g \in G)$ 到重构点云的欧式距离,由式(13)得到:

$$e_{g \rightarrow R} = \min_{r \in R} \|g - r\| \quad (13)$$

总体评价 (Overall) 的计算如式(14)所示,是精确度和完整度的平均值:

$$\text{Overall} = \frac{\text{Accuracy} + \text{Completeness}}{2} \quad (14)$$

2.2 优化器的选择

Fast-MVSNet 网络的优化器使用了 RMSProp,本文提

出使用 SmoothL1 作为损失函数,为配合网络达到更好的效果,优化器采用了 Adam。Adam 优化器是在 RMSProp 基础上做出梯度加权平均和偏差纠正的改进,随着梯度变得稀疏,Adam 优化器的效果更胜一筹。因此,使用 Adam 优化器不仅能够优化模型的损失和参数,也能够提高网络模型的鲁棒性。

2.3 实验配置

本文实验使用配置为:GPU 选用 NVIDIA GeForce GTX 3080Ti,操作系统选用 Ubuntu18.04,深度学习框架为 Pytorch 1.8.1,CUDA 11.1,cuDNN 8.0.5。

实验设置:训练使用的输入图片分辨率为 640×512 ,使用 Adam 优化器,初始学习率设置为 0.001 ,深度平面选为 $D=48$,在显卡上进行了 16 次的训练,选用分辨率为 1280×960 的图片作为测试集输入。

2.4 结果分析

为验证本文提出的网络在重建结果整体上均有提高,使改进后的网络模型与 Gipuma、MVSNet、R-MVSNet、CVP-MVSNet、PointMVSNet、Fast-MVSNet 和 Vis-MVSNet 在 DTU 数据集上进行精确性、完整性以及模型整体性能的比较,如表 1 所示为实验结果对比。

根据表 1 数据可知,本文提出的网络模型的总体评分优于 COLMAP、MVSNet、R-MVSNet、PointMVSNet、Fast-MVSNet、Vis-MVSNet、AA-RMVSNet 和 CVP-MVSNet 几个网络模型。在完整性上,本文提出的网络相比 COLMAP、

表 1 DTU 数据集实验结果对比表(数值越小越好)

Table 1 Comparison Table of DTU Dataset Experimental Results (Smaller values are better)

模型	Acc./mm	Comp./mm	OA./mm
COLMAP ^[15]	0.400	0.664	0.532
MVSNet ^[24]	0.396	0.572	0.462
R-MVSNet ^[16]	0.385	0.459	0.422
PointMVSNet ^[10]	0.342	0.411	0.376
Fast-MVSNet ^[13]	0.336	0.403	0.370
Vis-MVSNet ^[27]	0.369	0.361	0.365
AA-RMVSNet ^[18]	0.376	0.339	0.357
CVP-MVSNet ^[28]	0.296	0.406	0.351
本文	0.343	0.351	0.347

MVSNet、R-MVSNet、PointMVSNet、Fast-MVSNet、Vis-MVSNet、CVP-MVSNet 分别提高了 31.3%、22.1%、10.8%、6%、5.2%、1%、5.5%;对于模型总体评分,本文提出的网络比 Gipuma、COLMAP、MVSNet、R-MVSNet、PointMVSNet、Fast-MVSNet、Vis-MVSNet、AA-RMVSNet 和 CVP-MVSNet 分别提高 23.1%、18.5%、11.5%、7.5%、2.9%、2.3%、1.8%、1%、0.4%。

如图 3 所示即为本文整体改进后的网络与原始图像、MVSNet 网络在 DTU 数据集上的重建结果对比。对比原始图像以及 MVSNet 重建的结果图,可直观的看出本文提出的网络在重建结果的精确度上有很大的提升。

