视觉冲击的突出束调整

王珂，IEEE成员，马赛，陈俊兰，陆建波，IEEE研究员

*摘要*-最近，视觉显著性和注意力的哲学在机器人界开始流行起来。因此，本文旨在利用显著性预测模型在SLAM框架下模拟这一机制。与传统的SLAM算法在优化过程中将所有特征点同等重要相比，我们认为显著特征点在优化过程中应该发挥更大的作用。因此，本文提出了一种能同时捕捉场景语义和几何信息的显著性模型来预测显著性图。然后，在传统的束平差方法中，利用显著图的值作为特征点的权重，提出了显著束平差方法。在KITTI和EuRoc数据集上对最新算法进行了详尽的实验，结果表明，该算法在室内和室外环境下均优于现有算法。最后，我们将使我们的显著性数据集和相关的源代码开源，以便将来的研究。

*索引项*-视觉里程表，深度学习，大满贯，

显著性预测。

# 一、 导言

|  |
| --- |
| S码 |

LAM是指在未知环境中估计机器人的姿态，同时建立传感器感知的环境模型[1]，[2]。在过去的几十年里，SLAM已经在计算机视觉和机器人领域进行了研究，并且仍然是各种应用中不可或缺的模块，从自动车辆[3]、[4]和医疗机器人[5]、[6]到增强和虚拟现实[7]、[8]。对实时、准确和健壮的解决方案的不懈追求导致了许多方法的涌现[9]-[13]。

一般来说，这些方法可以分为两类：基于几何的方法和基于深度学习的方法。在过去的几十年中，我们在基于几何的方法上取得了令人印象深刻的进展，并在精确和实时性方面表现出了卓越的性能[14]-[16]。然而，这些方法缺少的一个最重要的特征是能够从大规模数据集中自动学习知识。因此，这些方法很难从现有的大规模数据中获益。近年来，深度学习显示出很强的处理高维数据中冗余信息的能力。有很多人试图利用深度学习技术来实现SLAM/VO

[17]–[20]. 因此，本文采用了几何与深度相结合的方法

|  |
| --- |
| K、 Wang和S.Ma在中国重庆大学汽车工程学院工作，400044，同时在中国重庆大学机械传动重点实验室工作，400044（电子邮件：网址：kw@cqu.edu.cn, masai@cqu.edu.cn).  J、 陈在重庆师范大学经济与管理学院工作，重庆401331，中国（电子邮件：网址：nwpujunlan@163.com).  J、 Lu在美国密歇根州迪尔伯恩市福特汽车公司从事研究和高级工程（电子邮件：jlu10@ford.com).  2005年4月19日收到稿件；2015年8月26日修订。（通讯作者：王珂） |

基于学习的SLAM技术，因为它们之间有很好的互补性。

最近，我们目睹了视觉SLAM方法从基于几何模型的方法到基于深度学习的方法的不断发展[21]。基于深度学习的方法在解决动态对象、光照变化等挑战性问题上表现出了良好的能力。然而，这些算法的进步仅仅依赖于深度学习本身的成功，而没有探索深度学习与SLAM之间的内在机制关系。有人尝试[22]–[25]使用语义信息来改善对SLAM的感知，例如使用语义先验信息来过滤动态对象。一个很难解决的问题是，语义信息不能完全确定一个对象是动态的还是静态的，就像汽车可以移动还是静态一样。这些算法将所有选定的特征点在定位和映射任务中同等重要。

人类执行定位和映射任务的方式非常不同，它关注最显著的对象或特征，并对不同的特征或对象赋予不同的权重[26]。这被称为主动方法，它根据特定的任务，主动寻找视野中最显著的特征和对象，以满足任务的需要[27]。例如，当人们在一个新的环境中走动时，他们通常会关注显著的地标来定位位置。因此，最近，显著性和注意力的哲学在机器人界开始流行起来[28]-[30]。例如，显著性DSO[29]使用显著性模型来模拟定性人类视觉，在显著区域中选择特征点。该框架中使用的显著性模型使用SalGAN[31]在CVL-UMD数据集上进行训练。然而，这种显著性模型并不能完全描述SLAM系统应该关注的所有问题，这使得SLAM系统无法在室外环境中工作。这主要是由于人类注视数据集的单焦点和中心偏差造成的。

因此，我们首先在KITTI[32]的基础上提出了一个开源的显著数据集，即显著KITTI。不同于只关注人类注视的数据集，我们提出的数据集还考虑了几何和语义信息。然后，基于显著性KITTI数据集，利用DI-Net[33]得到显著性模型来预测SLAM系统应该关注的区域。最后，与传统的束调整（BA）方法相比，我们认为显著特征点在优化过程中应该扮演更重要的角色，就像人类处理视觉信息的方式一样。因此，我们提出了一种显著束调整（SBA）方法来模拟这一过程。总之，我们的主要贡献如下：

|  |
| --- |
|  |
|  |  |

图1。我们框架的算法概述。红色部分：贡献。

•我们为室内和室外环境提出了一个突出的SLAM框架，可应用于各种应用。

•我们提出了一种生成显著数据集的方法，我们还将使我们的显著KITTI数据集开源，以促进研究。

•我们提出了一种模拟人类视觉系统的显著束调整（SBA）方法。

•我们提出的各种实验环境下的性能对比方法。

本文的剩余部分安排如下：在第二节中，我们回顾了显著性预测和SLAM的相关工作。第三节介绍了该方法，包括显著性预测和SBA。第四节给出了实验结果，第五节给出了结论。

# 二。相关工作

## A、 显著性预测

将人类视觉显著性机制引入到计算机视觉中，不仅可以将计算资源分配给重要对象，而且可以产生更像人类视觉感知的结果。因此，视觉显著性预测已经成为一个研究热点，受到了广泛的关注[26]、[27]、[34]、[35]。根据人类视觉和注意机制，显著性预测算法可分为自下而上和自上而下两种。Itti等人提出了最早的显著性预测工作之一[36]。他们利用特征融合理论和引导搜索模型，提出了一种基于视觉显著性机制的模型。此后，计算机视觉界提出了大量的视觉显著性计算模型，预测了人眼停留在场景中某一位置的可能性[37]-[39]。这些经典的算法主要涉及自下而上的视觉显著性机制，即利用颜色、亮度和边缘特征来预测显著性图。

