中图法分类号:TP391.41 文献标识码: A 文章编号: 1006-8961(2024)05-1346-18

论文引用格式:Li Y and Wu X Q. 2024. Survey of multilevel feature extraction methods for RGB-D images. Journal of Image and Graphics, 29(05): 1346-1363(李洋,吴晓群. 2024. 面向 RGB-D 图像的多层特征提取算法综述. 中国图象图形学报, 29(05): 1346-1363)[DOI: 10.11834/jig. 230351]

# 面向RGB-D图像的多层特征提取算法综述

## 李洋,吴晓群\*

1. 北京工商大学计算机学院,北京 100048; 2. 食品安全大数据技术北京市重点实验室,北京 100048

摘 要: RCB-D图像包含丰富的多层特征,如底层的线特征、平面特征,高层的语义特征,面向 RCB-D 图像的多层特 征提取结果可以作为先验知识提升室内场景重建、SLAM(simultaneous localization and mapping)等多种任务的输出 质量,是计算机图形学领域的热点研究内容之一。传统的多层特征提取算法一般利用 RGB 图像中丰富的颜色、纹 理信息以及深度图像中的几何信息提取多层特征,此类提取算法依赖输入 RGB-D 图像的质量,而受采集过程中环 境和人为因素的影响,很难得到高质量的 RGB-D 图像。随着深度学习技术的快速发展,基于深度学习的多层特征 提取算法突破了这一限制,涌现出一批高质量的研究成果。本文对面向 RGB-D 图像的多层特征提取算法进行综 述。首先,汇总了现有的常用于多层特征提取任务的 RGB-D 数据集和相关算法的质量评价指标。然后,按照特征 所处的不同层次,依次对线、平面和语义特征相关算法进行了总结。此外,本文还对各算法的优缺点进行比较并结 合常用算法质量评价标准进行了定量分析。最后,讨论了当前多层特征提取算法亟待解决的问题并展望了未来发 展的趋势。

关键词: RGB-D图像;多层特征;线特征;平面特征;语义特征;特征提取

### Survey of multilevel feature extraction methods for RGB-D images

Li Yang, Wu Xiaoqun<sup>\*</sup>

School of Computer Science and Engineering, Beijing Technology and Business University, Beijing 100048, China;
 Beijing Key Laboratory of Big Data Technology for Food Safety, Beijing 100048, China

**Abstract:** RGB-D images contain rich multilevel features, such as low-level line, planar, and high-level semantic features. These different levels of features provide valuable information for various computer vision tasks. Computer vision algorithms can extract meaningful information from RGB-D images and improve the performance of various tasks, including object detection, tracking, and indoor scene reconstruction, by leveraging these multilevel features. Terms such as feature and contour lines can be used when describing existing line features in a single RGB-D image. Line features provide crucial information regarding the spatial relationships and boundaries in the input image, aiding in the understanding and interpretation of input data. Plane and surface are used to describe planar features and those refer to flat or nearly flat regions in the RGB-D image. Terms such as instance and semantic labels can be used when describing an object. Instance labels refer to unique identifiers or labels assigned to individual instances or occurrences of objects in an image, while semantic labels represent the broad class or category to which an object belongs. Semantic labels provide a high-level understanding of the objects in the image, grouping them into meaningful categories that indicate the general type of object present. Traditional

收稿日期:2023-06-18;修回日期:2023-09-20;预印本日期:2023-09-27

<sup>\*</sup>通信作者:吴晓群 wuxiaoqun@btbu.edu.cn

基金项目:国家自然科学基金项目(62272014)

Supported by: National Natural Science Foundation of China (62272014)

methods for extracting line features often utilize color, texture information of RGB image, and geometric information in the depth image to extract feature and contour lines. The extraction of planar features involves clustering to extract sets of points with similar properties, further facilitating planar feature extraction. Semantic feature extraction aims to assign specific semantic categories to each pixel in the RGB-D input, and most of the methods used for this task are implemented based on deep learning. The multilevel feature extraction results for RGB-D images can be used as prior knowledge aids such as indoor scene reconstruction, scene understanding, object recognition, and other tasks to improve the quality of network output. Multilevel feature extraction for RGB-D images is also one of the popular topics in the field of computer graphics. With the development and popularization of commercial depth cameras, acquiring RGB-D data has become increasingly convenient. However, the quality of captured RGB-D data is often compromised by environmental and human factors during the acquisition process. This phenomenon leads to issues such as noise and depth absence, which, in turn, negatively affects the quality of multilevel feature extraction results to some extent. These problems are detrimental to traditional methods, but the emergence of deep learning approaches has overcome these issues to a certain extent. With the rapid development of deep learning technology, numerous high-quality research results have emerged for multilevel feature extraction tasks based on deep learning. The commonly used RGB-D datasets for multilevel feature extraction tasks, such as NYU v2 and SUN RGB-D, are summarized in this paper. These datasets contain diverse scene data, comprising RGB images paired with corresponding depth images. Taking NYU v2 as an example, the dataset includes 1 499 RGB-D images, derived from 464 distinct indoor scenes across 26 scene classes. After introducing the datasets, this paper provides a summary of commonly used evaluation criteria for assessing the quality of line, planar, and semantic features. Detailed explanations are presented for the computation method of each evaluation criterion. When reviewing line feature extraction methods, a comprehensive summary based on traditional and deep learning approaches is presented. Detailed explanations of the principles, advantages, and limitations of different methods are provided. Furthermore, quantitative comparisons of the extraction results from several different methods are conducted. When summarizing planar feature extraction methods, a comprehensive overview is provided from two perspectives: traditional and deep learning-based planar feature extraction methods. Relevant research papers are gathered, and a quality comparison of planar feature extraction methods is then conducted. Additionally, detailed explanations of the advantages and limitations of each method are provided. A comprehensive review of deep learning-based semantic feature extraction methods is presented in this paper from two aspects: fully-supervised and semi-supervised learning-based semantic feature extraction methods. Relevant research papers are also summarized. When comparing different semantic feature extraction methods, this paper used evaluation metrics such as pixel accuracy (PA), mean PA (MPA), and mean intersection over union (mIoU) to measure the quality of the extraction algorithms. The results of the quantitative comparisons revealed that semantic feature extraction methods oriented toward RGB-D data exhibit superior extraction quality. These comparison results prove that feature extraction methods designed specifically for RGB-D data can achieve better results compared to methods that only utilize RGB data. The incorporation of depth information in RGB-D data facilitates accurate and robust extraction of semantic features, leading to enhanced performance in various tasks such as scene understanding and object recognition. Data annotation has certainly been a challenge for feature extraction methods based on deep learning. Annotating large-scale datasets requires considerable time and human resources. Researchers have been actively seeking ways to reduce the workload of data annotation or maximize existing annotated data to overcome these challenges. Therefore, unsupervised, semi-supervised, and transfer learning are widely investigated to leverage unlabeled or sparsely labeled data for feature extraction. Finally, the problems of the current multilevel feature extraction algorithm that must be addressed are discussed to provide guidance to the future development trend at the end of this paper.

Key words: RGB-D images; multilevel features; line features; planar features; semantic features; feature extraction

### 0 引 言

RGB-D图像包含各种不同的特征,这些特征大 致可以分为以下几类:基于点/像素的特征描述,如 颜色、位置、方向、梯度和曲率等;基于线的特征描 述,如特征线、轮廓线等;基于面/区域的描述,如平 面、曲面等;基于物体的描述,如实例标签、语义标签 等。本文调研的线特征、平面特征和语义特征从线、 面和物体3个层次描述RGB-D图像特征。从RGB-D 图像中提取高质量的线、平面和语义特征等,为后续 RGB-D的图像配准、重建等提供更准确的先验信 息,从而更好地应用于后续自动驾驶(Chen等, 2018b) SLAM (simultaneous localization and mapping)(Yang等, 2020)以及三维场景重建(Dzitsiuk 等,2017)等应用中。同时,如何从RGB-D图像中提 取高质量的线、平面和语义特征也是当前计算机图 形学领域中的热点研究问题之一。因此,本文深入 调研了面向RGB-D图像的线、平面和语义特征提取 方法。受限于商业深度相机的采集精度,RGB-D图 像中的深度图像分辨率较低,同时,受采集过程中环 境和人为因素的影响, RGB-D图像往往存在噪声、 缺失等问题。这导致在具体的应用,例如三维场景 重建中,重建的三维场景存在缺失、裂缝和局部精度 低等问题。因此,提取高质量RGB-D图像的多层特 征,合理利用先验知识对于提高三维场景重建等任 务的质量显得尤为重要。

主流的基于结构光、双目视觉或时间飞行法 (time of flight,TOF)的商业深度相机,受现实场景中 物体间遮挡、物体材质等因素的影响,无法获取理论 上的高精度深度图像。相对于低分辨率、低精度的 深度图像,RGB图像的质量高,因此,大多数线特征 提取算法,多数平面、语义特征提取算法是面向 RGB图像的。RGB图像中包含丰富的自然特征,如 亮度、颜色和纹理等,面向RGB图像的特征提取算 法通常利用这些自然特征提取颜色、纹理差异较大 区域的线特征,颜色、纹理相似区域的平面特征和高 层语义特征。RGB图像中缺乏几何信息,特征提取 时,面向颜色、纹理整体差异较小的场景无法准确提 取线、平面和语义特征。深度图像包含物体在一个 方向上的深度信息,能够体现物体间的几何差异,为 了提高特征提取质量,将深度图像作为算法的补充 信息十分必要。