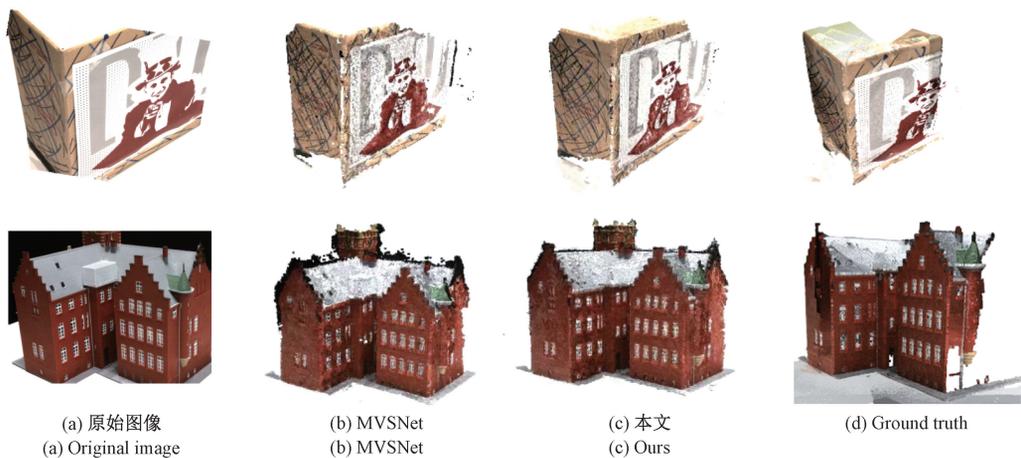


图 3 DTU 数据集点云图可视化结果对比

Fig. 3 Comparison of visualization results of DTU dataset point cloud maps

同时,为了更好的说明 EGF-MVSNet 网络能够提升目标物体重建的完整度,本文选取了 scan9 场景, Fast-MVSNet、本文网络以及 Ground Truth 的点云结果如图 4 所示。对图 4 中的实线框部分放大,并且采用虚线框起,明显的观察到本文提出的网络重建的结果中,点云结果的完整度更高。

由于数据集中含有建筑、雕塑、食物等等几类物体,在图像获取时可能会发生建筑上玻璃光反射或是物体的颜色辨识度低、特征不够明显的现象,这些现象将会导致网络对物体特征的识别不全面,大大降低重建的结果。本文在特征提取网络中加入 ECA 注意力层,能够更加的关注特征通道的信息。提出的 GC 正则化网络不仅可以

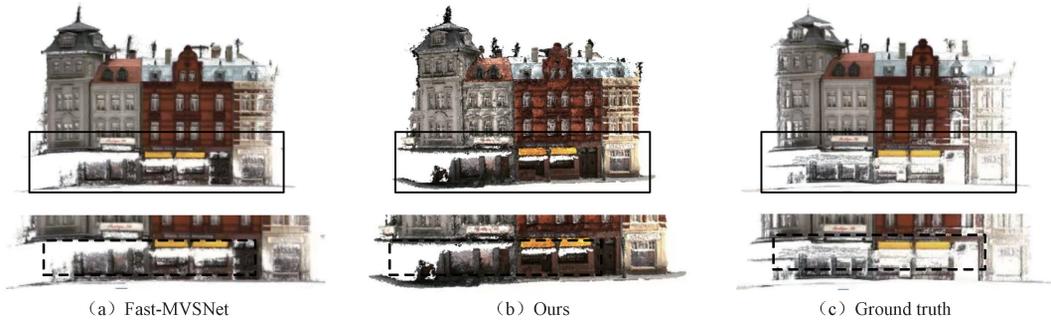


图 4 scan9 场景的重建点云结果对比

Fig. 4 Comparison of reconstructed point cloud results in the scan9 scene

考虑到上下文和通道间的信息,保留更多的特征信息。同时网络中采用由粗到细的方式获取深度图、使用 SmoothL1 损失函数和 Adam 优化器,提高网络模型的性能,输出高分辨率的深度图,保证物体的特征获取的更完全,提高重建的完整性。实验证明,本文提出的改进网络能够使重建结果更完整,与原始图像的契合度更高。

网络在 DTU 数据集上对显卡内存的消耗越小、网络的评价指标参数越小说明网络的性能越好。本文改进网络与经典重建网络在网络内存消耗以及网络各评价指标的对比结果,如图 5 所示。相比于经典的 Fast-MVSNet 网络,本文提出的改进后网络内存的消耗较大,但网络的整体评分以及完整度评分均优于 Fast-MVSNet 网络,说明改进后的网络性能更高,重建结果的质量更高,能够更好地还原物体的信息。

