3D Object Proposals for Accurate Object Class Detection

用于精确目标类别检测的 3D 目标提案

论文引用：

Chen X, Kundu K, Zhu Y, et al. 3d object proposals for accurate object class detection[C]//Advances in Neural Information Processing Systems(NIPS). 2015: 424-432.

ABSTRACT

The goal of this paper is to generate high-quality 3D object proposals in the context of autonomous driving. Our method exploits stereo imagery to place proposals in the form of 3D bounding boxes. We formulate the problem as minimizing an energy function encoding object size priors, ground plane as well as several depth informed features that reason about free space, point cloud densities and distance to the ground. Our experiments show significant performance gains over existing RGB and RGB-D object proposal methods on the challenging KITTI benchmark. Combined with convolutional neural net (CNN) scoring, our approach outperforms all existing results on all three KITTI object classes.

摘要

本文的目标是在自动驾驶的背景下生成高质量的 3D 目标建议。我们的方法利用立体图像以 3D 边界框的形式表示提案。我们将问题表述为最小化编码物体尺寸先验，地平面以及几个已知深度特征的能量函数，这些特征推断包括自由空间，点云密度和到地面的距离。我们的实验显示，在具有挑战性的 KITTI 基准测试中，现有的 RGB 和 RGB-D 对象建议方法的性能显着提升。结合卷积神经网络（CNN）评分，我们的方法优于所有三个 KITTI 对象类的所有现有结果。

1. Introduction

Due to the development of advanced warning systems, cameras are available onboard of almost every new car produced in the last few years. Computer vision provides a very cost effective solution not only to improve safety, but also to one of the holy grails of AI, fully autonomous self-driving cars. In this paper we are interested in 2D and 3D object detection for autonomous driving.

1. 引文

由于先进的警告系统的发展，几乎所有在过去几年生产的新车上都有摄像机。计算机视觉提供了一种非常具有成本效益的解决方案，不仅可以提高安全性，还可以提供人工智能的全自动驾驶汽车。在本文中，我们对自动驾驶的 2D 和 3D 物体检测感兴趣。

With the large success of deep learning in the past years, the object detection community shifted from simple appearance scoring on exhaustive sliding windows [1] to more powerful, multi-layer visual representations [2, 3] extracted from a smaller set of object/region proposals [4, 5]. This resulted in over 20% absolute performance gains [6, 7] on the PASCAL VOC benchmark [8].

随着过去几年深度学习的巨大成功，对象检测社区从详尽的滑动窗口[1]上的简单外观评分转变为从较小的一组对象/区域提案[4,5]中提取的更强大的多层视觉表示[2,3]。这导致 PASCAL VOC 基准[8]测试的绝对性能增加超过 20％[6,7]。

The motivation behind these bottom-up grouping approaches is to provide a moderate number of region proposals among which at least a few accurately cover the ground-truth objects. These approaches typically over-segment an image into super pixels and group them based on several similarity measures [4, 5]. This is the strategy behind Selective Search [4], which is used in most state-of-the-art detectors these days. Contours in the image have also been exploited in order to locate object proposal boxes [9]. Another successful approach is to frame the problem as

这些自下而上的分组方法背后的动机是提供适度数量的区域提案，其中至少有一些准确地涵盖了实际真值对象。这些方法通常将图像过度分割为超级像素，并基于几种相似性度量对它们进行分组[4,5]。这是选择性搜索背后的策略[4]，目前在大多数最先进的探测器中使用。图像中的轮廓也被利用以定位对象提案框[9]。另一种成功的方法是将问题框定为能量最小化，其中参数化的能量族表示分组的各种偏差，从而产生多种不同的解决方案[10]。

energy minimization where a parametrized family of energies represents various biases for grouping, thus yielding multiple diverse solutions [10].

