隐式3D表征学习的星表障碍物检测方法

杨文飞[1]，姜 涵[1]，潘晓扬[2]，李茂登[3,4]，周 晔[1]，张天柱[1,2]

（1. 中国科学技术大学 信息科学技术学院，合肥 230031；2. 深空探测实验室 未来技术研究院，合肥 230088；
3. 深空探测实验室，北京 100097；4. 北京控制工程研究所，北京 100094）

摘 要：传统的基于图像的障碍物检测只能定位其在图像平面的二维位置，需再结合双目立体匹配获取深度信息才可确定障碍物的实际空间位置，双目立体匹配具有计算量大且面临复杂环境匹配准确性下降的难题，提出一种基于隐式3D表征学习的星表障碍物检测方法。该方法将每个点潜在的三维坐标编码为图像特征，生成的隐式三维空间特征能有效建立2D图像到3D空间的隐式转换，从而直接预测障碍物的空间位置。 并在“勇气号”（Spirit）采集的火星地表图像进行了实验验证，结果表明所提出的方法可有效地识别地外天体表面岩石障碍物的位置和尺寸，检测准确率达到了85.5%。所提方法为星表障碍物的检测提供了新思路，有望为月球/火星探测器自主巡视探测提供技术支撑。

关键词：地外星表障碍物检测；3D位置编码；3D目标检测

中图分类号：TP18 **文献标识码：**A **文章编号：**2096-9287(2025)02-0172-07

DOI:10.15982/j.issn.2096-9287.2025.20240044

引用格式：杨文飞，姜涵，潘晓扬，等. 隐式3D表征学习的星表障碍物检测方法[J]. 深空探测学报（中英文），2025，12（2）：172-178.

Reference format: YANG W F，JIANG H，PAN X Y，et al. Implicit 3D representation learning for extraterrestrial obstacle detection[J]. Journal of Deep Space Exploration，2025，12（2）：172-178.

引 言

巡视器可在地外天体表面进行移动、探测和采样，是代替人类在地外天体开展科学探测活动最理想的工具之一。然而巡视器在地外天体表面移动时，不可避免地会遇到岩石等各种障碍物，如何准确地识别和定位这些障碍物对于保障巡视器的安全、提升探测效能至关重要。

当前地外天体表面障碍物的检测主要有基于激光雷达的主动扫描和基于光学相机的被动感知方法。由于激光雷达能耗高、体积大、质量重，主要应用于地外天体着陆过程中的落区感知和障碍物规避[1]。与之相比，光学相机具有功耗低、体积小、重量轻、稳定性高等优点，是地外天体巡视器广泛应用的导航设备。2004年美国国家航空航天局（National Aeronautics and Space Administration，NASA）发射的“勇气号”（Spirit）和“机遇号”（Opportunity）火星车[2]共安装了9个光学相机，其中6个都用于导航避障。美国第3代[3]火星车“好奇号”（Curiosity）配备了一对导航相机和4对避障相机用于火星表面地形的测量。此外，中国首台火星车“祝融号”配备了两对黑白避障相机作为障碍物检测的主要传感器[4]。因此，国内外学者和工程团队在基于相机输入的障碍物检测技术方面做了大量的探索。目前基于图像的障碍物检测主要分为基于人工特征和基于深度学习的检测方法。基于人工特征的检测方法通常依赖于图像的颜色、纹理、边缘等低层次特征进行障碍物的识别和定位[5]。这类方法在处理复杂场景时，可能会受到光照变化、阴影、遮挡等因素的影响，导致识别效果不佳。此外，对于地外天体表面形态各异的岩石障碍物，提取有效的人工特征并进行准确分类也是一项挑战。

近年来，随着深度学习技术的飞速发展，基于深度学习的障碍物检测方法逐渐成为研究热点。例如，Colon等[6]利用卷积UNet网络对输入图中的障碍物进行检测，Liu等[7]在此基础上将卷积UNet网络改进为基于Transformer的UNet网络，提升了对障碍物的检测性能。然而在当前的深空探测任务中，基于图像的障碍物检测大多只能定位障碍物在图像平面的二维位置，需再结合其它测量手段如双目相机之间立体匹配[8-11]获取深度信息才能确定障碍物的3D空间位置。同时现有匹配精度较高的方法面临计算量大的难题，严重影响