如何从 RGB-D 图像中提取高质量的多层特征 已成为近年来的一个研究热点。早期,张茗奕等人 (2017)从全局拟合和特征点生长两个角度对 RGB 图像的线特征提取算法进行了总结;田萱等人 (2019)从基于区域分类和基于像素分类两个方向对 基于深度学习的图像语义分割方法进行了全面综 述;现有的文献中,罕有面向 RGB-D 图像相关的总 结文献。本文阐述了面向 RGB-D 图像的多层特征 提取任务的重要性,对现有的面向 RGB-D 图像的 线、平面和语义特征算法进行了综述。

本文旨在对面向 RGB-D 图像的多层特征提取 算法的研究现状进行归纳总结,图1展示了来自 NYU v2(New York University)(Silberman等,2012)数 据集中的 RGB-D 图像以及对应的线、平面和语义特 征的可视化结果。本文汇总了目前常用的 RGB-D 数据集,对数据集的采集年份、数据的分类和数量等 进行了详细说明,并总结了多层特征提取算法相关 的评价指标。在总结面向 RGB-D 图像的线、平面和 语义特征提取算法时,对各算法的优缺点进行了对 比说明,同时按照总结的评价指标进行了定量分析。 最后,本文总结了当前领域亟待解决的问题,并展望 了未来发展的方向。



对应的多层特征

Fig. 1 RGB-D image from NYU v2(Silberman et al. ,2012) and corresponding multi-level features

((a)RGB image;(b)depth image;(c)line feature; (d)planar feature;(e)semantic feature)

## 1 RGB-D 数据集和特定任务评价 指标

评价多层特征提取任务中各算法的性能时,不 仅需要统一、科学的评价标准,还需要标准的RGB-D数据集。本节调研了多个可用于多层特征提取任 务的RGB-D数据集并汇总了特定任务的相关评价 标准。

### 1.1 RGB-D数据集

本节对几种常见的 RGB-D 数据集进行了详细 介绍,并在表1中汇总了可用于多层特征提取任务 的多个 RGB-D 数据集和下载地址。

#### 1.1.1 NYU v2数据集

NYU v2(Silberman等,2012)数据集是多个室内 场景视频序列的合集,由详细标注的数据子集、原始 数据集和处理数据标注的工具箱组成。标注数据子 集中包含1449幅RGB-D图像和对应的标注图像。

### 1.1.2 SUN RGB-D数据集

SUN RGB-D(Song等, 2015)数据集是由4个不

同深度相机采集 10 335 幅 RGB-D 图像得到的数据 集,这个密集标注的数据集包括 146 617个 LabelMe 风格的 2D 多边形注释,64 595个具有精确对象方向 的 3D 边界框。

1.1.3 ScanNet数据集

ScanNet(Dai等,2017)是一个 RGB-D 视频数据 集,ScanNet 包含 21 个场景类型、1 513 个场景,有 250 万幅图像,其中 1 201 个场景常作为训练数据, 其余的 312 个场景数据常作为测试数据。

#### 1.2 特定任务评价指标

评估算法性能时,需要使用客观评价指标以保 证评估结果的有效性,常见的有算法运行时间、内存 消耗以及精度等,其中精度是关键的评价指标。本 节汇总了多层特征提取算法在精度上常用的评价 指标。

1.2.1 线特征提取任务的评价指标

提取线特征时,常用精度(precision,P)、召回率 (recall,R)和交并比(intersection over union, IoU)评 估提取算法质量。P用于评估算法检测的正例样本 的准确性,统计了线特征提取结果中正确数量所占

名称	年份	分辨率/像素	说明
RGB-D Object Dataset (Lai 等, 2011)	2011	640 × 480	包含51个类、300个不同对象,25万幅RGB-D图像 (http://rgbd-dataset.cs.washington.edu/dataset)
3D Object Dataset(Janoch等,2011)	2011	640 × 480	包含超过50个不同对象,取自75个场景的849幅RGB-D图像 (http://www.kinectdata.com)
NYU v1(Silberman和Fergus,2011)	2011	640 × 480	取自7个场景类别、64个不同室内场景,有2347幅详细标注的RGB-D 图像(https://cs.nyu.edu/~silberman/datasets/nyu_depth_v1.html)
NYU v2(Silberman等,2012)	2012	640 × 480	取自26个场景类别、464个不同室内场景,有1449幅详细标注的 RGB-D图像(https://cs.nyu.edu/~silberman/datasets/nyu_depth_v2.html)
TUM RGB-D(Sturm等,2012)	2012	640 × 480	包含两个室内环境中采集的39个RGB-D序列和传感器真实运动轨迹 数据(https://cvg.cit.tum.de/data/datasets/rgbd-dataset/download)
SUN3D(Xiao等,2013)	2013	640 × 480	取自41座建筑、254个不同空间,有415个RGB-D序列,其中有8个详 细标注序列(https://sun3d.cs.princeton.edu)
SUN RGB-D(Song等,2015)	2015	不固定	使用4种深度图像采集设备,取自47个场景类别,有10335幅RGB-D 图像(https://rgbd.cs.princeton.edu/challenge.html)
SceneNet RGB-D (McCormac 等, 2017)	2017	320 × 240	取自16895个室内场景,有500万幅标注的RGB-D图像 (https://robotvault.bitbucket.io/scenenet-rgbd.html)
AVD(Ammirato等,2017)	2017	不固定	取自9个独立场景,有20916幅RGB-D图像以及对33个常见实例的 54247个边框标注数据(https://paperswithcode.com/dataset/avd)
ScanNet(Dai等,2017)	2017	640 × 480	取自707个不同空间、1513个室内场景,有250万幅标注的RGB-D数据(https://github.com/ScanNet/ScanNet)

表1 RGB-D数据集 Table 1 RGB-D dataset

比例;R用于评估算法对真实值(ground truth,GT)的 检测覆盖率,统计了线特征提取结果中正确数量占 GT的比例;IoU用于评估算法检测结果和GT的重合 程度,是线特征提取结果和GT的交集与并集的 比例。

1.2.2 像素级特征提取任务的评价指标

平面、语义特征提取作为像素级特征提取任务, 常用的质量评价指标有像素准确率(pixel accuracy, PA)、平均像素准确率(mean pixel accuracy, MPA)、 IoU、平均交并比(mean intersection over union, mIoU)、P、R、信息变化度量(variation of information, VI)、Rand指数(rand index, RI)和分割覆盖率(segmentation covering, SC)。

PA用于评估预测结果的像素准确率,是平面或 语义特征提取结果中准确提取的像素占总像素的比 值;IoU用于评估平面或语义特征提取结果和GT的 重合程度,是提取结果和GT的交集与并集的比值; MPA和mIoU是在PA和IoU的基础上推导产生的, 通过独立计算某平面(或某类语义)特征的PA或IoU 值后再取均值得到;P是平面(或语义)特征结果中 正确提取像素数量占像素总数的比例;R是平面(或 语义)特征正确提取像素数量占GT的比例;VI是评 估平面特征提取结果和GT之间差异的指标;RI与P 类似,也是评估平面特征提取结果和GT一致程度的 指标;SC是评估平面特征提取结果对GT的覆盖程 度指标。

PA的计算方法为

$$f_{\rm PA} = \frac{\sum_{i=1}^{N} TPN_i}{\sum_{i=1}^{N} PN_i}$$
(1)

式中,*PN*<sub>i</sub>为第*i*个平面(或第*i*类语义)中像素的真实 数量,*TPN*<sub>i</sub>为第*i*个平面(或第*i*类语义)中被正确预 测的像素数量。

MPA 和 mIoU 的计算方法为

$$f_{\rm MPA} = \frac{\sum_{i=1}^{N} \frac{TPN_i}{PN_i}}{N}$$
(2)

$$f_{\rm mlou} = \frac{\sum_{i=1}^{n} \frac{IPN_i}{PN_i + FPN_i}}{N}$$
(3)

式中,N为图像中平面(或语义)特征数量,PN<sub>i</sub>和 TPN<sub>i</sub>与式(1)中含义一致,FPN<sub>i</sub>为被错误预测为第*i*  个平面(或第i类语义)的像素数量。

VI的计算方法为

$$f_{\rm vI} = H(\boldsymbol{R}) + H(\boldsymbol{G}) - 2MI(\boldsymbol{R},\boldsymbol{G}) \tag{4}$$

式中,H(R)是平面提取结果的熵,H(G)是GT的熵, MI(R,G)是平面提取结果和GT之间的互信息 (mutual information)。

RI的计算方法为

$$f_{\rm RI} = \frac{TP + TN}{TP + FP + FN + TN} \tag{5}$$

式中,*TP*(true positive)表示正确提取的线特征数量, *FP*(false positive)表示错误提取的线特征数量,*FN* (false negative)表示未被提取到的线特征数量。*TN* (true negative)表示不属于该平面特征的像素数量。

SC的计算方法为

$$f_{\rm SC}(S',S) = \frac{1}{N} \sum_{R \in S} |R| \times \max_{R' \in S'} \frac{|R \cap R'|}{|R \cup R'|} \qquad (6)$$

