

如何入门多视角3D目标识别？超详细最新综述来袭！

原创

CV开发者都爱看的

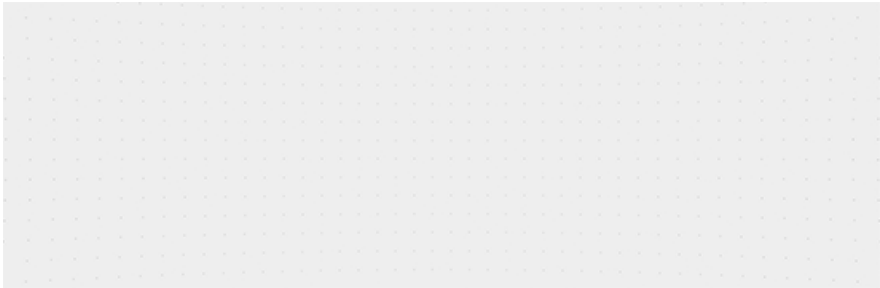
极市平台

2021-08-03 22:00:00

手机阅读

👁

↑ 点击蓝字 关注极市平台



作者 | 祁少华、宁欣、杨国为等

编辑 | 极市平台

Y

壹伴图



极市平台

extreme

月发文数目: **

月平均阅读: **

文章工具

已发文

采集图文

合成多

采集样式

查看详

极市导读

本文将对多视角3D目标识别方法的主要进展、代表性研究成果进行介绍，比分析、汇总多数方法的实验性能，总结了多视角3D目标识别研究中的难点和热点、发展趋势。 >>加入极市CV技术交流群，走在计算机视觉的最前沿

Original research article

Modulating self-similar vector bisolitons

Yan Zhou ^a✉, Xiaoyan Lin ^a, Meisong Liao ^b, Guoying Zhao ^c, Yongzheng Fang ^c

^a School of Science, Shanghai Institute of Technology, Shanghai 201418, China

^b Key Laboratory of Materials for High Power Laser, Shanghai Institute of Optics and Fine Mechanics, Chinese Academy of Sciences, Shanghai 201800, China

^c School of Materials Science and Engineering, Shanghai Institute of Technology, Shanghai 201418, China

Received 28 June 2021, Revised 7 July 2021, Accepted 7 July 2021, Available online 11 July 2021.

本文将重点对近年来基于深度学习的多视角3D目标识别方法的主要进展和部分具有代表性的研究成果进行介绍，通过详细的对比分析，以及汇总现有的绝大多数方法的实验性能，总结了多视角3D目标识别研究中的难点和热点，以及可能的发展趋势。

论文下载地址：

<https://doi.org/10.1016/j.displa.2021.102053>

目录：

- 0 引言
- 1 背景
- 2 视点设置与视图输入模式
- 3 多视角3D目标识别的深度学习方法
- 4 性能
- 5 总结和展望

摘要

3D目标识别广泛应用于诸多领域。深度学习越来越多地被用来解决3D视觉问题。基于深度学习技术的多视角3D目标识别，由于可以直接使用成熟的分类网络作为骨干网络，并且来自多个视角的视图可以互补目标的细节特征，因此成为研究的热点之一。然而，该领域仍然存在一些挑战。最近，已经提出了许多方法来解决与该主题相关的问题。本文对应用于多视角3D目标识别的深度学习方法的最新进展进行了全面的回顾和分类，汇总了这些方法在几个主流数据集上的结果，提供了有见地的总结，并提出了有启发性的未来研究方向。

0 引言

随着AI等领域的快速发展，各种环境下的目标识别需求变得迫切而重要。因此，目标识别领域发展迅速。随着3D采集技术和3D物体数据集的增加和成熟，3D目标识别已成为目标识别领域重要且热门的研究方向之一。深度学习是机器学习的一个新领域。由于卷积神经网络的发展，目标识别进入了一个新的阶段。深度学习和CNN在3D目标分类中得到了广泛的应用，并取得了优异的性能。

基于深度学习的3D目标识别方法根据其输入模型可分为三个方向，即基于体素的方法、基于点集的方法和基于视图的方法。在基于体素的方法中，目标被表示为3D网格，并通过三维网络进行分析。在基于点集的方法中，一个对象被表示为一组无序的点，并使用点云进行预测。这两种方法也可以统称为基于模型的方法。基于模型的方法使用3D卷积滤波器来卷积三维空间中的3D形状，从而直接从3D数据生成3D表示。