收稿日期：2024-07-04 修回日期：2024-12-04

基金项目：国防基础科学研究（JCKY2021130B016）

避障算法的实时性。此外在纹理单一或光照差异大的探测环境时，现有立体匹配方法准确性会大幅下降，严重影响避障算法的鲁棒性。因此，基于匹配的障碍物检测算法难以满足未来国际月球科研站等大范围星表巡视任务需求，亟需开展更高效、鲁棒的障碍物识别方法。

受地面自动驾驶领域方法的启发[12-14]，本文提出了一种基于2D图像输入的障碍物3D位置检测方法。与传统预测2D位置的方法相比，该方法无需利用双目立体匹配等手段来获取目标的深度信息，可基于双目相机的图像直接输出障碍物3D位置。本文提出了一种基于隐式三维表征学习的障碍物检测方法。该方法通过建立从相机光心出发的锥形3D空间，并在该空间内利用3D位置编码将输入的2D特征建模为隐式的3D空间表征，使得神经网络建立起从2D图像到3D空间的隐式转换、编码和表达能力，从而能直接感知输入图像所代表的三维空间并预测障碍物所处的空间位置。为验证该方法的有效性，在“勇气号”和“机遇号”所采集的火星地表图像数据集[1]进行实验，结果表明所提方法能够有效地识别地外天体表面的岩石障碍物的3D位置和尺寸，为巡视器的自主避障软件设计提供了新的思路，具有一定的应用价值。

1　隐式3D表征学习方法

1.1　整体框架

本文提出了一种基于隐式3D表征学习的星表障碍物检测方法，方法总体框架如图1所示。给定巡视器的双目图像对(I_L, I_R)，首先采用特征提取器（ResNet50[15]或ResNet101[15]）提取双目图像的2D特征。在3D坐标生成器中，将相机视锥空间离散化为一个3D网格。然后，通过相机参数将网格的坐标进行转换，生成3D世界空间的坐标。生成的3D坐标与2D多视图特征一起输入到隐式3D特征编码器，从而获得隐式的3D表征。最后，将3D表征进一步输入到Transformer解码器，并与由查询生成器生成的对象查询进行交互，更新后的对象查询用于预测对象类别以及3D边界框。

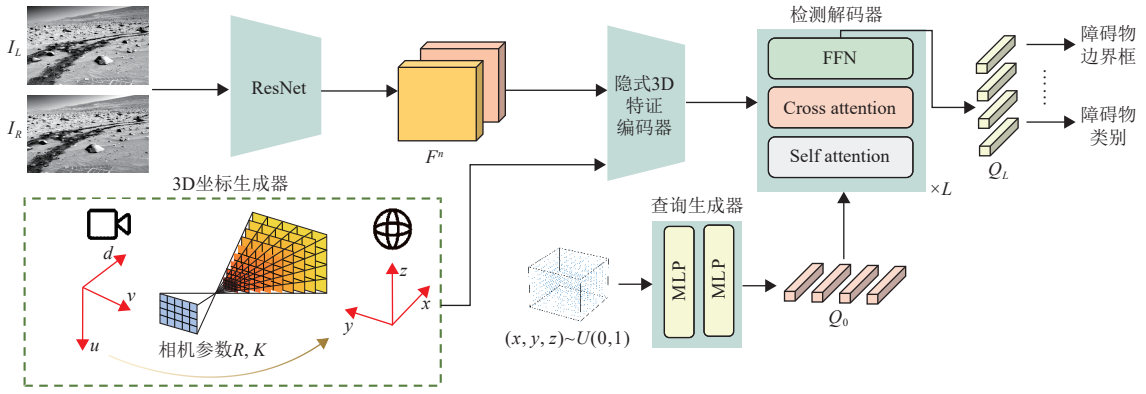

图 1　方法框架图

Fig. 1　Framework of proposed method

1.2　隐式3D 语义建模

本文提出在相机视锥空间构建一个连续的隐式语义场，用于高效表征障碍物的空间分布。该语义场不依赖显式点云或体素表示，而是通过将图像特征与三维空间位置编码进行融合，获得对空间位置具有感知能力的语义特征表示。具体而言图像平面的每一个像素点(x,y)都对应着相机视锥空间的一条三维射线，其参数化形式可表示为