式中,**R**′是实际提取结果S′中的某个平面,R是真实 结果S(即GT)中与R′对应的平面,N是像素的总数 量,|**R**|是平面中的像素个数。

## 2 面向 RGB-D 图像的线特征提取 方法

线特征是RGB-D图像中的重要特征之一,常用 于描述RGB-D输入中的几何、纹理差异等信息,作 为图像的底层特征在车道线检测、场景重建等领域 都有广泛应用。张茗奕等人(2017)对RGB图像的 线特征提取算法已经进行了总结,本节不再重复综 述。本节将从传统和基于深度学习两个角度对面向 RGB-D图像的线特征提取方法进行汇总。

### 2.1 传统的线特征提取方法

使用传统的方法提取线特征时,大多数方法仅利用了RGB-D中RGB图像丰富的颜色、纹理差异信息,仅有少部分方法同时利用RGB-D中深度图像丰富的几何信息。

本节从面向2D和2.5D数据两个角度对传统的 线特征提取方法进行总结。鉴于张茗奕等人(2017) 对RGB图像的线特征提取算法已经进行了综述,本 文对面向2D数据的线特征提取方法进行总结时,在 忽略深度图像本身几何意义的前提下,将这些方法 应用在深度图像上,仅利用深度图像中的颜色、纹理 差异提取线特征。 2.1.1 面向2D数据的线特征提取方法

面向 2D 数据的线特征提取方法本质上是将 RGB-D 图像中的深度图像作为 2D 图像处理,提取深 度图像中纹理差异大的边界。

von Gioi等人(2010)提出的LSD(line segment detector)算法可以将单通道的深度图像作为2D灰 度图像进行处理,计算灰度梯度值和梯度方向后,通 过筛选两者差异小的同类像素点并输出线段支撑区 域,对线段支撑区域进一步处理即可得到特征线。 LSD算法对噪声较为敏感,深度噪声较大时,同一条 特征线容易被过度分割为多条特征线。

为了提取更加完整的特征线,Cho等人(2018) 提出了新的线特征提取算法Linelet,计算像素梯度 值之后,Linelet在水平和垂直方向对非极大梯度值 像素进行抑制。划分线段合集时在两个方向对具有 相同属性且距离较近的线段划分线段合集,通过概 率模型估计每条直线的角度后结合梯度方向和梯度 值进一步筛选提取结果。为了提高特征线的鲁棒 性,Linelet在最后会聚合可能来自同一条直线的 线段。

为了提升特征线的提取速度同时降低噪声对结 果的影响,Akinlar和Topal(2011)提出了EDLines (edge drawing lines)算法。EDLines计算速度是LSD 算法的11倍,抗噪能力也有一定提升。EDLines性 能提升的关键点在于锚点的设置,锚点是一定范围 内梯度最大的边界点,锚点的设置降低了点的个数 也降低了噪声的影响。在锚点点集中提取特征线 时,EDLines结合水平和垂直方向梯度值确定锚点 之间的连接关系得到像素链,在像素链中进一步提取特征线。Zhang等人(2021)提出的AG3line同样是依据梯度值和梯度方向计算锚点,与EDLines不同,AG3line通过梯度方向和线的几何形状初步提取特征线,特征线优化阶段利用梯度值和特征线的几何形状进一步优化结果。

与 EDLines、AG3line 中锚点的设置相似, Nieto 等人(2011)提出的 LSWMS(line segment detection using weighted mean shift)使用参数方程估计像素点 的值并结合切片采样算法选取一定邻域内梯度显 著、方向一致的像素点作为候选点,这种采样策略同 样提升了运行速度。

线特征提取时,合理设置阈值是保证提取结果 的关键,Lu等人(2015)提出能自适应设置 Canny (Canny,1986)高、低阈值的线特征提取算法 Canny-Lines,其通过自适应阈值的 Canny 算法鲁棒地提取 边缘图,从边缘图中具有最大梯度值的像素点开始 结合梯度方向连接同类点获取较短的初始线段,随 后在初始线段的水平和垂直方向上迭代地合并同方 向线段,直至输出特征线。

然而,由于没有合理利用深度图像中的深度值, 在面对深度值差异小的场景时,2D深度图像相邻像 素间梯度差异小,无法保证提取结果质量。图2展 示几种线特征提取算法在NYU v2数据集上的提取 结果。

2.1.2 面向 2.5D 数据的线特征提取方法

与面向2D数据的线特征提取方法不同,面向 2.5D数据的线特征提取方法充分利用了深度图像





Fig. 2 Line feature extraction results of depth images ((a) RGB images; (b) depth images;(c) ground truth; (d) LSD; (e) EDLines; (f) CannyLines; (g) AG3line)

的几何差异。Bose和Richards(2016)提出的快速深度边界检测(fast depth edge detection, FDED)算法就利用了深度图像中的几何差异提取线特征,但是FDED没有利用RGB-D图像中的RGB信息。

与FDED不同,Yang等人(2015)算法先利用深 度突变提取遮挡区域,再利用法向突变提取折叠区 域,随后结合 RGB 图像的 Canny 检测结果提取遮挡 区域的遮挡边和折叠区域的折叠边。同时,为了提 高提取精度,该算法保留了遮挡区域和折叠区域中 足够大且细长的区域。保留的这些区域几何差异 大,大概率是线特征潜在区域,但颜色、纹理差异可 能较小导致 Canny 无法提取出边界信息。最后,在 遮挡边和折叠边中分别使用聚类算法优化提取线特 征结果,该算法充分利用了 RGB-D 中的颜色、纹理 和几何信息。

Cao等人(2017)提出的SFLine算法也是典型的 面向2.5D数据的线特征提取方法。SFLine先利用 RGB-D图像的颜色、纹理差异结合网格边缘检测器 提取强度边,随后利用深度图像计算深度法向,再结 合深度法向筛选几何强度边,并在强度边所在网格 内依据点到强度边的距离和法向选取3D特征点,最 后筛选同类特征点并拟合特征线。SFLine同样充分 利用了RGB-D中的颜色、纹理和几何信息。

此外,Choi等人(2013)还将 RGB-D 图像转换为 点云后提取线特征。在3D点云中,该算法通过点的 深度连续性提取物体间有遮挡关系的深度不连续区 域的线特征;通过计算曲率获取高曲率的几何边缘 线特征;此外,该算法还保留了点云的边界作为点云 边界线特征;最后,该算法使用 Canny 算子提取 RGB 图像的 2D 线特征,将最终结果投影并保留到 3D 空 间中。近些年,Lu等人(2019)和 Hu等人(2022)则 直接在点云上提取线特征,省略了 RGB-D 图像到点 云数据的转换过程。

### 2.2 基于深度学习的线特征提取方法

随着深度学习在计算机图形学领域的广泛应 用,结合深度学习的线特征提取方法也随之兴起。

现有的基于深度学习的线特征提取方法多是将 RGB图像作为算法输入,如Xue等人(2019)提出基 于U-Net(Ronneberger等,2015)网络的改进算法、 Zhang等人(2019)提出深度提取网络PPGNet(pointpair graph network)以及Xu等人(2021)提出基于编 码一解码结构的线特征提取网络LETR(line segment Transformer)。Xue等人(2019)网络结构结合 了DeepLab v3+(Chen等,2018a)中带孔空间金字塔 池化(atrous spatial pyramid pooling,ASPP)模块和 ResNet(He等,2016)的跳跃连接模块;PPGNet网络 创新性地将特征线提取问题转化为图推理问题,检 测模块检测所有节点后在这些节点上推理、识别节 点间连通性;LETR 网络采用了由粗到细的检测策 略,在粗解码阶段仅关注特征线的潜在区域,在细解 码阶段通过多尺度编码特征的交互提取细节信息。

从结果来看,无论是Xue等人(2019)提出的提取网络、PPGNet网络还是LETR网络都不能有效区分几何边界和纹理边界。事实上这是网络输入数据维度的局限性导致的必然结果,几何边界信息来源于深度图像,缺乏深度图像输入的网络只能依赖于纹理差异提取特征线,这种情况下无法做到有效区分几何边界和纹理边界。

### 2.3 线特征提取方法对比

表2对面向2D数据的线特征提取方法在NYU v2数据集上的表现进行了定量比较,其中结果参考 的真实值为人工标注;表3对传统的线特征提取方 法中各类方法实现原理、优缺点进行了对比说明; 图3展示了几种传统方法在2D数据和2.5D数据上 的线特征提取结果,其中的真实值同样是人工标注的 数据。从图3的结果来看,传统方法中,以RCB-D图

Table 2	Quantitative comparison of	several line leature	extraction meth	ious with 2D m	Jui
算法	年份	输入	Р	R	IoU
LSD(von Gioi等,2010)	2010	深度图像	0.46	0.19	0.17
EDLines(Akinlar和Topal,	2011) 2011	深度图像	0.58	0.20	0.18
AG3line(Zhang等,2021)	2021	深度图像	0.51	0.21	0.18
CannyLines(Lu等,2015)	2015	深度图像	0.66	0.42	0.34

表 2 几种面向 2D 数据的线特征提取方法定量对比 2 Opentitative comparison of several line feature extraction methods with 2D input