近年来，基于深度学习的显著性预测模型得到了很大的改进[40]-[43]。Vig等人在显著性预测中应用深度学习的早期尝试是[44]。它们利用神经网络和支持向量机学习的特征对贴片进行显著性和非显著性分类。此后，人们提出了许多有效而复杂的模型，并在精度方面取得了很大的提高。例如，DeepNet[45]、SALICON Net[46]、SalGAN[31]、DeepVS[40]、ACL Net[26]和DeepFix[41]探索了不同的网络架构和学习技术，以学习更具代表性的特性，使系统更接近人类。

|  |
| --- |
| **原始图像几何信息语义信息显著图（Ours）显著图（Original）**  图2。我们提出的语义凝视与人类凝视的比较只是根据事实。 |

在自主车辆领域，人们已经开始关注驾驶员显著性预测来解析驾驶员的注意行为，以及道路场景来预测潜在的不安全驾驶行为[28]，[30]。Deng等人[28]收集了驾驶场景中的眼球跟踪数据集，并提出了一种卷积-反卷积神经网络来预测驾驶场景中的显著性。Pal等人[30]提出了一种语义增强的注视检测方法，通过考虑深度、车速和行人过街意图来预测显著区域。然而，这些模型更多地关注车辆、行人和道路等，由于世界静态假设，这些区域的特征点在SLAM/VO任务中不稳定，鲁棒性较差。我们希望显著性模型更多地关注纹理丰富和稳定的区域，如道路标线、交通灯、交通标志，而忽略动态对象。为此，我们提出了一种结合几何和语义信息的显著性预测模型来解决这一问题。

## B、 猛击

一般来说，基于几何的方法可以分为两类：基于特征的方法和直接方法。基于特征的方法通过从图像中提取和匹配感兴趣的点来估计摄像机的姿态，而直接方法直接利用图像中的像素强度来估计摄像机的姿态，使测光误差最小化。MonoSLAM[47]是第一个使用扩展卡尔曼滤波（EKF）和Shi-Tomasi点的实时视觉SLAM系统。PTAM[48]采用非线性优化方法代替基于滤波器的方法作为后端优化方法，提出并实现了跟踪映射过程的并行化。ORB-SLAM[15]使用更稳定有效的ORB特性，并使用三个线程来构建一个完整的SLAM：跟踪、映射和循环检测线程。该系统仍然是目前最成功的SLAM系统之一，并在[9]，[16]中进行了扩展。由于在没有纹理的环境中提取不到足够的特征点，许多算法也尝试使用直线和平面特征[49]-[51]。

直接法不提取特征点，而是利用像素强度通过最小化光度误差来估计摄像机的姿态。DTAM[52]是第一个使用直接方法生成密集三维地图的系统。然而，它需要商用GPU来执行计算。为了提高效率，SVO[14]通过直接提取特征来估计摄像机的姿态和三维结构。然后，LSDSLAM[12]扩展了这项工作，它可以在大规模环境中生成半稠密映射。DSO[11]通过联合优化所有涉及的模型参数，包括几何参数、相机姿态和内部参数，实现了精度和效率之间的权衡。这项工作的扩展包括立体声[12]和环路关闭[53]。

近年来，基于深度学习的SLAM/VO凭借神经网络强大的学习能力取得了很大的进步[54]。PoseNet[55]是最早使用CNN端到端估计摄像机姿态的作品之一。此后，许多研究如雨后春笋般涌现[10]、[19]、[56]、[57]。Deep VO[58]使用RNN来建模运动动态和图像序列之间的关系。ESP-VO[19]接着扩展了这项工作，增加了姿态估计的不确定性估计。为了避免收集大量的数据，许多研究试图使用无监督的技术[59]-[61]。[62]提出了一种时间（前后）和空间（左右）约束来学习一致的姿势和三维结构，更好地利用非监督性。Almalioglu等人[63]使用生成对抗网络（GANs）进行相机姿态估计。

此外，许多研究者试图将基于几何的方法与基于深度学习的方法相结合[24]、[25]、[64]。[64]利用基于深度学习的语义分割和场景解析技术来减少图像静态和动态区域的模糊性。DynaSLAM[65]，[66]利用深度学习和多视图几何分割动态对象。DS-SLAM[25]将语义分割和运动一致性检查相结合，减少了动态对象的影响，提高了动态环境下的准确性。之后，显著DSO[29]利用视觉显著性和注意机制在显著区域提取特征点。这种方法提高了DSO的性能。然而，这个框架只能在室内环境下工作。因此，我们利用我们提出的显著性模型和显著性BA方法，将此方法推广到户外环境。

# 三、 方法论

## A、 系统总体

图1提供了我们提出的框架的简化说明，该框架由两个组件组成：基于几何的SLAM管道和基于学习的显著性预测模块。显著性预测模块生成与SLAM数据集相对应的显著性图。然后，利用显著性映射作为输入，帮助SLAM算法选择显著特征点，提高了算法的精度和鲁棒性。在我们提出的框架中，我们采用ORB-SLAM3[9]作为主干。在第三节的其余部分中，我们首先描述了如何建立一个显著性数据集，并在第三节B中使用它来获得显著性预测模型。接下来，在第三节C中，我们描述了如何使用我们提出的显著性模型预测的显著性映射来设计显著性束调整（SBA）。

|  |
| --- |
|  |
|  |  |

图3。视觉显著图预测的流水线。

## B、 显著性预测

视觉显著性或视觉注意机制是指模仿人类视觉系统，从自然场景中选取最显著、最感兴趣的区域或点，在不同的任务下进行进一步处理。在动态和奇异的环境中，忽略了动态和奇异的区域。近年来，有许多基于深度学习的方法来预测自然场景中的显著区域，并取得了很好的效果。然而，这些显著性预测方法并不能完全描述VO/SLAM系统需要注意的所有问题，这主要是由于训练数据集存在中心偏差。原因是这些方法只使用原始的人类凝视

信息，而人类的目光会停留在车辆前方的道路上，因为这是车辆行驶的地方。但是，这还不够，因为SLAM/VO还需要关注远离图像中心的区域。仅仅依靠人眼跟踪器获得的凝视数据并不能帮助捕捉所有这些重要的线索。