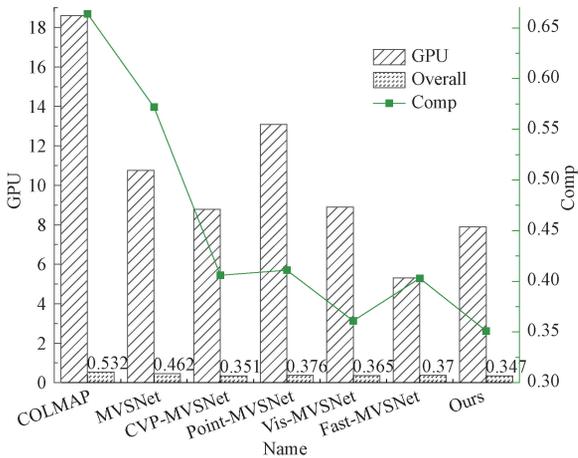


图 5 内存消耗及网络评价指标对比

Fig. 5 Comparison chart of memory consumption and network evaluation indicators

3 消融实验

为验证本文提出的融合 ECA 注意力层的特征提取

网络、GC 正则化网络、SmoothL1 损失函数以及 Adam 优化器能够提升网络的性能,获得质量更高的重建结果,本文进行了多组消融实验,从定量评价指标等多个方面分析。如表 2 所示,为采用融合 ECA 注意力层的特征提取网络和使用 8 层 2DCNN 特征提取网络以及损失函数与优化器改进前后网络的实验对比结果。采用 Acc.、Comp. 和 OA. 作为对比依据,评价指标的数值越小,说明网络的性能越好。

表 2 特征网络、损失函数+优化器改进前后在 DTU 数据集上评分对比

Table 2 Comparison of scores on DTU datasets before and after improvement of feature networks, loss functions, and optimizers

方法	Acc./mm	Comp./mm	OA./mm
2DCNN	0.336	0.403	0.370
ECA+2DCNN	0.352	0.377	0.365
ECA+GC+L1+RMSProp	0.329	0.387	0.358
ECA+GC+SmoothL1+Adam	0.343	0.351	0.347

根据表 2 数据显示,采用融合 ECA 注意力层的特征提取网络能够提升网络的整体性能,尤其是提高了网络重建的完整性。改进了损失函数和优化器的网络,完整性评价指标和网络整体性评价指标的数值均减小,说明提出的改进部分能够提升网络重建的完整性。这也证实了提出改进的网络具有更好的性能,能使物体的重建结果与原图像的贴合度较高。

在特征提取网络中加入注意力层,会增加网络对内存的消耗,为尽可能地减小网络对内存地消耗,降低运行地时间,本文提出了 GC 正则化网络。采用该网络可以降低正则化过程中对显卡内存的消耗,同时可以增加对上下文信息和通道间关联性的注意。本文在 DTU 数据集上进行了 4 组实验,以网络对内存的占用和运行时间为评估参数,实验结果如图 6 所示。

本文分别进行了 2DCNN+3DCNN、2DCNN+GC、ECA

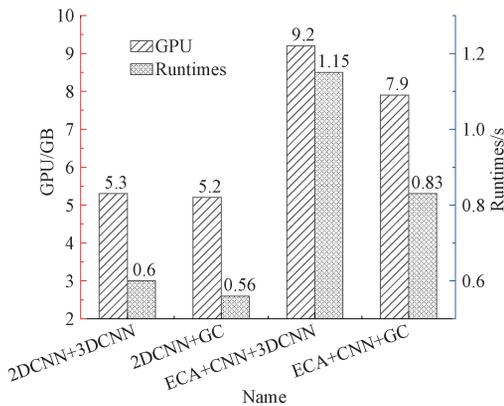


图6 网络改进前后实验对比数据

Fig. 6 Comparison data graph of experiments before and after network improvement

+2DCNN+3DCNN、ECA+2DCNN+GC 四组实验,通过图6显示的对比数据可知,采用改进的 GC 正则化网络能够有效的降低网络对内存的消耗,同时提升网络的运行速度,提高网络的性能。