注:加粗字体表示各列最优结果。

			1		
算法	年份	输入	原理	优点	缺点
LSD(von Gioi等,2010)	2010	2D数据	利用像素的梯度值和梯度 方向进行区域生长	提取精确,有效区分边界	检测速度慢,抗噪能力 弱,线段容易被过度分 割,鲁棒性低
Linelet(Cho等,2018)	2018	2D数据	通过连接水平和垂直方向 的极大值点绘制边缘	边缘明显,层次分明场景 检测效果好	运行速度慢,曲线场景 提取效果差
EDLines(Akinlar 和 Topal, 2011)	2011	2D数据	通过筛选局部极值点后绘 制边缘	检测速度快,具有良好的 抗噪能力	局部极值筛选机制导致 梯度值小的区域检测效 果不佳
AG3line(Zhang等,2021)	2021	2D数据	通过梯度方向和线的几何 形状提取特征线	鲁棒性强,运行速度快	会错误检测出假阳性 结果
LSWMS(Nieto等,2011)	2011	2D数据	使用改进的 Mean-Shift 算 法快速提取直线	能够降低输入数据中的噪 音的影响	鲁棒性较差
CannyLines(Lu等,2015)	2015	2D数据	通过自适应调整Canny的 高低阈值提取边界	提取结果鲁棒性强	部分结果检测不完全
Yang等人(2015)	2015	2.5D数据	结合深度、法向差异和RGB 纹理差异提取边界	有一定抗噪能力	几何差异不明显的场景 提取效果差
SFLine(Cao等,2017)	2017	2.5D数据	利用深度几何差异提取尖 锐的边缘特征	对于几何特征尖锐的场景 检测效果好	几何差异不明显的场景 提取效果差
PPGNet(Zhang等,2019)	2019	2D数据	通过图推理得到特征点的 连接性绘制线段	提出的图推理框架普适 性强	推理连接关系模块算法 时间复杂度高
LETR(Xu等,2021)	2021	2D数据	通过由粗到细检测模块提 取直线	设置细解码网络架构补充 细节信息	不能有效筛选伪线段 结果

表 3 线特征提取方法对比 Table 3 Comparison of line feature extraction methods



(c) RGB result of EDLines; (d) RGB result of CannyLines; (e) RGB result of AG3line;

(f) depth input; (g) depth result of LSD; (h) depth result of EDLines; (i) depth result of CannyLines;

(j) depth result of AG3line; (k) ground truth; (l) result of SFLine (Cao et al. ,2017))

像作为输入的SFLine线特征提取结果明显优于LSD 等方法。这样的结果是意料之中的,SFLine合理利用 了RGB-D图像中的颜色、纹理和几何信息,能够筛选 和提取颜色、纹理差异大且几何差异明显的线特征; 而LSD等方法仅利用了RGB图像中的信息,且在深 度图像上运用这些方法时忽略了深度图像本身的几 何意义,在这种情况下输入数据的信息量少、维度低, 因此无法有效提取接近真实值的结果。

### 3 平面特征提取方法

平面特征是 RGB-D 图像的底层特征之一,本节 从传统和基于深度学习两个角度调研了常用的平面 特征提取方法。

### 3.1 传统平面特征提取方法

传统面向 RGB-D 图像的平面特征提取任务关 键点在于如何利用 RGB-D 图像的纹理信息和深度 信息。在平面内纹理过于复杂的情况下,利用纹理 信息提取平面会导致同一个平面被分割成多个,提 取结果的完整性得不到保障。利用深度信息提取平 面能有效避免这一问题,但是受深度图像精度和深 度噪声的影响,在平面交界区域分割精度依旧不高, 平面结果的边缘粗糙。这种情况下,如何利用纹理 信息处理粗糙边缘是提高提取结果精度的关键。

本节将传统平面特征提取方法分为基于区域增长的方法和基于随机抽样一致的方法,并在表4中对各算法的相关原理和优缺点进行了汇总。 3.1.1 基于区域增长的方法

区域增长方法的核心思想是将具有相似性质的 像素点合并为一类。基于区域增长算法提取平面的 关键在于种子点的选择策略,一旦指定种子点后, 种子点将作为增长起点,与邻域内其他像素点对比 并作为合并参考的标准点。性质相似的像素点合 并后会继续向外围生长,直到无法合并更多像素点 为止。

Jin 等人(2019)提出的平面特征提取算法 DPD (depth-driven plane detection)就是基于区域增长实现的,DPD算法包含基于区域增长的平面提取和一 个两阶段的优化策略两部分。提取平面时,DPD选 取深度图像中没有深度缺失的区域结合线性最小二 乘法拟合平面,选取种子点时,平面中点到平面距离 的误差是重要的参考标准。对于提取平面遇到的过 度增长和增长不足的情况,DPD算法在优化阶段会 重新分配平面相邻的区域至综合误差最小的平面。 与DPD算法类似,Han等人(2021)提出的算法同样 对相邻平面交界区域进行了二次优化,以解决过度 分割问题。

RGB-D图像包含深度信息,通过深度相机参数 可以转换RGB-D图像的表达形式,将RGB-D图像转 换为点云数据,随后在点云上提取平面特征,提取结 果同样可以通过深度相机参数映射到2D平面。 Roychoudhury等人(2021)提出的基于区域增长的实 时平面提取算法就是面向点云数据的,该算法选取 曲率连续、与四邻域点法向差异小的点作为种子点, 增长时目标点到平面间距离误差、种子点和目标点 间的法向误差都是重要的合并评判标准。

基于区域增长的平面提取方法实现简单,无需 人工干预处理流程,但仍存在不足。深度噪声在图 像采集过程中不可避免,而区域增长算法对噪声比 较敏感,这会导致目标平面被过度分割或者分割不 足;此外,区域增长算法对参数同样敏感,种子点的 选择、阈值的设置都会影响提取结果。

3.1.2 基于随机抽样一致的方法

随机抽样一致(random sample consensus, RANSAC)算法的基本思想是从数据中随机选择一 些样本,然后用这些样本来拟合平面。初步拟合平 面后,RANSAC计算数据集中每一个点到平面的距 离,将距离小于给定阈值的点标记为内点,反之标记 为外点,内点数量是评估平面质量的重要依据,内点 数量达到一定的阈值时,就认为平面拟合效果良好。 RANSAC作为一种经典的参数估计算法,广泛应用 于计算机图形学领域。

Silberman 等人(2012)提出的面向 RGB-D 图像 的 3D 平面提取算法首先对深度图像中的点沿着水 平和垂直双方向进行采样,然后再运用 RANSAC 算 法拟合平面。对平面二次优化的过程中,点的深度 法向与平面法向差异、点到平面的距离和点的 RGB 梯度值都是重要的参考指标。

Li 等人(2017)、Yue 等人(2018)和 Xu 等人 (2019)提出的算法都是在 RGB-D 图像转换之后的 点云数据上实现的。Li 等人(2017)提出的基于正态 分布变换(normal distributions transform, NDT)单元 改进的 RANSAC 算法,每次迭代过程中, NDT 单元都 作为一个最小的采样样本,确保采样的正确性。 Yue 等人(2018)和Xu 等人(2019)也都在RANSAC 的基础上做了不同程度的改进。

RANSAC有一定的抗噪能力,能够处理包含噪 声和异常值的数据,但是RANSAC算法十分依赖阈 值和迭代次数的设置,算法计算量也比较大。

### 3.2 基于深度学习的平面特征提取方法

随着深度学习技术在各个领域的广泛应用,利 用深度学习框架提取平面特征也取得一定的研究成 果。PlaneNet(Liu 等, 2018)和PlaneRCNN(Liu 等, 2019)都是基于深度学习的平面特征提取方法。 PlaneNet 是基于深度神经网络(deep neural network, DNN)实现的, PlaneNet依据RGB输入率先推断一组 平面参数,然后再结合平面分割掩码和预测的深度 图像完成平面提取。PlaneRCNN则是基于 Mask R-CNN (region convolutional neural network) (He 等, 2017)实现的, PlaneRCNN的输入也是单幅RGB图 像,PlaneRCNN 同样会预测深度图像并依据深度图 像辅助实现平面特征提取。此外, PlaneAE(planar reconstruction via associative embedding) (Yu 等, 2019)将关联嵌入的思想应用到了平面特征提取中; PlaneTR (plane recovery with structure-guided Transformer)(Tan等,2021)在提取平面特征时创新性地 使用了线特征先验信息指明平面的结构,这些工作 都取得了较好的结果。

事实上,单幅RGB图像的信息量是有限的,Xie 等人(2022)提出的 PlanarRecon 网络和 Liu 等人 (2022)提出的 PlaneMVS (planer reconstruction from multi-view stereo)网络都拓宽了网络输入,PlanarRecon 融合多帧的分割碎片实现全局平面重建,而 PlaneMVS则利用多视图之间的几何优势实现平面 提取。如PlaneNet、PlaneRCNN等运用深度学习的 方式预测深度图像是有误差的,将深度图像直接作 为网络输入可以减小这种误差。图4展示了 PlaneRCNN(Liu等,2019)和PlaneMVS(Liu等,2022) 的提取结果,可以明显看到,红框中标出部分 PlaneMVS 比 PlaneRCNN 的提取结果更好。无论是 PlanarRecon 中对多帧分割碎片的融合还是 PlaneMVS中对多视角数据和对应相机姿态的合理 利用都是为了补充输入数据的信息量。RGB-D图 像信息量丰富,既包含丰富的纹理信息又包含几何 信息,未来面向RGB-D图像的深度平面提取方法将 成为研究热点。