虽然它们利用了3D物体的空间信息，但实际应用受到计算量的限制。基于视图的方法从不同的视点将3D目标渲染成2D图像，并使用2D卷积滤波器对这些视图进行卷积。基于视图的方法不依赖于复杂的3D特征，并且容易捕获输入视图。它们拥有大量的数据，可以利用成熟的高级网络框架。在遮挡的情况下，不同视点的视图可以互补目标的细节特征，达到很好的识别效果。

自MVCNN问世以来的六年中，基于多视角技术的三维目标识别方法的研究取得了长足的发展。与现有的综述不同，我们的综述侧重于多视角三维目标识别的深度学习方法，主要包括2018-2021年的最新研究成果，创新性地将本综述分为（i）视点设置和输入模式，（ii）主干网络和特征提取，特征融合和视点选择机制，如图1所示。

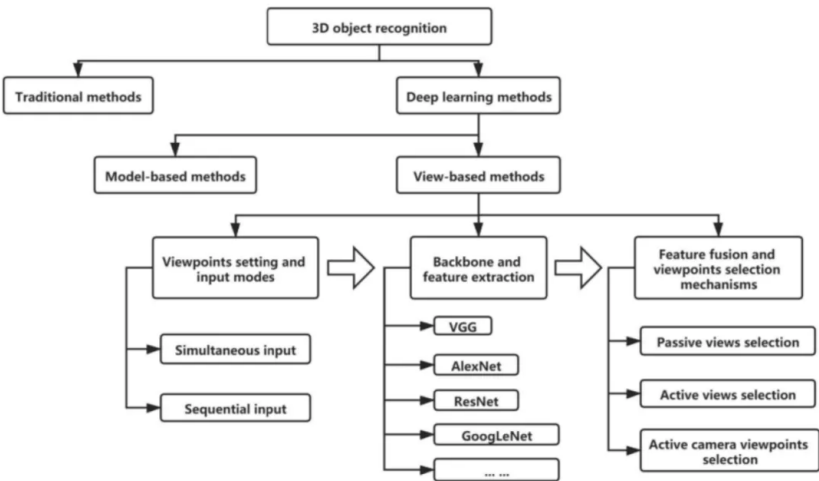


Fig. 1. Various 3D object recognition methods based on the multi-view approach.

图1

其中，与二维图像分类最大的区别在于视点设置和输入方式，以及特征融合和视点选择机制。二维图像分类通常不需要考虑这两点，而是分别处理数据集中的每幅图像。在基于多视角的三维目标识别中，从多个角度获取物体的特征，并将其进一步融合以提高识别精度。因此，特征融合成为该领域的研究热点。本文重点研究了现有方法的视点设置和特征融合，并在主流数据集上汇总性能进行了比较。

本文综述了基于多视角的三维目标识别深度学习方法，主要内容如下:第2节介绍了主流的数据集和评价指标。第3节概述了现有多视图三维目标识别方法中使用的视点设置和输入模式。第4节从特征提取和融合策略两个方面综述了基于多视角的三维目标识别方法。第5节总结了大多数方法在几个主流数据集上的性能。第6节总结了基于多视角的三维目标识别的特点，并对未来的研究方向进行了简要的讨论。

1 背景

本节介绍了主流数据集和过去六年中对基于多视角的三维目标识别进行的各种研究中使用的评估指标。

1.1 数据集

表1列出了在多视角3D目标识别的不同研究中使用的主要数据集。

Table 1
Datasets used in studies on multi-view 3D object recognition.

Datasets for 3D Shape Classification							
Name	Year	Samples	Classes	Training	validation	Test	Type
RGB-D [59]	2011	250000	51	-	-	-	Real-World
ModelNet40 [13]	2015	12311	40	9843	-	2468	Synthetic
ModelNet10 [13]	2015	4899	10	3991	-	908	Synthetic
ShapeNetCore [60]	2015	51162	55	35764	5133	10265	Synthetic
MIRO [30]	2018	160	12	-	-	-	Real-World

1.2 评价指标

随着视觉内容数量和维度的不断增加，对视觉内容进行快速分类和检索是许多应用的重要任务。通常，基于多视角的3D目标识别方法使用两种类型的评价指标，分类任务和检索任务。3D目标检索涉及在数据集中找到指定的查询目标，而3D目标分类涉及识别给定3D对象的类别。