4 结论

本文在基础的 Fast-MVSNet 网络上提出了 EGF-MVSNet 多视图三维重建网络。在基础的 8 层 2DCNN 特征提取网络中加入 ECA 注意力层,提升了网络对通道特征的关注度。使用 GC 代价体正则化网络,解决了原始网络中使用 3DCNN 对代价体进行正则化时产生的计算量较大、设备内存消耗严重的问题。为使本文提出的 EGF-MVSNet 多视图三维重建网络具有更高的使用价值,提高网络的重建完整度,提出了使用 SmoothL1 损失函数,使网络在训练后期收敛到更高的精度。使用 Adam 优化器配合改进的网络,提高了网络的鲁棒性和泛化能力。实验结果表明,本文提出的网络在内存消耗上明显低于其他网络,对比 MVSNet 网络在完整性上提高了 22.1%,模型总体评分提高了 11.5%。

在研究过程中发现,每种 MVS 网络都具有自己适用性较高的领域,这就使得网络应用起来较为麻烦。下一步会针对提高网络的泛化能力进行挖掘研究,通过优化、改进网络模型提高其多领域的三维重建,让研究成果具备更好的泛化性,从而让其适用价值更高。

参考文献

[1] 谢琪琦,辛月兰,曾曦. 基于注意力机制的多视图三维重建[J]. 激光杂志,2023,44(1):136-142.
XIE Q Q, XIN Y L, ZENG X. Multi view 3D reconstruction based on attention mechanism [J]. Laser Journal,2023,44(1):136-142.

[2] 颜深,张茂军,樊亚春,等. 大规模室外图像 3 维重建技术研究进展[J]. 中国图象图形学报,2021,26(6):1429-1449.
YAN SH, ZHANG M J, FAN Y CH, et al. Research progress in large-scale 3D reconstruction technology for outdoor images [J]. Chinese Journal of Image and Graphics, 2021,26(6):1429-1449.

[3] 李哲,卢健,杨腾飞. 基于深度学习的三维点云重建方法[J]. 国外电子测量技术,2021,40(3):1-5.
LI ZH, LU J, YANG T F. A 3D point cloud reconstruction method based on deep learning [J]. Foreign Electronic Measurement Technology, 2021,40(3):1-5.

[4] 谢亮亮,屠大维,张旭,等. 深海原位激光扫描双目立体视觉成像系统[J]. 仪器仪表学报,2020,41(6):106-114.
XIE L L, TU D W, ZHANG X, et al. Deep sea in situ laser scanning binocular stereo vision imaging system [J]. Chinese Journal of Scientific Instrument, 2020,41(6):106-114.

[5] GALLUP D, FRAHM J M, MORDOHAJ P, et al. Real-time plane-sweeping stereo with multiple sweeping directions [C]. 2007 IEEE Conference on Computer Vision and Pattern Recognition. IEEE, 2007: 1-8.

[6] YANG R, POLLEFEYS M. Multi-resolution real-time stereo on commodity graphics hardware [C]. 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. IEEE, 2003.

[7] SHU B, QIU X, WANG Z. Image based modeling via plane sweep based surface growing [C]. 2009 IEEE International Conference on Shape Modeling and Applications. IEEE, 2009: 15-20.

[8] 薛俊诗,易辉,吴止媛,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.
XUE J SH, YI H, WU ZH H, et al. A hybrid multi view 3D reconstruction method based on scene image segmentation [J]. Journal of Automation, 2020,46(4):782-795.

[9] LUO K, GUAN T, JU L, et al. P-mvsnet: Learning patch-wise matching confidence aggregation for multi-view stereo [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019: 10452-10461.

[10] 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.