基于深度学习的平面特征提取模型经过大量数 据训练之后通常具有一定的泛化能力,能够在现实 世界复杂环境下的未知场景中提取平面特征,对于 RGB-D输入数据中存在的噪声、缺失等问题有一定 的容忍度;而传统方法则非常依赖 RGB-D输入的质 量,噪声、缺失等问题都可能导致出现分割不足或过 度分割的问题。因此,在面向复杂输入数据的情况 下基于深度学习方法的表现更胜一筹。此外,如 PlanarRecon等基于深度学习的方法能够做到实时 提取,而如基于随机抽样一致的传统方法则随着迭 代次数的增加效率变得非常低。然而,平面特征提 取模型训练时需要大量的计算资源,同时也非常依 赖训练样本数据的质量。



图 4 平面特征提取方法 PlaneRCNN(Liu等,2019)和 PlaneMVS(Liu等,2022)的提取结果 Fig. 4 The results of planar feature extraction method PlaneRCNN (Liu et al., 2019) and PlaneMVS (Liu et al., 2022)((a) RGB; (b) PlaneRCNN;

(c) PlaneMVS)

#### 3.3 平面特征提取方法对比

本文对平面特征提取工作中不同方法的工作原 理、优缺点和代表工作进行了总结,结果如表4所 示。此外,本文还总结了几种平面特征提取方法的 定量分析结果,结果如表5所示。

### 4 语义特征提取方法

语义特征提取即语义分割是当下计算机图形学 领域的关键研究点之一,语义分割在场景重建、场景 理解和自动驾驶等诸多领域的应用前景十分广阔。 1356

JOURNAL C	F IMAGE AND	GRAPHICS
-----------	-------------	----------

Table 4 Comparison of planar reactive extraction methods								
类别	代表工作	原理	优点	缺点				
基于区域增长	Jin等人(2019);Han等人(2021); Roychoudhury等人(2021)	通过相邻点的共同 特性进行平面提取	能快速有效提取连 通区域	提取结果依赖种子 点的选择策略				
基于随机抽样一致	Silberman 等人(2012);Li 等人(2017); Yue等人(2018);Xu等人(2019)	通过迭代采样输入 数据估计平面参数	提取结果鲁棒性好	有限迭代次数内可 能得不到最优解				
基于深度学习	Liu等人(2018);Liu等人(2019);He等人 (2017);Xie等人(2022);Liu等人(2022)	通过深度网络编码 输入特征预测平面	场景适应能力强, 分割精度高	过小平面提取易丢 失问题				

表4 平面特征提取方法对比

Table 4 Comparison of planar feature extraction methods

表5 几种平面特征提取方法定量对比

Table 5	Quantitative	comparison	of several	planar	feature	extraction	methods
---------	--------------	------------	------------	--------	---------	------------	---------

皙汢	在心	S	ScanNet数据集			NYU v2数据集		
<del>异</del> 位	平切	VI	RI	SC	VI	RI	SC	
PlaneNet(Liu等,2018)	2018	1.259	0.858	0.716	1.813	0.753	0.558	
PlaneAE(Yu等,2019)	2019	1.025	0.907	0.791	1.380	0.888	0.519	
PlaneRCNN(Liu等,2019)	2019	1.809	0.880	0.810	1.596	0.839	0.612	
PlaneTR(Tan等,2021)	2021	0.767	0.925	0.838	1.110	0.898	0.726	
PlanarRecon(Xie等,2022)	2022	3.622	0.898	0.247	-	-	-	

注:加粗字体表示各列最优结果,"-"表示该项对比数据缺失。

本节汇总了近几年基于深度学习的语义分割方法, 并划分为基于全监督学习的语义分割方法和基于弱 监督学习的语义分割方法两类。

#### 4.1 基于全监督学习的语义分割方法

全监督学习的语义分割方法需要像素级标注的 训练数据,包括 RGB-D 图像和语义标签图像,通过 深度相机可以轻易获取 RGB-D 图像,但语义标签图 像获取难度较大,多为人工标注。在分割网络训练 阶段,精确标注的训练数据有利于提高分割网络的 分割精度。

Long等人(2015)提出的全卷积神经网络(fully convolutional network, FCN)是基于深度学习的高精 度语义分割的开山之作, FCN 在卷积神经网络(convolutional neural network, CNN)(LeCun 等, 1998)的 基础上使用卷积层代替全连接层,保留了像素的空 间位置信息,因此提高了分割精度。此后 RefineNet (Lin 等, 2017)、SegNet(Badrinarayanan 等, 2017)、AdapNet++(Valada等, 2020)和栗风永等人(2023)方 法等都尝试对单幅 RGB 图像进行语义分割,也都取 得了一定成果。

面向RGB-D图像的语义分割方法,如Jiang等人

(2018)提出的语义分割网络RedNet、Seichter等人 (2021)提出的ESANet都会在特征编码阶段将编码 后的RGB图像特征和深度图像特征相融合,以此拓 宽输入数据维度。ESANet通过一个深度编码器提 取几何信息,ESANet中的RGB编码器会对几何信息 进行二次编码,实现特征融合。基于编码一解码网 络架构的RedNet在特征融合之后都通过跳跃连接 将编码器与解码器中对应的特征图相融合,弥补深 层特征图缺失的细节信息。与ESANet和RedNet不 加思考地融合特征信息不同,Hu等人(2019)提出的 注意力互补网络(attention complementary module,ACM)可以选择性地从RGB图像和深度 图像中收集特征信息。

事实上,将深度图像中的几何信息作为RGB图像的补充信息可以大大提高深度分割网络的性能, 分割网络设计的关键在于2D外观特征和几何信息 有效整合。Lee等人(2017)提出的语义分割网络 RDFNet并非简单地融合RGB图像和深度图像的特征,RDFNet将残差学习的核心思想运用到语义分 割上,通过多模态特征融合块和多层次特征细化块 获得多层次 RGB-D 特征。结果预测时,利用得到的多层次 RGB-D 特征实现精准预测。图5展示了 RDFNet和面向 RGB数据的语义分割方法 RefineNet 的结果,RDFNet的结果明显更接近GT。

此外,Wang和Neumann(2018)提出的深度感知 CNN网络(depth-aware CNN, D-CNN)利用像素间信 息传播过程中深度相似性,在CNN网络中增加深度 感知卷积和深度感知平均池化操作,利用像素间信 息传播过程中的深度相似性,将几何信息巧妙地融 入到CNN网络中。Chen等人(2019)提出的3D邻域 卷积(3D neighborhood convolutions, 3DN-Conv)创新 性地将卷积的感受野和局部深度结构相关联,将一 定深度范围内的区域纳入卷积,提高了结果的分割 精度。孙启超等人(2022)提出的分割网络通过引入 轻量级多模态自适应卷积生成模块实现了多尺度自 适应卷积核的动态生成,在卷积过程中充分利用了 RGB-D图像的内在信息,实现了颜色特征和深度特 征的高效融合。Deng等人(2015)提出的语义分割 网络在条件随机场(conditional random field, CRF)上 增加全局对象共现约束、相对高度关系约束和局部 支撑关系约束,通过这种方式精准利用RGB-D图像 中丰富的几何结构信息。段立娟等人(2021)提出一



图 5 RefineNet(Lin等,2017)和RDFNet(Lee等,2017) 语义分割结果



种基于注意力感知和语义感知的网络模型 ASNet (attention-aware and semantic-aware network),通过引 入注意力感知多模态融合模块和语义感知多模态融 合模块,有效地融合了多层次的 RGB 特征和深度特 征。为了更有效地利用输入图像中的深度数据, 赵经阳等人(2022)在语义分割网络中引入了深度信 息引导的特征提取模块,该模块通过自适应调整深 度数据引入比例,可以在必要时将深度数据嵌入到 提取的特征中以提高语义分割精度。

无论是简单地融合 RGB 特征和深度特征,还是 如上文中提出的在网络中引入处理深度图像的特定 模块,将深度图像中的深度特征融合到语义分割网 络对于提高分割精度大有裨益。

### 4.2 基于半监督学习的语义分割方法

传统基于全监督学习方法中,需要手动标注语 义标签图像用于模型训练,这是一项费时费力的任 务。而基于半监督学习的方法则允许语义分割网络 使用少量标记数据和大量未标记数据训练模型。这 种方法降低了数据标注的成本,成为近些年的热点 研究工作之一。

