

基于深度学习的自动化小行星搜索

孙畅, 邹桐, 周佳颖

(南京外国语学校, 江苏省南京市玄武区北京东路 30 号)

摘 要

为了以后能摆烂而创造了一个模板，为了展现转行效果而开始啊对对对对对对对对对对对对

关键词：摆烂、啊对对对

Abstract

Attention! If you input "different", the computer will output "different", but if you input "dif{}ferent", the computer will output "different"

目录

1	引言	3
2	数据集构建	4
2.1	数据选取	4
2.2	图像预处理	5
2.3	训练目标	6
3	深度学习模型构建	6
3.1	研究背景	6

1 引言

小行星，一般指太阳系内的一类天体。该类天体类似行星环绕太阳运动，但体积和质量比行星小得多。小行星一般被认为是由太阳系形成时期的微行星演变而来，是目前发现数量最多的太阳系天体。

了解小行星的位置和轨道参数细节，对于人类在地球上的生存安全有极为重要的意义。1994 年，苏梅克-列维九号彗星在分裂成 21 颗碎片后撞击木星，其中最大的一颗碎片直径达 35 公里，产生了明显的撞击坑。据估计，这次撞击相当于 10 亿颗原子弹同时爆炸的当量，对木星大气的影响直到三个月后才基本恢复。如果这样的撞击发生在地球上，将会对地球大气产生极为显著的影响。自此之后，IAU 建立起完备的小行星观测、报备、核验系统，确认了相当数量的小行星，大大增加了我们对于地球所处的宇宙环境的了解。



Figure 1: 宇宙中的小行星示意图

此外，包括小行星在内的太阳系小天体已成为人类了解太阳系的起源和发展的重要观测对象，对于人类揭开恒星系演化过程具有重要的意义。

截至目前，尽管已有相当多的太阳系内的小行星被发现、编号，仍然有大量的太阳系内小天体未被发现。世界各地的先进天文观测望远镜每天都在产生体量极大的观测数据，但是对于观测数据的处理、筛选、分析，仍然处于相对较低的水平。尽管在观测图像中存在相当数量的小行星踪迹，但由于拍摄出的照片体量相当大，且分析手段仍存在

优化空间，所以仍可能有大量已经被拍摄到的小行星被我们忽略。

传统的小行星检测算法往往从小行星本身的物理特性出发，结合本观测站的观测条件，精心设计出特定的检测算法。然而，这类算法在不同的观测设备、观测方法和观测条件下迁移较为困难，且几乎所有的大型观测项目都不提供开源的小行星检测算法，在实际使用中存在较多限制。

一些大型观测项目逐渐意识到，在现有条件下，小行星观测数据往往存在图像分析能力不够、精确度不高的问题。于是，他们发起了一系列公众科学项目 (如 IASC 和 Hubble Asteroid Hunter)，将望远镜产生的数据分发给参与项目的志愿者，借助社会力量来补充现有的数据分析手段。在这些公众科学项目中，参与项目的志愿者往往需要通过肉眼进行小行星检测，总体效率很低，且准确度无法保证。

针对以上问题，本文提出了使用深度学习方法来进行小行星搜寻工作。我们拟采取对经典的图像分类卷积神经网络进行微调的方法，实现高效、自动化小行星搜索工作。

2 数据集构建

2.1 数据选取

从观测的角度来说，地面观测通常受到昼夜更替、晴夜数量、夜天光情况和大气视宁度等因素的影响，对小行星的巡天观测条件比较苛刻。所以，选取拍摄质量更高的空间望远镜数据或将成为更好的选择。哈勃空间望远镜作为部署在大气层外的先进光学望远镜，具有视场大、观测能力强、干扰较少、不受地表天气因素约束等优势，是进行小行星搜寻的良好器材。

为能够捕获到星际中微弱的电磁波信息，哈勃空间望远镜采用的是对指定区域的长时间曝光 (平均每张照片曝光 35 分钟) 的观测方法。在长时间曝光下，距离更远的恒星、星系、星云往往不会发生明显的位置变化，所以在图像中能够看到一个清晰的像。由于小行星是太阳系内天体，离望远镜距离要近很多，在一次曝光中会产生明显的位移，所以表现在照片上是一条线状轨迹。我们展示了几张哈勃望远镜的观测数据 (Figure ???)。