因此，为了解决这一问题，我们采用了与文献[30]相似的策略，将几何信息和语义信息结合起来，制作了一个显著性数据集来训练显著性模型，用语义注视代替人类注视地面的真实性。我们提出的显著性数据集基于KITTI目标检测数据集[32]。具体来说，我们首先提取图像的几何信息，如特征点、线和平面。这是因为经典的SLAM/VO方法通常集中在具有丰富几何信息的区域。接下来，我们使用SDC Net[67]在感兴趣的对象周围生成分段掩码。我们选取13个类别（红绿灯、交通标志、道路、建筑物、人行道、停车场、轨道、栅栏、桥梁、杆、杆组、植被、地形）来过滤几何信息，这些类别中的区域通常包含显著、稳定和健壮的特征。因此，动态区域中的特征不会出现在我们的显著数据集中，例如移动的车辆、人、骑手等。图2显示了我们提出的语义注视和人类注视地面真相的比较。最后，基于我们提出的显著性数据集，利用显著性图预测算法DIAlgorithm 1，得到显著性模型。

输入：输入图像I，输入激光雷达点L（可选）

输出：预测最终显著性图*S码*（最终）

1： 为每个图像提取几何信息。

2： =SDC−Net（I）；：分段任务。*性虐待性虐待*

3： 用于过滤几何信息。4： 建立显著性数据集，训练显著性预测模型（DI-Net）。*性虐待*

5： =DI-净（I）*S码*（初始化）

6： =MonoDepth（I）orD=DFuseNet（L+I）；：深度图。*D级D级*

7：对于（席，XJ）我做

8： b（最终）=aS（初始）（x）/D（x）+b*S码我，x轴日本我，x轴日本*

9:



Net[33]，并使用它来预测初始显著性图。

此外，我们还考虑到物体与自我载体之间的距离会影响显著性。就像人类更关注离自己更近的物体一样，因为他们更容易与离自己更近的物体互动。因此，我们使用单目深度估计网络MonoDepth[68]来生成深度图以获得正确的显著性图。此外，如果有激光雷达或其他深度传感器，我们可以使用深度完成网络获得更精确的深度图。图3显示了显著性预测的管道。算法中总结了生成显著图的所有步骤

1.

## C、 凸束调整

|  |
| --- |
| 图4。在SALICON和显著KITTI数据集中训练的显著性模型显著性预测的比较。 |

在第III-B节中，我们使用显著性预测模型得到每一帧的显著性图。在本节中，我们将描述如何在姿态估计过程中使用显著性映射。在经典的SLAM或VO系统中，通常在跟踪线程（仅运动BA）中对摄像机的自我运动进行BA（Bundle Adjustment）优化，在局部映射线程（local BA）中对关键帧和映射点的局部窗口进行优化。如果循环闭包可用，它还会优化循环闭包（完整BA）后的所有关键帧和贴图点。在传统的BA算法中，所有的特征点都被同等对待，使得最显著的特征点不能做更多的工作。因此，本文提出了显著束调整方法，使显著特征点充分发挥其价值。其定义如下：

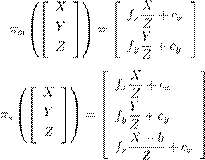
仅运动BA：优化摄像机方向∈SO（3）和平移∈R，使匹配的3D点∈Rin世界坐标与摄像机像素坐标中的关键点∈R（单目摄像机）或∈R（立体摄像机）之间的重投影误差最小化，其中∈X是所有匹配的集合。**R t Xxx公司**3*我*3 *我米*2 *我s码*3 *我*

2

{R，t}=argminRX

**放射性同位素***,*∑i∈X

其中是鲁棒Huber代价函数和与关键点尺度相关联的协方差矩阵。重投影函数定义如下：*ρ*第*π*(·)



其中（fx，fy）是焦距，（cx，cy）是主点，是基线。显著权重，定义为：As2（席，彝）+b，（·）是显著图中的像素值，为常数值。*b类威斯康辛州威斯康辛州S码a、 b类*

局部BA：优化一组共可见的关键帧KL和这些关键帧PL中的所有映射点。将Xk定义为PL中的点和关键帧中的关键点之间的匹配集，优化问题可以定义为：*k公司*

=argmin（Ek，j）X X X*ρ*

**XRt公司***我,我,我k公司*∈KL∪KF j∈Xk



# 四、 评价

在本节中，我们从两个方面验证了我们提出的算法。首先，我们比较了分别使用SALICON[69]和signific KITTI数据集训练的显著性模型的算法。其次，我们在KITTI[32]和EuRoc[70]数据集上对我们提出的算法进行了评估，并与现有的算法进行了比较。在实验过程中，它运行在一台电脑上，电脑上有nvidiageforcegtx1650gpu、amdryzen5-2600cpu和8gb内存

Linux。

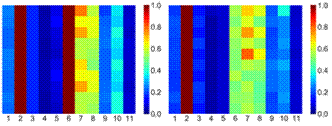
## A、 用于评估的数据集

KITTI数据集：包含从城市、农村和高速公路地区等场景收集的真实世界图像数据。每幅图像可以包含多达15辆汽车和30名行人，具有不同程度的遮挡和截断。对于里程计/SLAM，该数据集提供了自动车辆场景下的22个序列，总长度为39.2km，并提供了11个具有地面真值的序列（00-10序列）。

EuRoc数据集：是一个用于室内MAV的单目、立体和IMU数据集。它包含两个场景，一个是工业环境，另一个是普通的房间环境。这个数据集提供了11个具有基本真值的序列。根据纹理质量、场景亮度和运动速度将这些序列划分为三个难度等级。

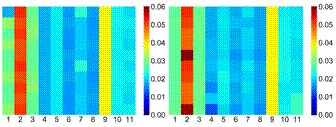
## B、 显著性模型评价

为了验证本文提出的显著性KITTI数据集的有效性，我们首先设计了一个实验，分别对SALICON和显著性KITTI数据集上训练的显著性预测模型生成的显著性图进行比较。我们在图4中提供了一些有代表性的结果。对于SALICON数据集上训练的模型，当图像中没有重要对象时，注意力集中在图像的中心，从而忽略了其他重要信息。相反，在显著KITTI数据集上训练的模型能够成功地捕获这些重要信息。此外，这些模型可以减少