$$r(x,y;t) = o + t \cdot d_{x,y} \tag{1}$$

其中：o为相机光心；t为在射线方向的深度参数；$d_{x,y}$为该像素对应的方向向量。

在射线空间定义一个连续映射函数：$\varphi : \mathbb{S} \to \mathbb{R}^C$，其中，$\mathbb{S}$为射线域，即所有起点为相机光心、方向合法的射线集合，$\varphi(r)$为对射线的空间编码。在此基础上，构建一个由射线编码与图像特征共同生成的三维语义表示为

$$\hat{F}(r) = f(\varphi(r), F(x,y)) \tag{2}$$

其中：$F(x,y)$为图像特征图中像素点(x,y)的特征向量；$f : \mathbb{R}^C \to \mathbb{R}^C$为连续的特征融合函数。由于$\varphi$和$f$为连续函数，图像特征$F$在空间中具有局部平滑性，因此，复合函数$\hat{F}$在射线空间中亦保持连续性。

$$\| r - r' \| \to 0 \Rightarrow \| \hat{F}(r) - \hat{F}(r') \| \to 0 \tag{3}$$

从而确保了语义场在空间表达的连续性与一致

性，为后续的障碍物检测提供了理论基础。

1.3 3D坐标生成器

在3D坐标生成器中，本文首先将相机视锥空间离散成三维网格。具体而言网格中的每个点可表示为$p_{i,j}=(u_i\times d_j, v_i\times d_j, d_j), j=1,\cdots,D$，其中$(u_i,v_i)$代表的是图像平面的像素坐标，$\{d_j\}_{j=1}^{D}$为一系列预定义的深度值。则每个网格点$p_{i,j}$对应的3D坐标$P_{i,j}^{n}=(x_{i,j},y_{i,j},z_{i,j},1)$可根据每个相机的投影矩阵计算。

$$[x_{i,j},y_{i,j},z_{i,j},1]^{\mathrm{T}}=K_n^{-1}\cdot R_n^{-1}d_{\mathrm{j}}\cdot[u_i,v_i,1]^{\mathrm{T}} \tag{4}$$

其中：$n\in\{L,R\}$，$K_n\in\mathbb{R}^{3\times4}$，$R_n\in\mathbb{R}^{4\times4}$分别为左右目相机标记、相机的内参和外参，$i\in\{1,2,\cdots H\times W\}$，$H$、$W$分别为2D特征图的高和宽。进一步，对生成的3D坐标进行正则化。

$$\begin{cases} x_{i,j}=(x_{i,j}-x_{\min})/(x_{\max}-x_{i,j}) \\ y_{i,j}=(y_{i,j}-y_{\min})/(y_{\max}-y_{i,j}) \\ z_{i,j}=(z_{i,j}-z_{\min})/(z_{\max}-z_{i,j}) \end{cases} \tag{5}$$

其中：$[x_{\min},x_{\max},y_{\min},y_{\max},z_{\min},z_{\max}]$为3D空间需检测的区域范围。正则化后的坐标集合可写为

$$P^n=\{P^n(u_i,v_i),i=1,2,\cdots H\times W\} \tag{6}$$

其中：$P^n(u_i,v_i)=[P_{i,1}^n\ \ P_{i,2}^n\ \ \cdots\ \ P_{i,D}^n]$为一个$D\times4$维的向量。

1.4 隐式3D位置编码器

本模块的结构如图2所示，其目的是将输入的2D特征建模与3D位置信息相关联，获得隐式的3D空间表征。具体来说，给定一对双目图像$I_n,n\in\{L,R\}$，分别采用特征提取器和3D坐标生成器生成对应的2D特征F^n和3D坐标P^n，首先将P^n输入到多层感知器，并且转换成3D位置嵌入。然后，通过一个1 × 1卷积层将2D特征进行转换，并且与3D位置编码相加，形成位置感知的隐式3D表征。