基于半监督学习的语义分割方法在网络训练阶 段使用的是不精确标注的训练数据,这本身就不利 于提高网络分割精度,业界学者为了提高分割精度 拓展了输入数据维度,将深度信息也输入到网络中。 Maheshwari等人(2023)提出的半监督学习框架M3L 就在网络训练过程中利用线性融合(linear fusion, LF)模块融合RGB-D多层特征,融合的结果将应用 于多模态均值(Multi-modal mean, EMA)模块来监督 有缺失标记的数据,这样不仅提高了语义分割性能 还提高了面向真实未标记场景数据的鲁棒性。 Zhang等人(2022)提出的基于对称编码一解码结构 的语义分割网络ERF-AC-PSPNet也是通过融合多层 特征来保证分割精度,ERF-AC-PSPNet新增残差连 接层和扩展卷积层用于下采样,通过注意力互补模 块融合多层特征图。Stekovic等人(2020)提出的基 于U-Net网络结构实现的半监督语义分割网络S4-Net则是利用场景间的几何一致性处理没有标签可 用的场景。

面向 RGB-D 数据的半监督语义分割网络可以 通过深度图像输入获取几何信息,一定程度上弥补 了数据标注不全带来的影响,保证了语义分割精度。

### 中国图象图形学报

JOURNAL OF IMAGE AND GRAPHICS

/%

### 4.3 语义特征提取方法对比

全监督学习和半监督学习的语义特征提取方法 有着各自的优缺点,表6进行了详细对比。表7展示 了几种面向 RGB、RGB-D 图像的语义分割算法在 NYU v2数据集和 SUN RGB-D 数据集上的定量比较结果,可以看出,面向 RGB-D 图像的语义分割算法具有更高的分割精度。

<b>火</b> 0 固固KGD-D国际的旧关时间近极为亿利比									
Table 6         Comparison of semantic feature extraction methods with RGB-D input									
类别	代表工作	原理	优点	缺点					
基于全监督学习	RedNet(Jiang等,2018);ESANet(Seichter等,2021); ACNet(Hu等,2019);RDFNet(Lee等,2017); D-CNN(Wang和Neumann,2018);3DN-Conv(Chen等, 2019);孙启超等人(2022);Deng等人(2015);ASNet (段立娟等,2021)	利用像素级标注 数据训练网络	分割效果好, 结果精度高	训练集数据标注 工作量大					
基于半监督学习	M3L(Maheshwari等,2023);ERF-AC-PSPNet(Zhang等, 2022);S4-Net(Stekovic等,2020)	利用粗粒度图像 标注训练网络	训练集数据标 注工作量小	监督标签弱,难 以保证分割精度					

**主ィ 面向 DCD D 図** 使的语义特征坦取专注对比

#### 表7 几种语义分割方法定量对比

#### Table 7 Quantitative comparison of several semantic segmentation methods

NYU v2数据集 SUN RGB-D数据集 算法 年份 输入  $\mathbf{PA}$ MPA PA MPA mIoU mIoU Deng等人(2015) 2015 RGB-D 63.8 31.5 \_ \_ RDFNet(Lee 等,2017) 2017 RGB-D 76.0 62.8 50.1 81.5 60.1 47.7 RedNet(Jiang等,2018) 2018 RGB-D \_ 60.3 \_ \_ 81.3 47.8 D-CNN(Wang和Neumann, 2018) 2018 RGB-D 60.3 39.3 27.8 72.4 38.6 29.7 ACNet(Hu 等,2019) 2019 RGB-D \_ \_ 48.3 48.1 3DN-Conv(Chen等,2019) 2019 RGB-D 52.4 39.3 \_ \_ \_ \_ ESANet(Seichter 等, 2021) 2021 **RGB-D** \_ \_ 50.3 \_ \_ 48.17 ASNet(段立娟 等,2021) 2021 RGB-D 77.6 64.7 51.3 82.6 62.7 51.1 孙启超等人(2022) 2022 RGB-D 77.1 64.2 50.8 50.6 82.5 62.0 FCN(Long等,2015) 2015 RGB 65.4 34 \_ \_ \_ 46.1 SegNet(Badrinarayanan等, 2017) RGB 2017 \_ \_ 72.63 44.76 31.84 \_ RefineNet(Lin等,2017) 2017 RGB 73.6 58.9 46.5 80.6 58.5 45.9 AdapNet++(Valada等,2020) 2020 RGB 38.4

注:加粗字体表示各列最优结果,"-"表示该项对比数据缺失。

### 5 亟待解决的问题及发展趋势

#### 5.1 亟待解决的问题

#### 5.1.1 线特征提取问题

商业深度相机采集的RGB-D图像大多存在噪声、缺失等问题,传统方法中,低质量RGB-D输入中

的缺失无法保证线特征结果的鲁棒性,而噪声引起 的深度法向突变也会导致算法检测出错误的结果, 这些问题在车道线检测、自动驾驶等实际应用中安 全隐患极大。虽然通过去噪、深度补全等算法预处 理RGB-D输入可以降低影响,但也增加了算法用 时,无法满足对于实时性要求较高的应用。此外,当 提取结果中的一条线特征被过度分割成多条时,现 有算法缺少线特征整合策略来进一步提高算法提取 结果的鲁棒性。

新兴的基于深度学习的线特征提取方法是一条 新的道路,但大规模完整标注的RGB-D数据集的缺 乏限制了面向RGB-D图像的线特征提取方法的 发展。

5.1.2 平面特征提取问题

低质量的 RGB-D 输入同样可能导致在平面特征提取时出现过小平面缺失、边界不明确和过度分割等问题。在机器人导航和场景理解等任务中,这样的平面特征输出结果会增加算法对于地面理解的困难程度。这种情况下,结合线特征明确平面特征的边界可以作为一个提升平面特征提取质量的研究方向。此外,基于区域增长方法的种子点选择、合并标准设定,基于随机抽样一致方法参数、迭代次数设定都会影响平面提取质量,而在不同的场景中设定这些参数时并没有普适的标准。

5.1.3 语义分割问题

语义分割时,小物体和物体边界区域分割精度 一般不高。这种情况下,以室内场景重建为例,可能 导致重建的场景缺乏更多的细节信息,例如无法有 效区分桌上的杯子和苹果等。语义分割精度不高是 因为语义分割网络在扩大感受视野的同时会丢失更 多的细节信息,这可能导致语义分割网络无法得到 高精度的语义分割结果,因此,网络设计时能否借鉴 ResNet的设计模式,使用类似跳跃连接的操作补全 深层特征图中缺失的细节信息以提高分割精度。此 外,能否在分割网络中融合线、平面特征先验信息, 用于提升不同语义物体边界的分割精度也是值得研 究的方向。虽然大多数的分割算法都使用精准标注 的 RGB-D 数据训练网络,但数据标注费时费力,短 期内很难高效地完成训练数据标注工作。

### 5.2 未来发展趋势

传统的线特征、平面特征提取算法主要利用 RGB-D输入中的颜色、纹理和几何信息,RGB-D输 入中存在的噪声、缺失等问题都会影响提取结果的 质量。现有商业级深度相机采集的RGB-D数据质 量较低,这限制了传统的特征提取算法的发展。与 此同时,基于深度学习的多层特征提取算法表现却 越来越好。全监督学习的特征提取网络十分依赖精 确标注的训练数据,训练集数据标注是一项费时费 力的工作,而半监督学习的深度网络通过混合使用 少量标记数据和大量未标记数据就可以训练网络。 除半监督学习外,自监督学习(self-supervised learning)、无监督学习(unsupervised learning)的深度特征 提取网络对训练数据要求也比较低,就这一点而言, 半监督、自监督和无监督学习的多层特征提取网络 将成为未来发展的趋势。此外,基于生成对抗网络 (generative adversarial networks,GAN)的深度特征提 取方法也为多层特征提取提供了一个全新的解决方 案,GAN利用生成器和判别器反复对抗的过程提高 网络的精度,是一个全新的思路。新兴起的迁移学 习主张利用已有的经验解决相似的任务,利用数 据、任务和模型间的相似性,将训练好的内容应用到 新的任务上,这样可以节约大量的训练时间,提高效 率。因此基于GAN和迁移学习的多层特征提取也 是一个非常值得探索的方向。

### 6 结 语

本文全面综述了面向 RGB-D 图像的多层特征 提取算法相关的数据集、评价指标、研究现状、亟待 解决的问题以及未来发展的趋势。依据特征所处的 不同层次,从底层的线特征、平面特征和高层的语义 特征3个方面进行了分类汇总。传统的线特征、平 面特征提取算法十分依赖RGB-D输入的质量,RGB-D图像中存在的噪声、缺失等问题导致特征提取结 果的质量较低。随着深度学习技术的快速发展,基 于深度学习的多层特征提取算法为面向 RGB-D 图 像的特征提取开辟了一条全新的道路。本文从传统 和基于深度学习两个角度对RGB-D图像底层的线 特征、平面特征进行了综述,对各算法的技术原理、 优缺点进行了详细说明;此外,还从全监督和半监督 两个角度总结了面向 RGB-D 图像的语义特征提取 方法。最后总结了当前面向 RGB-D 图像的多层特 征提取算法存在的难以解决的问题,展望了未来发 展趋势。

### 参考文献(References)

Akinlar C and Topal C. 2011. EDLines: a real-time line segment detector with a false detection control. Pattern Recognition Letters, 32(13): 1633-1642 [DOI: 10.1016/j.patrec.2011.06.001]

Ammirato P, Poirson P, Park E, Košecká J and Berg A C. 2017. A data-

Vol. 29, No. 5, May 2024

set for developing and benchmarking active vision//Proceedings of 2017 IEEE International Conference on Robotics and Automation. Singapore, Singapore: IEEE: 1378-1385 [DOI: 10.1109/ICRA. 2017.7989164]