|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 表一显著性预测模型的比较。   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 顺序。 |  | | RPE（单眼） | | |  | | |  | | | RPE（立体声） | | |  | | | RMSE[m] | | 平均值[m] | | | 性病科 | | | RMSE[m] | | | 平均值[m] | | | 性病科 | | | DI网 | 我们的 | DI网 | 我们的 | DI网 | | 我们的 | DI网 | | 我们的 | DI网 | | 我们的 | DI网 | | 我们的 | | 0 | 0.2345 | 0.1871 | 0.1456 | 0.1253 | 0.1839 | | 0.1389 | 0.0290 | | 0.0275 | 0.0195 | | 0.0190 | 0.0215 | | 0.0201 | | 1 | – | – | – | – | – | | – | 0.0506 | | 0.0503 | 0.0453 | | 0.0438 | 0.0226 | | 0.0305 | | 2 | 0.1578 | 0.1522 | 0.1257 | 0.1237 | 0.0954 | | 0.0937 | 0.0291 | | 0.0284 | 0.0233 | | 0.0232 | 0.0174 | | 0.0159 | | 3 | 0.0380 | 0.0374 | 0.0303 | 0.0287 | 0.0229 | | 0.0236 | 0.0178 | | 0.0172 | 0.0153 | | 0.0148 | 0.0090 | | 0.0086 | | 4 | 0.0852 | 0.0693 | 0.0681 | 0.0597 | 0.0504 | | 0.0352 | 0.0212 | | 0.0203 | 0.0189 | | 0.0179 | 0.0096 | | 0.0091 | | 5 | – | 0.5288 | – | 0.3917 | – | | 0.3549 | 0.0166 | | 0.0165 | 0.0129 | | 0.0128 | 0.0105 | | 0.0099 | | 6 | 0.6626 | 0.6246 | 0.4490 | 0.4245 | 0.4871 | | 0.4575 | 0.0207 | | 0.0185 | 0.0146 | | 0.0140 | 0.0146 | | 0.0122 | | 7 | 0.5754 | 0.5407 | 0.4117 | 0.3843 | 0.4019 | | 0.3800 | 0.0168 | | 0.0162 | 0.0139 | | 0.0137 | 0.0094 | | 0.0089 | | 8 | 0.2192 | 0.2160 | 0.1591 | 0.1580 | 0.1506 | | 0.1470 | 0.0392 | | 0.0389 | 0.0256 | | 0.0253 | 0.0297 | | 0.0296 | | 9 | 0.3551 | 0.2858 | 0.2150 | 0.1948 | 0.2820 | | 0.2090 | 0.0218 | | 0.0214 | 0.0187 | | 0.0183 | 0.0112 | | 0.0113 | | 10 | 0.0673 | 0.0665 | 0.0542 | 0.0537 | 0.0399 | | 0.0390 | 0.0206 | | 0.0200 | 0.0153 | | 0.0151 | 0.0139 | | 0.0134 | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | |

               单眼（SALICON）单眼（Ours）



                  立体声（SALICON）立体声（我们的）

图5。彩色正方形表示KITTI数据集每个序列中10个不同执行的RMSE。

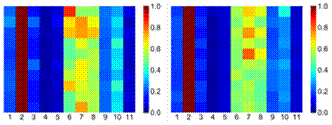
动态对象，因此具有高显著性的点通常是稳定和健壮的点，如图4中的红框所示。

然后在我们提出的框架中使用这两个显著性模型来评估其性能。在这个实验中，我们设置了两个传感器配置：单目和立体相机。我们使用evo[71]来评估具有RPE（相对姿态误差）的算法。结果列于表一，相应的热图如图5所示（每个矩形显示不同的运行）。

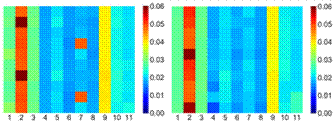
如表1所示，我们可以发现使用在显著KITTI数据集上训练的显著性模型的算法比使用在SALICON数据集上训练的显著性模型的算法获得更准确的结果，在大多数情况下，尤其是在单目传感器配置中，获得了更大的精度。在序列05中，在SALICON数据集上训练的显著性模型丢失，而我们提出的显著性模型在单目结构下可以很好地工作。这是因为大多数框架没有突出的物体，导致注意力集中在图像的中心。它们不能捕获周围区域的其他重要信息，也不能减少动态对象的影响。

## C、 KITTI数据集的评价

在本节中，我们设计了一个实验来验证SBA在ORB-SLAM3中的有效性。在单目和立体相机配置中，我们也使用evo来评估带有RPE（相对姿态误差）的算法，并绘制轨迹。实验结果如表二所示，显示了不同序列的RPE。此外，图6显示了每个评估轨迹误差的全套结果。



             ORB-SLAM3（单眼）Ours（单眼）



                ORB-SLAM3（立体声）Ours（立体声）

图6。彩色正方形表示KITTI数据集每个序列中10个不同执行的RMSE。

为了总结性能，我们在每个序列中运行这些算法10次。由于中值不受分布序列的最大值或最小值的影响，在一定程度上提高了中值对分布序列的代表性。因此，我们选择了10次处决的中位数作为最终结果

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 表二  与ORB-SLAM3算法的比较。   |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 顺序。 |  | | RPE（单眼） | | |  | | |  | | | RPE（立体声） | | |  | | | RMSE[m] | | 平均值[m] | | | 标准[m] | | | RMSE[m] | | | 平均值[m] | | | 标准[m] | | | ORB3 | 我们的 | ORB3 | 我们的 | ORB3 | | 我们的 | ORB3 | | 我们的 | ORB3 | | 我们的 | ORB3 | | 我们的 | | 0 | 0.2087 | 0.1871 | 0.1384 | 0.1253 | 0.1562 | | 0.1389 | 0.0289 | | 0.0275 | 0.0194 | | 0.0190 | 0.0213 | | 0.0201 | | 1 | – | – | – | – | – | | – | 0.0516 | | 0.0503 | 0.0450 | | 0.0438 | 0.0243 | | 0.0305 | | 2 | 0.1629 | 0.1552 | 0.1233 | 0.1237 | 0.1048 | | 0.0937 | 0.0288 | | 0.0284 | 0.0234 | | 0.0232 | 0.0169 | | 0.0159 | | 3 | 0.0384 | 0.0374 | 0.0300 | 0.0287 | 0.0239 | | 0.0236 | 0.0176 | | 0.0172 | 0.0153 | | 0.0148 | 0.0088 | | 0.0086 | | 4 | 0.0714 | 0.0693 | 0.0583 | 0.0597 | 0.0407 | | 0.0352 | 0.0207 | | 0.0203 | 0.0186 | | 0.0179 | 0.0092 | | 0.0091 | | 5 | 0.6090 | 0.5288 | 0.4229 | 0.3917 | 0.4319 | | 0.3549 | 0.0167 | | 0.0165 | 0.0131 | | 0.0128 | 0.0105 | | 0.0099 | | 6 | 0.6294 | 0.6246 | 0.4267 | 0.4245 | 0.4624 | | 0.4575 | 0.0250 | | 0.0185 | 0.0149 | | 0.0140 | 0.0192 | | 0.0122 | | 7 | 0.5785 | 0.5407 | 0.4104 | 0.3843 | 0.4075 | | 0.3800 | 0.0172 | | 0.0162 | 0.0140 | | 0.0137 | 0.0098 | | 0.0089 | | 8 | 0.2242 | 0.2160 | 0.1646 | 0.1580 | 0.1521 | | 0.1470 | 0.0392 | | 0.0389 | 0.0256 | | 0.0253 | 0.0297 | | 0.0296 | | 9 | 0.3413 | 0.2858 | 0.2304 | 0.1948 | 0.2517 | | 0.2090 | 0.0222 | | 0.0214 | 0.0189 | | 0.0183 | 0.0116 | | 0.0113 | | 10 | 0.0672 | 0.0665 | 0.0541 | 0.0537 | 0.0397 | | 0.0390 | 0.0207 | | 0.0200 | 0.0152 | | 0.0151 | 0.0140 | | 0.0134 | | 平均。 | 0.2931 | 0.2711 | 0.2059 | 0.1944 | 0.2071 | | 0.1879 | 0.0262 | | 0.0250 | 0.0203 | | 0.0198 | 0.0159 | | 0.0154 | |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |     图7。我们提出的算法在KITTI数据集的序列00、01、05、07、09上的结果是，每个序列有五个样本图像。 |