$$PE^n=\varphi(P^n) \tag{7}$$

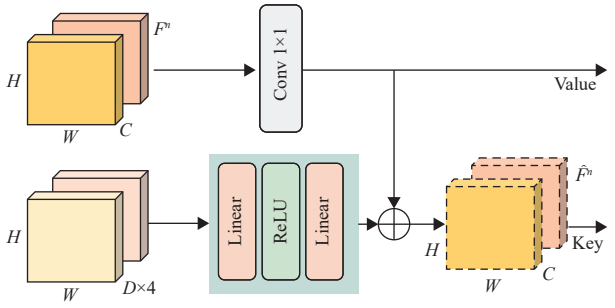

图 2 隐式3D编码器的结构

Fig. 2 The structure of 3D position encoder

$$\hat{F}^n=Conv_{1\times1}(F^n)+PE^n \tag{8}$$

生成位置感知的3D特征和原始的2D特征分别作为后续检测解码器的键分量和值分量。

为展示三维位置编码的效果，本文在左视图选取两个位置编码点，并计算这两个点与所有视角的位置编码之间的相似性，结果如图3 所示，与这些点相邻的区域通常具有更高的相似性。如在左视图选取右侧的点时，双目相机获取的图片重合度较高，因此右视图中的对应位置也具有较高的响应值。这表明3D 位置编码在三维空间隐式地建立了不同视角之间的位置关联关系。

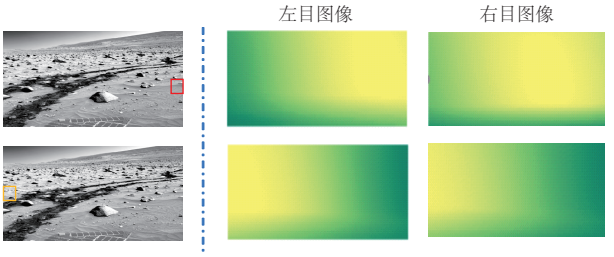

图 3 3D位置编码相似性

Fig. 3 The similarity of 3D position embedding

1.5 查询生成器和解码器

原始的DETR[16]直接使用一组可学习的参数作为初始对象查询。为缓解3D场景中收敛困难的问题，首先在3D世界空间使用均匀分布0~1初始化一组可学习的锚点。然后3D锚点的坐标被输入到一个具有两个线性层的小型多层感知机，并生成初始对象查询Q_0。对于解码器而言，遵循DETR的标准Transformer解码器，其中包括L个解码器层。解码器层的交互过程可表述为

$$Q_l=\Omega_l\left(Q_{l-1},\hat{F}^n,F^n\right),\ l=1,\cdots,L \tag{9}$$

其中：Ω_l为解码器的第l层，$Q_l\in\mathbb{R}^{M\times C}$为第$l$层的更新对象查询，其中$M$和$C$分别为查询的数量和通道数。在每个解码器层中，对象查询通过多头注意力和前馈网络与位置感知的3D 隐式表征进行交互。经过迭代交互，更新后的对象查询可用于预测地表障碍物的3D位置。

1.6 检测头和损失

检测头主要包括两个分支，用于分类和回归。来自解码器更新后的对象查询被输入到检测头，并预测对象类别的概率及3D边界框，其中回归分支相对于锚点的坐标预测相对偏移量。设$y=(c,b)$和$\hat{y}=(\hat{c},\hat{b})$分别表示真实值集合和预测值集合。匈牙利算法[17]用于真实标签与预测之间的标签分配。假设σ为最优分配函数，则3D物体检测的损失可总结为

$$L(y,\hat{y}) = \lambda_{cls}L_{cls}(c,\sigma(\hat{c})) + L_{reg}(b,\sigma(\hat{b})) \tag{10}$$

$$L_{cls}(c,\sigma(\hat{c})) = \begin{cases} -\alpha(1-p)^{\gamma}\log(p), & if\ \ \sigma(\hat{c}) = 1 \\ -(1-\alpha)p^{\gamma}\log(1-p), & \text{otherwise} \end{cases} \tag{11}$$

$$L_{reg} = |b - \sigma(\hat{b})| \tag{12}$$

其中：p为预测的分类概率；γ、α、λ_{cls}分别为预先定义的超参数。

2　实验验证及分析

2.1　数据预处理

“勇气号”和“机遇号”火星车通过立体导航相机（Navcam）等获取了大量的火星地面图像。本研究以“勇气号”火星车为例，从 MER Analyst’s Notebook网站（http：//an.rsl.wustl.edu/mer/mera/mera.htm）下载200对导航相机拍摄的双目图像对以及派生的三维点云数据构建数据集，图4展示了其中的5组图像。这些数据由美国喷气动力实验室（Jet Propulsion Laboratory，JPL）的多任务图像处理实验室（Multimission Image Processing Laboratory，MIPL）通过其软件流水线自动生成。根据三维点云数据，采用KITTI的标注格式为每张图片生成一个.txt格式的标注文档，每个标注物体的标签描述如表1所示。