Interestingly, the state-of-the-art R-CNN approach [6] does not work well on the autonomous driving benchmark KITTI [11], falling significantly behind the current top performers [12, 13]. This is due to the low achievable recall of the underlying box proposals on this benchmark. KITTI images contain many small objects, severe occlusion, high saturated areas and shadows. Furthermore, KITTI's evaluation requires a much higher overlap with ground-truth for cars in order for a detection to count as correct. Since most existing object/region proposal methods rely on grouping super pixels based on intensity and texture, they fail in these challenging conditions.

有趣的是，最先进的 R-CNN 方法[6]在自动驾驶基准 KITTI [11]上效果不佳，显着落后于当前最佳表现者[12,13]。这是由于该基准的基础框提案的可实现性较低。KITTI 图像包含许多小物体，严重遮挡，高饱和区域和阴影。此外，KITTI 的评估要求与汽车的真值重叠得更高，以便将检测视为正确。由于大多数现有的对象/区域提案方法依赖于基于强度和纹理对超像素进行分组，因此它们在这些挑战性条件下失败。

In this paper, we propose a new object proposal approach that exploits stereo information as well as contextual models specific to the domain of autonomous driving. Our method reasons in 3D and places proposals in the form of 3D bounding boxes. We exploit object size priors, ground plane, as well as several depth informed features such as free space, point densities inside the box, visibility and distance to the ground. Our experiments show a significant improvement in achievable recall over the state-of-the-art at all overlap thresholds and object occlusion levels, demonstrating that our approach produces highly accurate object proposals. In particular, we achieve a 25% higher recall for 2K proposals than the state-of-the-art RGB-D method MCG-D [14]. Combined with CNN scoring, our method outperforms all published results on object detection for Car, Cyclist and Pedestrian on KITTI [11]. Our code and data are online: http://www.cs.toronto.edu/~3dop.

在本文中，我们提出了一种新的目标提案方法，该方法利用立体信息以及特定于自动驾驶领域的上下文模型。我们的方法以 3D 形式出现，并以 3D 边界框的形式提出提案。我们利用物体尺寸先验，地平面以及几个已知深度特征，如自由空间，盒内点密度，能见度和到地面的距离。我们的实验表明，在所有重叠阈值和物体遮挡水平上，对现有技术的可实现有显着改善的召回，这表明我们的方法可以产生高度准确的目标提案。特别是，与最先进的 RGB-D 方法 MCG-D [14]相比，我们对 2K 提案的召回率提高了 25％。结合 CNN 评分，我们的方法优于所有公布的关于 KITTI 上汽车，自行车和行人的物体检测结果[11]。我们的代码和数据在线：http：//www.cs.toronto.edu/~3dop。

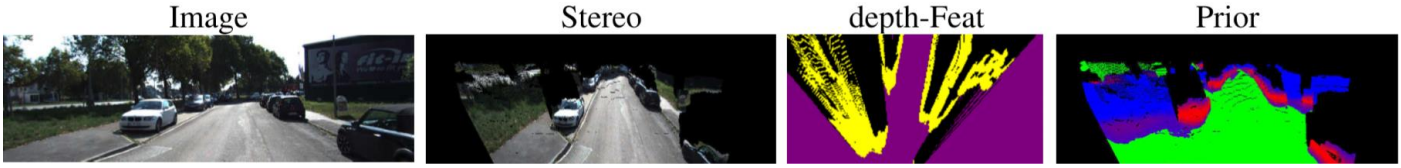

Figure 1. ***Features:*** *From left to right: original image, stereo reconstruction, depth-based features and our prior. In the third image, purple is free space(F in Eq.(2)) and occupancy is yellow(S in Eq.(1)). In the prior, the ground plane is green and red to blue indicates distance to the ground.*

图 1. 特征：从左到右：原始图像，立体图像重建，基于深度的特征和我们的先验。在第三幅图像中，紫色是自由空间(F in Eq.(2))，占据为黄色(S in Eq.(1))。在先验特征中，地平面是绿色，红色到蓝色表示到地面的距离。