结果。在单目和立体视觉配置中，我们的系统比ORB-SLAM3更精确，因为SBA使显著特征点充分发挥其价值。因此，本实验证明利用显著图可以使我们的算法在姿态估计方面有更多的优势。这背后的直觉是，视觉显著性包含高层次的语义信息，这使得特征更加健壮。我们提出的SBA方法进一步加强了这些显著点的作用。另外，我们的显著性模型推理一幅图像只需8ms，整个系统仍然可以在我们的平台上实时运行。我们在KITTI数据集上提出的方法与groundtruth的比较结果如图7所示，每个序列有5个样本图像。可以看出，我们提出的算法在各种环境下都取得了准确的结果。因此，与只在室内和静态环境下工作的突出DSO相比，

我们提出的算法更具实际应用价值。即使在室内环境中，我们提出的算法也比大多数最先进的技术具有更高的性能（详见第IV-D节）。

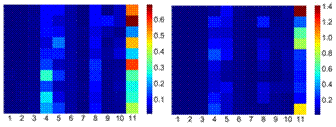
## D、 EuRoc数据集的评价

在第三节中，我们将在室内场景中评估我们的算法。

与第IV-C节不同，我们使用RMS ATE[72]测量精度，在纯单眼情况下，使用Sim（3）变换将估计轨迹与地面真值对齐。此外，我们还将我们的算法与其他最先进的算法进行了比较，如ORB-SLAM[15]、DSM[73]、DSO[11]、突出DSO[29]和ORB-SLAM3[9]。特别是对于显著DSO，该算法还利用显著性信息来估计姿态。由于DSO不能在室外工作

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| V101 V102 V103 V201 V202  图8。我们提出的算法在RuRoc数据集的序列MH01-05、V101-03和V201-02上的结果。  表III  与其他算法的比较。   |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | 序列 | MH01型 | MH02型 | MH03型 | MH04 MH05 V101型 | | | V102型 | V103型 | V201型 | V202型 | 二氧化三钒 | | 困难 | 电子 | 电子 | 米 | D-D-E | | | 米 | D级 | 电子 | 米 | D级 | |  |  |  |  | 单眼（ATE[m]） | | |  |  |  |  |  | | 圆满贯 | 0.0710 | 0.0670 | 0.0710 | 0.0820 | 0.0600 | 0.0150 | 0.0200 | - | 0.0210 | 0.0180 | - | | DSO公司 | 0.0460 | 0.0460 | 0.1720 | 3.8100 | 0.1100 | 0.0890 | 0.1070 | 0.9030 | 0.0440 | 0.1320 | 1.1520 | | 需求侧管理 | 0.0390 | 0.0360 | 0.0550 | 0.0570 | 0.0670 | 0.0950 | 0.0590 | 0.0760 | 0.0560 | 0.0570 | 0.7840 | | 凸极DSO | 0.0412 | 0.0435 | 0.1522 | 4.5629 | 0.0951 | 0.0963 | 0.6774 | 0.7562 | 0.0709 | 0.1322 | 1.2380 | | ORB-SLAM3型 | 0.0170 | 0.0170 | 0.0310 | 0.0660 | 0.0440 | 0.0330 | 0.0160 | 0.0370 | 0.0210 | 0.0220 | - | | 我们的 | 0.0144 | 0.0294 | 0.0261 | 0.0510 | 0.0408 | 0.0328 | 0.0124 | 0.0190 | 0.0129 | 0.0165 | - | |  |  |  |  | 立体声（ATE[m]） | | |  |  |  |  |  | | 球体-SLAM2 | 0.0350 | 0.0180 | 0.0280 | 0.1190     0.0600       0.0350 | | | 0.0200 | 0.0480 | 0.0370 | 0.0350 | – | | VINS融合 | 0.5400 | 0.4600 | 0.3300 | 0.7800     0.5000       0.5500 | | | 0.2300 | – | 0.2300 | 0.2000 | – | | ORB-SLAM3型 | 0.0250 | 0.0220 | 0.0270 | 0.0890     0.0580       0.0350 | | | 0.0210 | 0.0490 | 0.0320 | 0.0270 | 0.3610 | | 我们的 | 0.0137 | 0.0152 | 0.0207 | 0.1386     0.0878       0.0329 | | | 0.0169 | 0.0914 | 0.0213 | 0.0347 | 0.3526 | |

在不同的环境下，我们比较了它在室内环境下的性能，以证明我们提出的方法的有效性。结果如表三所示，相应的热图如图9所示（每个矩形显示不同的运行）。图8显示了EuRoc数据集上具有地面真值的一些轨迹结果。



                      单眼立体

图9。彩色序列表示EuRoc数据集每个序列中10个不同执行的ATE。

如表三所示，在大多数序列中，我们提出的算法比文献中的最新技术获得更准确的结果。我们还发现SalientDSO比DSO更精确，而我们的比ORB-SLAM3更精确。因此，将显著性模型与传统的单目视觉SLAM或VO相结合，可以提高系统的性能。然而，由于显著性DSO中使用的显著性模型存在中心偏差，该算法在室外环境中由于没有局部显著性目标进行显著性预测而导致跟踪丢失。相比之下，我们的算法在室内和室外都能很好地工作，并且在大多数序列中都能获得更精确的结果。然而，由于运动速度快，光照弱，我们提出的算法在复杂序列上的性能较差。这是因为我们的框架仍然采用ORB-SLAM3特征提取方法，没有利用显著性信息来改进特征点的提取过程，这也是我们今后工作的一个方向。此外，本实验所使用的显著性预测模型是在不包含室内场景的显著性KITTI数据集上训练的。这可能是结果不佳的另一个原因。