2 视点设置与视图输入模式

基于多视角的3D目标识别将复杂的3D分类任务简化为简单的2D分类任务，因此需要首先将3D目标渲染成2D图像。执行视点设置的这一步骤以获得具有不同视角的2D渲染图像，使得从不同视图捕获的特征可以是互补且相互关联的，并且可以帮助提高网络的识别性能。在现有的研究中，相机视点通常预先设置在不同的位置，并且从这些视点执行3D目标到2D图像的渲染。然后如此获得的图像被输入到网络中，并且使用2D卷积滤波器完成来自不同视点的所有视图的卷积。根据现有研究中的视图输入，我们将其分为同时输入和顺序输入。

2.1 同时输入

由Su等人提出的MVCNN是基于多视角的3D目标识别领域的开创性工作，其中测试了两个摄像机设置。在第一个相机设置中，假设目标沿着一致的轴(如Z轴)垂直定向。通过在对象周围以30°的间隔设置虚拟摄像机，可以收集12个渲染视图(图2)。相机从地面升起30°，指向目标的质心。第二个相机设置没有利用目标处于相同垂直方向的假设。在这种情况下，由于能够产生目标的良好代表性视图的视点不是先验的，因此他们在二十面体的20个顶点放置了20个虚拟摄影机（我们认为Su等人在文章中写错了，应该是十二面体，因为只有十二面体才有20个顶点）来围绕目标进行渲染。所有的摄像机都指向目标的质心。然后，每个摄像头分别沿通过摄像机和物体的质心的轴旋转0°, 90°, 180°, 和270°，生成四个渲染视图，共提供80个视图。网络训练采用同时输入，可以同时从所有摄像机向网络输入任意数量的视图。

2.2 顺序输入

为了解决视图中内容以及空间中视图之间的关系被忽略的问题，Han等人提出了SeqViews2SeqLabels的模型。它类似于GVCNN的视点设置，它围绕每个3D目标捕捉一个圆内的连续视图，形成由按顺序均匀分布在圆上的V个视图组成的视图序列。相机设置在地面以上30°的角度，并指向3D对象的质心。随机选择圆上的一个位置作为视图序列的开始，并以360°/ V的间隔在相同的序列方向上获取后续视图。虽然3D目标的顶部或底部和GVCNN一样不能完全覆盖，但Han等人采用的顺序视图方法能够更有效地捕获低级特征，同时保留视图之间的空间信息用于3D全局特征学习。

3 多视角3D目标识别的深度学习方法

受益于CNN的进步及其在2D图像分类中的成功，我们可以通过使用图像描述符的进步和微调3D模型投影的细节来学习2D图像分类的大量一般特征。用于多视图3D目标识别的深度学习方法可以直接使用由优秀的大规模2D公共数据集(如ImageNet)训练的成功的高级分类网络作为骨干网络。这节省了网络训练的时间，将复杂的三维分类任务简化为简单的2D分类任务，提高了网络

精度。在该方法中，使用在2D分类任务中表现良好的CNN从2D渲染图像中提取特征，并将提取的特征融合以获得用于分类的形状描述符。图5示出了近年来为多视角3D目标识别开发的深度学习方法的时间轴。

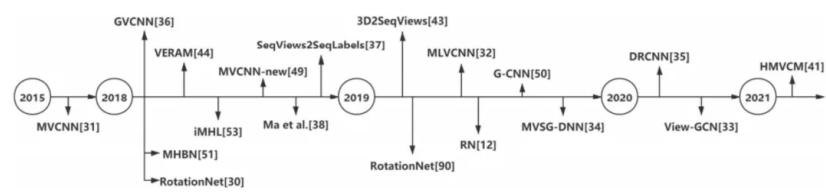


Fig. 5. Time axis of the deep learning methods developed for multi-view 3D object recognition in recent years.