哈勃官方团队为提高小行星搜索检测的准确度，发起了“Hubble Asteroid Hunter”公众科学项目。该项目借助公众力量和机器学习方法，发现了 1701 条小行星轨迹，其中 670 条被确认为已知的小行星。在项目结束后，哈勃团队公开了 1701 条数据的观测号、曝光时间、图像中小行星轨迹起始和终止位置等信息，为深度学习提供了较高质量

的训练数据来源。该数据集是目前我们已知的唯一公开的标记数据集。

我们提供了哈勃团队提供表格的局部行列 (Table 1)。表格中提供的小行星轨迹是使用天球坐标系上的 RA(经度) 和 Dec(纬度) 来描述的, 单位是度, 这是天文中常用的坐标系统, 但是不适合我们进行进一步的深度学习训练。在后续处理中, 我们将会把天球坐标系中的坐标转换成图像上的像素点坐标。



Figure 2: 哈勃空间望远镜

Table 1: Hubble Asteroid Hunter 数据集样例

哈勃望远镜观测号	轨迹曝光时间 (秒)	轨迹起始位置 RA	轨迹起始位置 Dec	轨迹终止位置 RA	轨迹终止位置 Dec
ib1901010	2520.0	146.7329823	10.0969372	146.72596	10.0971402
ib2r03020	1340.0	57.8625451	28.3094616	57.866537	28.3105017
ib4801010	1772.0	135.3464886	18.2331854	135.3422318	18.2361665
ib4803010	3686.0	135.0326681	22.5615199	135.0202912	22.549794
ib4803020	2799.0	135.0172415	22.5440068	135.0042467	22.5320243

2.2 图像预处理

尽管该数据集提供了含有小行星照片的详细参数信息, 但没有直接提供可以用于训练的哈勃望远镜观测照片。所以, 我们设计了一套处理流程, 实现由 Python 代码自动化完成数据批量下载、剪裁、导出并可视化等工作。以下是预处理工作的操作流程:

-
1. 读取 Hubble 团队提供的“tablea1.dat”文件，获得指定观测号、小行星位置等关键信息
 2. 使用 Python 中的 astroquery 第三方库，调用相关 API，下载 Mikulski Archive for Space Telescopes (MAST) 数据库中指定观测号的数据¹
 3. 提取天文标准数据格式 FITS 文件中的图像数据，处理成 Numpy Array 的格式
 4. 对小行星轨迹标记的位置进行坐标转化。将原先天球坐标系统中的坐标转换成图像上的像素坐标，方便后续处理
 5. 根据现有的小行星位置信息，引入 x, y 方向上的随机偏移，获取 1000*1000 像素点尺寸的包含小行星 (positive 类别) 数据集。在这种构造方式下，可以在确保图像上保留全部或部分的小行星轨迹的同时，实现图像上的小行星轨迹随机出现，模拟真实应用场景下的图像

哈勃团队只提供了含有小行星的图像观测号，而没有提供不含小行星的观测号。为保证不包含小行星 (negative 类别) 数据集的纯净性，我们从这些确保经过人工筛查的图像中，在远离小行星标记位置随机选取 1000*1000 像素点大小的图像，构成 negative 类别数据集。

2.3 训练目标

尽管“小行星搜索”看起来是一个目标检测问题，但这应该首先是一个图像分类问题。跟据文献给出的数据统计，在所有哈勃观测

3 深度学习模型构建

3.1 研究背景

近年来，随着深度学习，尤其是卷积神经网络的快速发展，越来越多的天文领域问题开始使用深度学习的方式进行解决。在图像分类领域，一系列令人振奋的成果涌现，大大提高了自动化图像分类技术的实现精度，拓宽了卷积神经网络的应用范畴。

1. 啊对

¹<https://astroquery.readthedocs.io/en/latest/mast/mast.html>

(a) 太对辣

(b) 好对捏

2. 啊对对

3. 啊对对对²

至臻无双 [1]

²变成光守护麻衣学姐

参考文献

- [1] Ronald A. Remillard and Jeffrey E. McClintock. X-Ray Properties of Black-Hole Binaries. *Annual Review of Astronomy and Astrophysics*, 44(1):49–92, 2006. __eprint: <https://doi.org/10.1146/annurev.astro.44.051905.092532>.