# 五、 结论

在本文中，我们提出了一个基于显著性的SLAM框架，基线来自ORB-SLAM3。受显著性和注意原理的启发，我们利用显著性预测模型来产生显著性信息来模拟人类视觉系统。基于显著性信息，提出了一种能充分发挥显著特征价值的SBA方法。我们对KITTI数据集和EuRoc数据集进行了全面的定性评估，以证明使用显著信息可以提高准确性和鲁棒性。此外，我们还开放了KITTI数据集的源代码。作为未来的研究方向，我们将提出一个紧密耦合的框架，利用显著性信息来改进特征提取线程，甚至映射线程。此外，如何从IMU或Lidar数据中提取显著性信息并应用于SLAM/VO也是未来的发展方向。

## 致谢

感谢国家自然科学基金（批准号：51605054）、重庆市人工智能技术重点技术创新项目（批准号：cstc2017rgznzdyfX0039）、重庆市社会科学规划项目（批准号：2018QNJJ16）的资助，中央高校基础研究基金（编号：2019CDXYQC003）。

## 参考文献

[1] C.Cadena、L.Carlone、H.Carrillo、Y.Latif、D.Scaramuzza、J.Neira、I.Reid和J.J.Leonard，“同时定位和绘图的过去、现在和未来：走向稳健感知时代”，Ieee机器人学报，第32卷，第6期，第1309-1332页，2016年。[在线]。可用：hGotoISIi://WOS:000389849700001

[2] 王国强，马世安，陈建民，卢俊卿，“面向复杂和新兴领域的深度视觉里程计的挑战与应用”，IEEE认知与发展系统学报，第1-12020页。

[3] 《自动驾驶模式识别研讨会》（the Ieee S.Abdamman，C.Abdamman，2018），第360页。[在线]。可用：hGotoISIi://WOS:

000457636800055https://ieeexplore.ieee.org/ielx7/8575058/8575239/

08575523.pdf？tp=&arnumber=8575523&isnumber=8575239

[4] S.W.Yang、S.A.Scherer、X.D.Yi和A.Zell，“微型飞行器自主导航的多摄像机视觉冲击”，机器人与自主系统，第93卷，第116-134页，2017年。[在线]。可用：hGotoISIi://WOS:000403027600010

[5] M.Turan、Y.Almalioglu、H.Gilbert、H.Araujo、T.Cemgil和M.Sitti，“内窥镜胶囊机器人的内窥镜融合：基于粒子滤波的带切换状态空间模型的多传感器数据融合”，2018 Ieee机器人学和自动化国际会议（Icra），第5393-5400页，2018年。[在线]。可用：hGotoISIi://WOS:000446394504010

[6] M.Turan、E.P.Ornek、N.Ibrahimli、C.Giracoglu、Y.Almalioglu、M.F.Yanik和M.Sitti，“内窥镜胶囊机器人的无监督里程计和深度学习”，2018 Ieee/Rsj智能机器人和系统国际会议（Iros），第1801-1807页，2018年。[在线]。可用：hGotoISIi://WOS:000458872701133

[7] 李宝莲、秦天琴、胡伯泰、朱福耀、沈圣杰，

“2017年国际单目视觉增强现实研讨会”，第11-21页。[在线]。可用：hGotoISIi://WOS:000435259700002

[8] K.Nagao、M.Yang、X.Cao和Y.Miyakawa，“建筑规模虚拟现实：扩展真实世界的另一种方式”，2019年IEEE多媒体信息处理和检索会议（MIPR）。IEEE，会议记录，第205-211页。

[9] C.Campos、R.Elvira、J.J.G.Rodr´ıguez、J.M.M.Montiel和J.D.Tardos，“Orb-slam3：视觉、视觉惯性和多地图slam的精确开源库”，arXiv预印本附件十四：2007.11898, 2020.

[10] G.Costante和M.Mancini，“数据驱动视觉里程计的不确定性估计”，IEEE机器人学报，第1-20页，2020年。

[11] J.Engel、V.Koltun和D.Cremers，“直接稀疏里程计”，IEEE模式分析和机器智能学报，第40卷，第3期，611-625页，2018年。[在线]。可用：https://www.ncbi.nlm.nih.gov/pubmed/28422651

[12] 《欧洲单目视觉会议》，J。斯普林格，《会议记录》，第834-849页。

[13] 薛国辉，王晓辉，李国强，王国强，王国强，查浩，“超越追踪：深度视觉里程计的记忆选择与姿态优化”，载IEEE计算机视觉与模式识别会议论文集，会议论文集，第8575-8583页。

[14] J.Engel、J.Sturm和D.Cremers，“单目相机的半密集视觉里程计”，《IEEE计算机视觉国际会议论文集》，第1449-1456页。

[15] R.Mur Artal、J.M.M.Montiel和J.D.Tardos，“Orb slam：一个多功能和精确的单眼slam系统”，IEEE机器人学报，第31卷，第5期，第1147-1163页，2015年。[在线]。可用：hGotoISIi://WOS:000364504400006

[16] R.Mur Artal和J.D.Tardos，“Orb-slam2：单目、立体和rgb-D相机的开源slam系统”，IEEE机器人学报，第33卷，第5期，第1255-12622017页。[在线]。可用：

hGotoISIi://WOS:000412235700019

[17] M.A.Esfahani，H.Wang，K.Y.Wu，和S.H.Yuan，“ABOLDEPIO:一种新型的自主车辆深度惯性里程计网络”，Ieee智能交通系统学报，第21卷，第5期，1941-1950页，2020年。[在线]。可用：hGotoISIi://WOS: 000532285400010

[18] A.Wang、Z.Fang、Y.Gao、S.Tan、S.Wang、S.Ma和J.N.Hwang，“深度和自我运动联合优化的对抗式学习”，IEEE图像处理学报，第29卷，第4130-4142页，2020年。[在线]。可用：https://www.scopus.com/inward/record。uri？eid=2-s2.0-85079493155&doi=10.1109%2fTIP.2020.2968751&partnerID=40&md5=fdcb22e110bbe00e99b10fd7f2b47801https:

//ieeexplore.ieee.org/document/8972902/

[19] S.Wang，R.Clark，H.K.Wen和N.Trigoni，“端到端，通过深度神经网络的序列到序列概率视觉里程计”，《国际机器人研究杂志》，第37卷，第4-5期，第513-5422018页。[在线]。可用：hGotoISIi://WOS: 000432134700008

[20] N.Yang，L.v.Stumberg，R.Wang和D.Cremers，“D3vo:单目视觉里程计的深度、深度姿势和深度不确定性”，《IEEE/CVF计算机视觉和模式识别会议论文集》，第1281-1292页。

[21]R.H.Li，S.Wang和D.B.Gu，“视觉大满贯从几何到深度学习的持续演变：挑战和机遇”，认知计算，第10卷，第6期，第875-889页，2018年。[在线]。

可用：hGotoISIi://WOS：00045334480001[22]第121–128页，2019年。

《2018年国际机器人技术会议》（IEEE64G.Tofarson）第23页，第64G页。[在线]。可用：hGotoISIi://WOS:000446394504135

[24]肖立华，王建国，邱小S，容子荣，邹学德，

“动态slam：动态环境下基于深度学习的语义单目视觉定位和映射”，《机器人与自主系统》，第117卷，第1-16页，2019年。[在线]。可用：hGotoISIi://WOS:000469898500001https://www.sciencedirect。com/science/article/abs/pii/S0921889018308029？通过%3Dihub

[25]余春贤、刘春贤、刘春杰、谢福国、杨勇、魏秋飞，“Ds slam：面向动态环境的语义视觉slam”，2018 Ieee/Rsj国际智能机器人与系统会议（Iros），第1168-1174页，2018年。[在线]。可用：hGotoISIi://WOS:000458872701043

[26]W.Wang，J.Shen，J.Xie，M.-M.Cheng，H.Ling，A.Borji，“重温深度学习时代的视频显著性预测”，IEEE模式分析和机器智能学报，2019年。

[27]A.Borji和L.Itti，“视觉注意力建模的最新技术”，Ieee模式分析和机器智能学报，第35卷，

第1期，第185-207页，2013年。[在线]。可用：hGotoISIi://WOS: 000311127700017

[28]T.Deng、H.Yan、L.Qin、T.Ngo和B.S.Manjunath，“驾驶员如何分配他们的潜在注意力？通过卷积神经网络驱动固定预测，“IEEE智能交通系统学报，第21卷，第5期，第2146-21542020页。

[29]H.-J.Liang，N.J.Sanket，C.Fermuller，和Y.Aloimonos，“Salientdso:关注直接稀疏里程计”，IEEE自动化科学与工程学报，第16卷，第4期，第1619-16262019页。

[30]A.Pal、S.Mondal和H.I.Christensen，“看正确的东西”自主驾驶的引导语义凝视，“IEEE/CVF计算机视觉和模式识别会议论文集，会议论文集，第11883-11892页。

[31]J.Pan、C.C.Ferrer、K.McGuinness、N.E.O&apos;Connor、J.Torres、E.Sayrol和X.Giro-i Nieto，“Salgan：生成性对抗网络的视觉显著性预测”，arXiv预印本附件十四：1701.01081, 2017.

[32]U.R.Geiger A，Lenz P，“我们准备好自动驾驶了吗？kitti视觉基准套件，“IEEE计算机视觉和模式识别会议（CVPR），2012年。

[33]S.Yang，G.Lin，Q.Jiang和W.Lin，“视觉显著性预测的扩展初始网络”，IEEE多媒体学报，2019年。

[34]A.Borji，“深度学习时代的显著性预测：成功与局限”，IEEE模式分析与机器智能学报，2019年。

[35]A.Borji、D.N.Sihite和L.Itti，“视觉显著性建模中人形模型一致性的定量分析：比较研究”，IEEE图像处理学报，第22卷，第1期，第55-69页，2012年。[36]L.Itti，C.Koch和E.Niebur，“快速场景分析中基于显著性的视觉注意模型”，IEEE模式分析和机器智能学报，第20卷，第11期，第1254-12591998页。

[37]X.Hou和L.Zhang，“显著性检测：光谱残差法”，2007年IEEE计算机视觉和模式识别会议。Ieee，会议记录，第1-8页。

[38]O.Le Meur、P.Le Callet、D.Barba和D.Thoreau，“一种自下而上视觉注意模型的连贯计算方法”，Ieee模式分析和机器智能学报，第28卷，第5期，802–8172006页。[在线]。可用：hGotoISIi://WOS: 000235885700011

[39]L.Zhang，M.H.Tong，T.K.Marks，H.Shan，G.W.Cottrell，“Sun:基于自然统计的贝叶斯显著性框架”，《视觉杂志》，第8卷，第7期，第32-32页，2008年。

[40]L.Jiang，M.Xu，T.Liu，M.Qiao，Z.Wang，“Deepvs:一种基于深度学习的视频显著性预测方法”，《欧洲计算机视觉会议论文集》，会议论文集，第602-617页。

[41]S.S.S.Kruthiventi、K.Ayush和R.V.Babu，“Deepfix：预测人眼注视的完全卷积神经网络”，IEEE图像处理学报，第26卷，第9期，4446-4456页，2017年。

[42]G.Leifman、D.Rudoy、T.Swedish、E.Bayro Corrochano和R.Raskar，“从深度学习凝视转换以改进视频显著性估计”，《IEEE计算机视觉国际会议论文集》，会议论文集，1698-1707页。

[43]W.Wang和J.Shen，“通过注意框预测和美学评估进行深裁剪”，IEEE计算机视觉国际会议论文集，会议论文集，第2186-2194页。

[44]E.Vig、M.Dorr和D.Cox，“自然图像显著性预测层次特征的大规模优化”，《IEEE计算机视觉和模式识别会议论文集》，第2798-2805页。

[45]J.Pan，E.Sayrol，X.Giro-i Nieto，K.McGuinness和N.E.O&apos;Connor，“显著性预测的浅层和深层卷积网络”，IEEE计算机视觉和模式识别会议论文集，会议论文集，第598-606页。

[46]第262–270页，2015年。

[47]A.J.Davison、I.D.Reid、N.D.Molton和O.Stasse，“单镜头碰撞：实时单镜头碰撞”，IEEE Trans-Pattern Anal-Mach Intell，第29卷，第6期，第1052-67页，2007年。[在线]。可用：https://www.ncbi.nlm.nih.gov/pubmed/17431302