图5

3.1 骨干网络和特征提取

骨干网络是在2D分类任务中表现良好的CNN。它被用于3D识别任务中网络开始时的特征提取和网络结束时的分类；因此，使其作为该方法的框架。3D识别任务首先使用由大规模2D数据集训练的成熟分类网络来提取不同视图中的视图级特征。

3.2. 特征融合

成熟的2D分类网络提取的视图级特征需要特征融合来生成全局描述符并提供精确的对象分类。对于这些方法来说，开发一个将多个视图特征聚集成一个不同的全局表示的过程是一个关键的挑战。原始方法使用从所有视点收集的所有视图。然而，由于处理所有视图需要很高的计算量，并且不是每个视图都有助于识别，因此一些方法引入了代表性视图选择机制，而另一些方法引入了下一视点选择机制。根据现有特征融合方法的不同选择机制，我们将其分为被动视图选择机制、主动视图选择机制和主动像机视点选择机制。主动选择机制可以使网络自动选择最佳视图，从而节省计算成本，提高网络性能。

4 性能

多视角3D目标识别的实验涉及到各种数据集，如ModelNet40/10、ShapeNetCore、RGB-D、MIRO等数据集，包括分类和检索。因为大多数方法使用ModelNet40/10，所以我们对使用ModelNet40/10数据集实现的不同网络的性能进行分类和总结。主要以图像为模态总结了各种方法的实验结果。包含体素、点云和其他模态的不同论文中提供的实验结果表也将一起总结。

5 总结与展望

在3D目标识别任务中，得益于CNN的进步及其在2D图像分类中的成功，基于视图的方法可以直接使用由优秀的大规模2D公共数据集训练的成功的高级分类网络作为骨干网络。这样节省了训练网络的时间，将复杂的3D分类任务简化为简单的2D分类任务，保证了网络的精度。

现有的基于多视角的3D目标识别方法的性能指标已经充分显示了其优越性，但仍有一些方面需要改进。例如，有效地利用多视点图像之间的关系，改进特征融合，已经成为当前研究工作的主流方向。例如，多视点3D目标识别中的被动视图选择方法计算量大，性能有限，而主动视图选择方法仍然需要输入所有视图，存在遮挡等实际问题。计算机视觉的发展要求网络能够自动选择目标的最佳视角，即主动像机视点选择。这些最佳视图将具有最丰富的图像信息和最高的可分辨性，这将有助于网络以尽可能少的输入视图获得最佳的识别性能，并在解决遮挡问题的同时降低移动机器人的成本。未来，主动像机视点的选择将与机器人的主动视觉密切相关，以便机器人能够找到最佳的视点来观察和识别目标。它将是多视点3D目标识别领域的一个很好的发展方向，具有非常重要的应用前景。

- 作者：
祁少华^{1,2}，宁欣^{1,3,5}，杨国为²，张丽萍^{1,3}，龙鹏¹，蔡薇薇⁴，李卫军^{1,5}
- 单位信息：


- 4、中南林业科技大学；
- 5、中国科学院大学。

• 引用：

1. Qi S, Ning X, Yang G, et al. Review of Multi-view 3D Object Recognition Methods Based on Deep Learning[J].

2. Displays, 2021: 102053. DOI: doi.org/10.1016/j.displa.2021.10205

如果觉得有用，就请分享到朋友圈吧！



极市平台

专注计算机视觉前沿资讯和技术干货，官网：www.cvmart.net

624篇原创内容

公众号

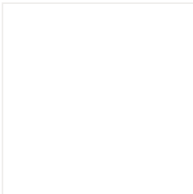
▲点击卡片关注极市平台，获取最新CV干货

公众号后台回复“CVPR21检测”获取CVPR2021目标检测论文下载～

极市干货

- 深度学习环境搭建：如何配置一台深度学习工作站？
- 实操教程：OpenVINO2021.4+YOLOX目标检测模型测试部署 | 为什么你的显卡利用率总是0%？
- 算法技巧（trick）：图像分类算法优化技巧 | 21个深度学习调参的实用技巧

CV技术社群邀请函 #
.....



△长按添加极市小助手

添加极市小助手微信（ID：cvmart4）

备注：姓名-学校/公司-研究方向-城市（如：小极-北大-目标检测-深圳）

即可申请加入极市目标检测/图像分割/工业检测/人脸/医学影像/3D/SLAM/自动驾驶/超分辨率/姿态估计/ReID/GAN/图像增强/OCR/视频理解等技术交流群

每月大咖直播分享、真实项目需求对接、求职内推、算法竞赛、干货资讯汇总、与 10000+来自港科大、北大、清华、中科院、CMU、腾讯、百度等名校名企视觉开发者互动交流～

[觉得有用麻烦给个在看啦~](#)

[阅读原文](#)

喜欢此内容的人还喜欢

15个目标检测开源数据集汇总
极市平台