- Badrinarayanan V, Kendall A and Cipolla R. 2017. SegNet: a deep convolutional encoder-decoder architecture for image segmentation. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(12): 2481-2495 [DOI: 10.1109/TPAMI.2016.2644615]
- Bose L and Richards A. 2016. Fast depth edge detection and edge based RGB-D SLAM//Proceedings of 2016 IEEE International Conference on Robotics and Automation. Stockholm, Sweden: IEEE: 1323-1330 [DOI: 10.1109/ICRA.2016.7487265]
- Canny J. 1986. A computational approach to edge detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI-8 (6) : 679-698 [DOI: 10.1109/TPAMI.1986.4767851]
- Cao Y P, Ju T, Xu J and Hu S M. 2017. Extracting sharp features from RGB-D images. Computer Graphics Forum, 36 (8) : 138-152 [DOI: 10.1111/cgf.13069]
- Chen L C, Zhu Y K, Papandreou G, Schroff F and Adam H. 2018a. Encoder-decoder with atrous separable convolution for semantic image segmentation//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: Springer: 833-851 [DOI: 10.1007/978-3-030-01234-2\_49]
- Chen X Z, Kundu K, Zhu Y K, Ma H M, Fidler S and Urtasun R. 2018b. 3D object proposals using stereo imagery for accurate object class detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(5): 1259-1272 [DOI: 10.1109/TPAMI. 2017.2706685]
- Chen Y L, Mensink T and Gavves E. 2019. 3D neighborhood convolution: learning depth-aware features for RGB-D and RGB semantic segmentation//Proceedings of 2019 International Conference on 3D Vision (3DV). Quebec City, Canada: IEEE: 173-182 [DOI: 10. 1109/3DV.2019.00028]
- Cho N G, Yuille A and Lee S W. 2018. A novel Linelet-based representation for line segment detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 40(5): 1195-1208 [DOI: 10. 1109/TPAMI.2017.2703841]
- Choi C, Trevor A J B and Christensen H I. 2013. RGB-D edge detection and edge-based registration//Proceedings of 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems. Tokyo, Japan: IEEE: 1568-1575 [DOI: 10.1109/IROS.2013.6696558]
- Dai A, Chang A X, Savva M, Halber M, Funkhouser T and Nießner M. 2017. ScanNet: richly-annotated 3D reconstructions of indoor scenes//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 2432-2443 [DOI: 10.1109/CVPR.2017.261]
- Deng Z, Todorovic S and Jan Latecki L. 2015. Semantic segmentation of

RGBD images with mutex constraints//Proceedings of 2015 IEEE International Conference on Computer Vision. Santiago, Chile: IEEE: 1733-1741 [DOI: 10.1109/ICCV.2015.202]

- Duan L J, Sun Q C, Qiao Y H, Chen J C and Cui G Q. 2021. Attentionaware and semantic-aware network for RGB-D indoor semantic segmentation. Chinese Journal of Computers, 44(2): 275-291 (段立 娟, 孙启超, 乔元华, 陈军成, 崔国勤. 2021. 基于注意力感知 和语义感知的 RGB-D 室内图像语义分割算法. 计算机学报, 44(2): 275-291) [DOI: 10.11897/SP.J.1016.2021.00275]
- Dzitsiuk M, Sturm J, Maier R, Ma L N and Cremers D. 2017. De-noising, stabilizing and completing 3D reconstructions on-thego using plane priors//Proceedings of 2017 IEEE International Conference on Robotics and Automation. Singapore, Singapore: IEEE: 3976-3983 [DOI: 10.1109/ICRA.2017.7989457]
- Han X N, Wang X H, Leng Y Q and Zhou W J. 2021. A plane extraction approach in inverse depth images based on region-growing. Sensors, 21(4): #1141 [DOI: 10.3390/s21041141]
- He K M, Gkioxari G, Dollár P and Girshick R. 2017. Mask R-CNN//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2980-2988 [DOI: 10.1109/ICCV. 2017.322]
- He K M, Zhang X Y, Ren S Q and Sun J. 2016. Deep residual learning for image recognition//Proceedings of 2016 IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas, USA: IEEE: 770-778 [DOI: 10.1109/CVPR.2016.90]
- Hu X X, Yang K L, Fei L and Wang K W. 2019. ACNET: Attention based network to exploit complementary features for RGBD semantic segmentation//Proceedings of 2019 IEEE International Conference on Image Processing. Taipei, China: IEEE: 1440-1444 [DOI: 10.1109/ICIP.2019.8803025]
- Hu Z T, Chen C, Yang B S, Wang Z Y, Ma R Q, Wu W T and Sun W L. 2022. Geometric feature enhanced line segment extraction from large-scale point clouds with hierarchical topological optimization. International Journal of Applied Earth Observation and Geoinformation, 112: #102858 [DOI: 10.1016/j.jag.2022.102858]
- Janoch A, Karayev S, Jia Y Q, Barron J T, Fritz M, Saenko K and Darrell T. 2011. A category-level 3-D object dataset: putting the kinect to work//Proceedings of 2011 IEEE International Conference on Computer Vision Workshops. Barcelona, Spain: IEEE: 1168-1174 [DOI: 10.1109/ICCVW.2011.6130382]
- Jiang J D, Zheng L N, Luo F and Zhang Z J. 2018. RedNet: residual encoder-decoder network for indoor RGB-D semantic segmentation [EB/OL]. [2023-06-03]. https://arxiv.org/pdf/1806.01054.pdf
- Jin Z, Tillo T, Zou W B, Zhao Y and Li X. 2019. Robust plane detection using depth information from a consumer depth camera. IEEE Transactions on Circuits and Systems for Video Technology, 29(2): 447-460 [DOI: 10.1109/TCSVT.2017.2780181]

- Lai K, Bo L F, Ren X F and Fox D. 2011. A large-scale hierarchical multi-view RGB-D object dataset//Proceedings of 2011 IEEE International Conference on Robotics and Automation. Shanghai, China; IEEE: 1817-1824 [DOI: 10.1109/ICRA.2011.5980382]
- LeCun Y, Bottou L, Bengio Y and Haffner P. 1998. Gradient-based learning applied to document recognition. Proceedings of the IEEE, 86(11): 2278-2324 [DOI: 10.1109/5.726791]
- Lee S, Park S J and Hong K S. 2017. RDFNet: RGB-D multi-level residual feature fusion for indoor semantic segmentation//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 4990-4999 [DOI: 10.1109/ICCV.2017.533]
- Li F Y, Ye B and Qin C. 2023. Mutual attention mechanism-driven lightweight semantic segmentation network. Journal of Image and Graphics, 28(7): 2068-2080 (栗风永,叶彬,秦川. 2023. 互注意力机 制驱动的轻量级图像语义分割网络.中国图象图形学报, 28(7): 2068-2080) [DOI: 10.11834/jig.211127]
- Li L, Yang F, Zhu H B, Li D L, Li Y and Tang L. 2017. An improved RANSAC for 3D point cloud plane segmentation based on normal distribution transformation cells. Remote Sensing, 9(5): #433 [DOI: 10.3390/rs9050433]
- Lin G S, Milan A, Shen C H and Reid I. 2017. RefineNet: multi-path refinement networks for high-resolution semantic segmentation//Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, USA: IEEE: 5168-5177 [DOI: 10.1109/ CVPR.2017.549]
- Liu C, Kim K, Gu J W, Furukawa Y and Kautz J. 2019. PlaneRCNN: 3D plane detection and reconstruction from a single image//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 4445-4454 [DOI: 10. 1109/CVPR.2019.00458]
- Liu C, Yang J M, Ceylan D, Yumer E and Furukawa Y. 2018. PlaneNet: piece-wise planar reconstruction from a single RGB image//Proceedings of 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, USA: IEEE: 2579-2588 [DOI: 10.1109/CVPR.2018.00273]
- Liu J C, Ji P, Bansal N, Cai C J, Yan Q A, Huang X L and Xu Y. 2022. PlaneMVS: 3D plane reconstruction from multi-view stereo// Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 8655-8665 [DOI: 10.1109/CVPR52688.2022.00847]
- Long J, Shelhamer E and Darrell T. 2015. Fully convolutional networks for semantic segmentation//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 3431-3440 [DOI: 10.1109/CVPR.2015.7298965]
- Lu X H, Liu Y H and Li K. 2019. Fast 3D line segment detection from unorganized point cloud [EB/OL]. [2023-06-03]. https://arxiv.org/pdf/1901.02532.pdf