[11] 李剑,陈宇航. 一种多视角高精度图片的深度估计方

- 法[J]. 北京邮电大学学报, 2021, 44(5): 101-106.
- LI J, CHEN Y H. A depth estimation method for multi perspective and high-precision images [J]. Journal of Beijing University of Posts and Telecommunications, 2021, 44(5): 101-106.
- [12] 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. WANG S Q, ZHANG J Q, LI L Y, et al. The application of MVSNet in 3D reconstruction of space objects [J]. China Laser, 2022, 49(23): 176-185.
- [13] YU Z, GAO S. Fast-mvsnet: Sparse-to-dense multi-view stereo with learned propagation and Gauss-Newton refinement [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 1949-1958.
- [14] MA X, GONG Y, WANG Q, et al. Epp-mvsnet: Epipolar-assembling based depth prediction for multi-view stereo [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 5732-5740.
- [15] 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.
- [16] YAO Y, LUO Z, LI S, et al. Recurrent MVSNet for high-resolution multi-view stereo depth inference [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 5525-5534.
- [17] YU A, GUO W, LIU B, et al. Attention aware cost volume pyramid based multi-view stereo network for 3D reconstruction [J]. ISPRS Journal of Photogrammetry and Remote Sensing, 2021, 175: 448-460.
- [18] WEI Z, ZHU Q, MIN C, et al. AA-RMVSNet: Adaptive aggregation recurrent multi-view stereo network [C]. Proceedings of the IEEE/CVF International Conference on Computer Vision, 2021: 6187-6196.
- [19] GAO S, LI Z, WANG Z. Cost volume pyramid network with multi-strategies range searching for multi-view stereo [C]. Advances in Computer Graphics: 39th Computer Graphics International Conference, CGI 2022, Virtual Event, September 12-16, 2022, Proceedings. Cham: Springer Nature Switzerland, 2023: 157-169.
- [20] GIANG K T, SONG S, JO S. Curvature-guided dynamic scale networks for multi-view stereo [J]. ArXiv preprint arXiv:2112.05999, 2021.
- [21] 陈亮, 张媛媛, 刘韵婷. 基于改进的 LSTM 的药品温湿度预测方法 [J]. 电子测量与仪器学报, 2019, 33(1): 106-112.
- CHEN L, ZHANG Y Y, LIU Y T. Drug temperature and humidity prediction method based on improved LSTM [J]. Journal of Electronic Measurement and Instrumentation, 2019, 33(1): 106-112.
- [22] 唐颖川, 黄姣茹, 钱富才. 基于深度学习与注意力机制的化工故障分类 [J]. 电子测量技术, 2022, 45(4): 168-174. TANG Y CH, HUANG J R, QIAN F C. Chemical fault classification based on deep learning and attention mechanism [J]. Electronic Measurement Technology, 2022, 45(4): 168-174.
- [23] GUPTA A, KUMAR S. A multiplicative Gauss-Newton minimization algorithm: Theory and application to exponential functions [J]. Applied Mathematics-A Journal of Chinese Universities, 2021, 36: 370-389.
- [24] YAO Y, LUO Z, LI S, et al. Mvsnet: Depth inference for unstructured multi-view stereo [C]. Proceedings of the European Conference on Computer Vision (ECCV), 2018: 767-783.
- [25] YI P, TANG S, YAO J. DDR-Net: Learning multi-stage multi-view stereo with dynamic depth range [J]. ArXiv preprint arXiv:2103.14275, 2021.
- [26] LI J, LU Z, WANG Y, et al. DS-MVSNet: Unsupervised multi-view stereo via depth synthesis [C]. Proceedings of the 30th ACM International Conference on Multimedia, 2022: 5593-5601.
- [27] ZHANG J, YAO Y, LI S, et al. Visibility-aware multi-view stereo network [J]. ArXiv preprint arXiv: 2008.07928, 2020.
- [28] YANG J, MAO W, ALVAREZ J M, et al. Cost volume pyramid based depth inference for multi-view stereo [C]. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020: 4877-4886.

作者简介



刘韵婷(通信作者), 博士, 副教授, 现为沈阳理工大学自动化与电气工程学院硕士生导师, 主要研究方向为无线传感器网络、人工智能技术和数据分析等。

E-mail: liuyunting0224@163.com

Liu Yunting (Corresponding author), received a Ph. D. degree. She is now an associate professor and a master's supervisor at the School of Automation and Electrical Engineering, Shenyang University of Technology. Her main research interests include wireless sensor networks, artificial intelligence technology, and data analysis, etc.