[48]G.Klein和D.Murray，“小型ar工作空间的并行跟踪和映射”，2007年第6届IEEE和ACM混合与增强现实国际研讨会论文集。IEEE计算机学会，会议记录，第1-10页。

[49]A.Vakhitov、V.Lempitsky和Y.Zheng，“直线和点特征三胞胎的立体相对姿势”，《欧洲计算机视觉会议论文集》（ECCV），《会议论文集》，第648-663页。

[50]张建勇，曾国强，查海波，“人工环境中平面和直线的结构感知撞击”，《模式识别快报》，第127卷，181-1902019页。[在线]。可用：hGotoISIi://WOS: 000493892700022

[51]Y.Zhao和P.A.Vela，“良好的线切割：实现线辅助vo/vslam的精确姿态跟踪”，《欧洲计算机视觉会议论文集》，第516-531页。

[52]R.A.Newcombe、S.J.Lovegrove和A.J.Davison，“Dtam：实时密集跟踪和绘图”，2011年计算机视觉国际会议。IEEE，会议记录，第2320-2327页。

[53]X.Gao，R.Wang，N.Demmel和D.Cremers，“Ldso：带回路闭合的直接稀疏里程计”，2018 IEEE/RSJ智能机器人和系统国际会议（IROS）。IEEE，会议记录，第2198-2204页。

[54]C.Chen，B.Wang，C.X.Lu，N.Trigoni和A.Markham，“定位和制图的深度学习调查：走向空间机器智能时代”，arXiv预印本附件十四：2006.12567, 2020.

[55]A.Kendall、M.Grimes和R.Cipolla，“Posenet:实时6自由度相机重定位的卷积网络”，2015 Ieee计算机视觉国际会议（Iccv），第2938-29462015页。[在线]。可用：hGotoISIi://WOS:

000380414100328https://ieeexplore.ieee.org/ielx7/7407725/7410356/

## 07410693.pdf文件？tp=&arnumber=7410693&isnumber=7410356&ref=

[56]B.Li，D.Zou，D.Sartori，L.Pei，和W.Yu，“Textslam:具有平面文本特征的视觉slam”，2020 IEEE机器人技术和自动化国际会议（ICRA）。IEEE，会议记录，第2102-2108页。

[57]H.Li，W.Chen，J.Zhao，J.-C.Bazin，L.Luo，Z.Liu，和Y.-H.Liu，“用于单目视觉里程测量的绝对相机姿态的鲁棒有效估计”，IEEE国际机器人与自动化会议（ICRA），2020年。IEEE，会议记录，第2675-2681页。

[58]S.Wang，R.Clark，H.Wen和N.Trigoni，“Deepvo:使用深度递归卷积神经网络实现端到端视觉里程计”，2017 IEEE机器人技术与自动化国际会议（ICRA）。IEEE，会议记录，第2043-2050页。

[59]V.M.Babu、K.Das、A.Majumdar和S.Kumar，“亡灵：

深度和自我运动估计的无监督深度网络，“2018

*Ieee/Rsj智能机器人与系统国际会议（Iros）*，第1082-1088页，2018年。[在线]。可用：hGotoISIi://WOS: 000458872701031

[60]R.H.Li，S.Wang，Z.Q.Long和D.B.Gu，“Undeepvo：通过无监督深度学习进行单目视觉里程测量”，2018 Ieee机器人与自动化国际会议（Icra），第7286–72912018页。[在线]。可用：hGotoISIi://WOS:000446394505079

[61]T.H.Zhou，M.Brown，N.Snavly和D.G.Lowe，“从视频中无监督学习深度和自我运动”，第30届Ieee计算机视觉和模式识别会议（Cvpr 2017），第6612–+，2017年。[在线]。可用：hGotoISIi://WOS:000418371406075

[62]C.Godard、O.Mac Aodha和G.J.Brostow，“具有左右一致性的无监督单目深度估计”，第30届Ieee计算机视觉和模式识别会议（Cvpr 2017），第6602–6611页，2017年。[在线]。可用：hGotoISIi://WOS:000418371406074

[63]Y.Almalioglu、M.R.U.Saputra、P.P.de Gusmao、A.Markham和N.J.A.P.A.Trigoni，“Ganvo:无监督的深单目视觉里程计和生成对抗网络深度估计”，2018年。

[64]N.Brasch、A.Bozic、J.Lallemand和F.Tombari，“高动态环境中的语义单眼碰撞”，2018 IEEE/RSJ智能机器人和系统国际会议（IROS）。《IEEE会议记录》，第393-400页。

[65]B.Bescos、C.Campos、J.D.Tardos和J.J.a.p.a.Neira，“Dynaslam´ii:紧耦合多目标跟踪和slam”，2020年。

[66]B.Bescos、J.M.Facil、J.Civera、J.J.I.R.Neira和A.Letters，“动态场景中的动态跟踪、映射和修复”，第3卷，第4期，4076-4083页，2018年。

[67]Y.Zhu，K.Sapra，F.A.Reda，K.J.Shih，S.Newsam，A.Tao和B.Catanzaro，“通过视频传播和标签松弛改进语义分割”，IEEE计算机视觉和模式识别会议论文集，会议论文集，第8856-8865页。

[68]C.Godard、O.Mac Aodha和G.J.Brostow，“具有左右一致性的无监督单目深度估计”，《IEEE计算机视觉和模式识别会议论文集》，会议论文集，第270-279页。

[69]M.Jiang，S.Huang，J.Duan，and Q.Zhao，“Salicon:语境中的显著性”，《IEEE计算机视觉与模式识别会议论文集》，第1072-1080页。

[70]M.Burri、J.Nikolic、P.Gohl、T.Schneider、J.Rehder、S.Omari、M.W.Achtelik和R.Siegwart，“欧洲微型飞行器数据集，”

*国际机器人研究杂志*，第35卷，第10期，第1157页-

1163, 2016. [在线]。可用：hGotoISIi://WOS:000382981300001

[71] 2017. [在线]。可用：https://github.com/MichaelGrupp/evo

[72]J.Sturm、N.Engelhard、F.Endres、W.Burgard和D.Cremers，“rgb-D slam系统评估基准”，2012 IEEE/RSJ智能机器人和系统国际会议。IEEE，会议记录，第573-580页。

[73]J.Zubizarreta、I.Aguinaga和J.M.J.I.T.o.R.Montiel，“直接稀疏映射”，2020年。