- Lu X H, Yao J, Li K and Li L. 2015. CannyLines: a parameter-free line segment detector//Proceedings of 2015 IEEE International Conference on Image Processing. Quebec City, Canada: IEEE: 507-511 [DOI: 10.1109/ICIP.2015.7350850]
- Maheshwari H, Liu Y C and Kira Z. 2023. Missing modality robustness in semi-supervised multi-modal semantic segmentation [EB/OL]. [2023-06-03]. https://arxiv.org/pdf/2304.10756.pdf
- McCormac J, Handa A, Leutenegger S and Davison A J. 2017. SceneNet RGB-D: can 5M synthetic images beat generic imagenet pre-training on indoor segmentation?//Proceedings of 2017 IEEE International Conference on Computer Vision. Venice, Italy: IEEE: 2697-2706 [DOI: 10.1109/ICCV.2017.292]
- Nieto M, Cuevas C, Salgado L and García N. 2011. Line segment detection using weighted mean shift procedures on a 2D slice sampling strategy. Pattern Analysis and Applications, 14 (2) : 149-163 [DOI: 10.1007/s10044-011-0211-4]
- Ronneberger O, Fischer P and Brox T. 2015. U-Net: convolutional networks for biomedical image segmentation//Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention. Munich, Germany: Springer: 234-241 [DOI: 10.1007/978-3-319-24574-4\_28]
- Roychoudhury A, Missura M and Bennewitz M. 2021. Plane segmentation in organized point clouds using flood fill//Proceedings of 2021 IEEE International Conference on Robotics and Automation. Xi'an, China: IEEE: 13532-13538 [DOI: 10.1109/ICRA48506. 2021.9561325]
- Seichter D, Köhler M, Lewandowski B, Wengefeld T and Gross H M. 2021. Efficient RGB-D semantic segmentation for indoor scene analysis//Proceedings of 2021 IEEE International Conference on Robotics and Automation. Xi' an, China: IEEE: 13525-13531 [DOI: 10.1109/ICRA48506.2021.9561675]
- Silberman N and Fergus R. 2011. Indoor scene segmentation using a structured light sensor//Proceedings of 2011 IEEE International Conference on Computer Vision Workshops. Barcelona, Spain: IEEE: 601-608 [DOI: 10.1109/ICCVW.2011.6130298]
- Silberman N, Hoiem D, Kohli P and Fergus R. 2012. Indoor segmentation and support inference from RGBD images//Proceedings of the 12th European Conference on Computer Vision. Florence, Italy: 746-760 [DOI: 10.1007/978-3-642-33715-4\_54]
- Song S R, Lichtenberg S P and Xiao J X. 2015. SUN RGB-D: a RGB-D scene understanding benchmark suite//Proceedings of 2015 IEEE Conference on Computer Vision and Pattern Recognition. Boston, USA: IEEE: 567-576 [DOI: 10.1109/CVPR.2015.7298655]
- Stekovic S, Fraundorfer F and Lepetit V. 2020. Casting geometric constraints in semantic segmentation as semi-supervised learning//Proceedings of 2020 IEEE Winter Conference on Applications of Computer Vision. Snowmass, USA: IEEE: 1843-1852 [DOI: 10.1109/

WACV45572.2020.9093571]

- Sturm J, Engelhard N, Endres F, Burgard W and Cremers D. 2012. A benchmark for the evaluation of RGB-D SLAM systems//Proceedings of 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems. Vilamoura-Algarve, Portugal: IEEE: 573-580 [DOI: 10.1109/IROS.2012.6385773]
- Sun Q C, En Q, Duan L J and Qiao Y H. 2022. RGB-D image semantic segmentation based on multi-modal adaptive convolution. Journal of Computer-Aided Design and Computer Graphics, 34(8): 1272-1282 (孙启超,恩擎,段立娟,乔元华. 2022. 基于多模态自适 应卷积的 RGB-D 图像语义分割. 计算机辅助设计与图形学学 报,34(8): 1272-1282) [DOI: 10.3724/SP.J.1089.2022.19132]
- Tan B, Xue N, Bai S, Wu T F and Xia G S. 2021. PlaneTR: structureguided Transformers for 3D plane recovery//Proceedings of 2021 IEEE/CVF International Conference on Computer Vision. Montreal, Canada: IEEE: 4166-4175 [DOI: 10.1109/ICCV48922. 2021.00415]
- Tian X, Wang L and Ding Q. 2019. Review of image semantic segmentation based on deep learning. Journal of Software, 30(2): 440-468 (田萱, 王亮, 丁琪. 基于深度学习的图像语义分割方法综述. 软件学报, 30(2): 440-468) [DOI: 10.13328/j.cnki.jos.005659]
- Valada A, Mohan R and Burgard W. 2020. Self-supervised model adaptation for multimodal semantic segmentation. International Journal of Computer Vision, 128(5): 1239-1285 [DOI: 10.1007/s11263-019-01188-y]
- von Gioi R G, Jakubowicz J, Morel J M and Randall G. 2010. LSD: a fast line segment detector with a false detection control. IEEE Transactions on Pattern Analysis and Machine Intelligence, 32(4): 722-732 [DOI: 10.1109/TPAMI.2008.300]
- Wang W Y and Neumann U. 2018. Depth-aware CNN for RGB-D segmentation//Proceedings of the 15th European Conference on Computer Vision. Munich, Germany: 144-161 [DOI: 10.1007/978-3-030-01252-6\_9]
- Xiao J X, Owens A and Torralba A. 2013. SUN3D: a database of big spaces reconstructed using SfM and object labels//Proceedings of 2013 IEEE International Conference on Computer Vision. Sydney, Australia: IEEE: 1625-1632 [DOI: 10.1109/ICCV.2013.458]
- Xie Y M, Gadelha M, Yang F T, Zhou X W and Jiang H Z. 2022. PlanarRecon: realtime 3D plane detection and reconstruction from posed monocular videos//Proceedings of 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans, USA: IEEE: 6209-6218 [DOI: 10.1109/CVPR52688.2022. 00612]
- Xu D, Li F H and Wei H X. 2019. 3D point cloud plane segmentation method based on RANSAC and support vector machine//Proceedings of the 14th IEEE Conference on Industrial Electronics and Applications. Xi' an, China: IEEE: 943-948 [DOI: 10.1109/

ICIEA.2019.8834367]

- Xu Y F, Xu W J, Cheung D and Tu Z W. 2021. Line segment detection using Transformers without edges//Proceedings of 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Nashville, USA: IEEE: 4255-4264 [DOI: 10.1109/CVPR46437.2021. 00424]
- Xue N, Bai S, Wang F D, Xia G S, Wu T F and Zhang L P. 2019. Learning attraction field representation for robust line segment detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 1595-1603 [DOI: 10.1109/CVPR.2019.00169]
- Yang B J, Chen E Q, Yang S Y and Bai W J. 2015. RGB-D geometric features extraction and edge-based scene-SIRFS//Proceedings of 2015 IEEE International Conference on Communication Software and Networks. Chengdu, China: IEEE: 306-311 [DOI: 10.1109/ ICCSN.2015.7296174]
- Yang N, Mi Z Q, Guo Y, Sadoun B and Obaidat M S. 2020. Fast local map construction of robot using semantic priors//Proceedings of 2020 International Conference on Communications, Computing, Cybersecurity, and Informatics. Sharjah, United Arab Emirates: IEEE: 1-5 [DOI: 10.1109/CCCI49893.2020.9256777]
- Yu Z H, Zheng J, Lian D Z, Zhou Z H and Gao S H. 2019. Singleimage piece-wise planar 3D reconstruction via associative embedding//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 1029-1037 [DOI: 10.1109/CVPR.2019.00112]
- Yue W L, Lu J G, Zhou W H and Miao Y B. 2018. A new plane segmentation method of point cloud based on mean shift and RANSAC// Proceedings of 2018 Chinese Control and Decision Conference. Shenyang, China: IEEE: 1658-1663 [DOI: 10.1109/CCDC.2018. 8407394]
- Zhang M Y, Liu X L and Xu D. 2017. Survey on line segment detection on images//Proceedings of the 36th Chinese Control Conference.
  Dalian, China: Control Theory Professional Committee of China Association of Automation: 1287-1293 (张茗奕,刘希龙,徐德.
  2017. 图像线段提取方法综述//第36届中国控制会议论文集.大 连,中国:中国自动化学会控制理论专业委员会: 1287-1293)
- Zhang Y J, Wei D and Li Y S. 2021. AG3line: active grouping and geometry-gradient combined validation for fast line segment extraction. Pattern Recognition, 113: #107834 [DOI: 10.1016/j.patcog. 2021.107834]
- Zhang Z H, Li Z X, Bi N, Zheng J, Wang J L, Huang K, Luo W X, Xu Y Y and Gao S H. 2019. PPGNet: learning point-pair graph for line segment detection//Proceedings of 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach, USA: IEEE: 7098-7107 [DOI: 10.1109/CVPR.2019.00727]
- Zhang Z Y, Deng H G, Liu Y, Xu Q G and Liu G. 2022. A semi-

supervised semantic segmentation method for blast-hole detection.
Symmetry, 14(4): #653 [DOI: 10.3390/sym14040653]

Zhao J Y, Yu C Q and Sang N. 2022. RGB-D semantic segmentation: depth information selection. Journal of Image and Graphics, 27(8): 2473-2486 (赵经阳, 余昌黔, 桑农. 2022. RGB-D 语义 分割: 深度信息的选择使用. 中国图象图形学报, 27(8): 2473-2486) [DOI: 10.11834/jig.210061]

### 作者简介

李洋,男,硕士研究生,主要研究方向为计算机图形学、数字 几何处理和图像处理。E-mail:li\_yang@st.btbu.edu.cn 吴晓群,通信作者,女,副教授,主要研究方向为计算机图形 学、数字几何处理和图像处理。 E-mail:wuxiaoqun@btbu.edu.cn