此外，美国喷气动力实验室还为每张图片提供了CAHVOR相机标定模型。该模型通过扩展传统的针孔相机模型来处理畸变，从而提供更高精度的相机标定结果。CAHVOR代表的含义为：**C**（Camera center）：相机的光心或投影中心，即相机镜头的几何中心；**A**（Axis）：相机光轴的方向，即相机的主光轴方向向量；**H**（Horizontal vector）：水平方向向量，用于定义图像的水平方向；**V**（Vertical vector）：垂直方向向量，用于定义图像的垂直方向；**O**（Optical axis vector）：光轴矢量，描述了从光心到图像平面的光线方向；**R**（Radial distortion coefficients）：径向畸变系数，用于描述镜头的径向畸变效应。

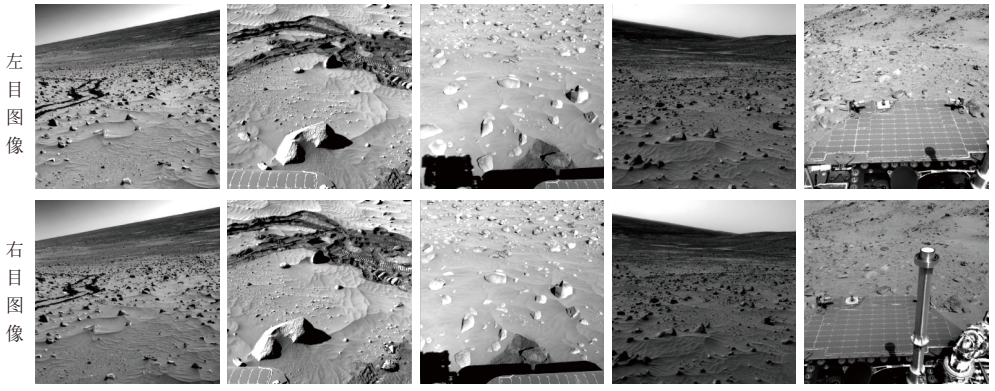

图 4　火星表面图像

Fig. 4　Mars surface image

表 1　数据标签描述

Table 1　Label data description

名称	描述
类别	物体类别
2D检测框	物体在图像上的2D检测框，可以用左上(c_l,c_t)和右下(c_r,c_b)的像素坐标表示
位置	相机坐标系下物体中心的3D坐标x,y,z
尺度	物体的3D尺度：高度h、宽度w、长度l

根据每张图片给出的CAHVOR模型可以计算得到对应的相机内参K和外参R[18]，其具体公式为

$$h_c = A \cdot H \tag{13}$$

$$v_c = A \cdot V \tag{14}$$

$$h_s = \|A \times H\| \tag{15}$$

$$v_s = \|A \times H\| \tag{16}$$

$$H_1 = \frac{H - h_c A}{h_s} \tag{17}$$

$$V_1 = \frac{V - v_c A}{h_c} \tag{18}$$

相机外参R可表示为

$$R = [H_1; V_1; A] \tag{19}$$

相机内参K可表示为

$$K = \begin{bmatrix} h_s & 0 & h_c \\ 0 & v_s & v_c \\ 0 & 0 & 1 \end{bmatrix} \tag{20}$$

2.2 实验过程

本文采用ResNet50和ResNet101[15]作为特征提取器，C5特征（第5阶段的输出）被上采样并与C4特征（第4阶段的输出）融合以产生最终的2D特征。在3D坐标生成器中，采用线性增加离散化（Linear Increase Discretion，LID）的方法沿着深度轴采样64个点。障碍物在X轴和Y轴的检测范围[−61.2 m, 61.2 m]，Z轴为[−10 m, 10 m]。使用AdamW[19]优化器进行训练，权重衰减为0.01。学习率初始化为2.0×10^{-4}，并采用余弦退火策略进行衰减。在3D空间中，实例的真实标签被随机旋转，旋转范围为[−22.5°, 22.5°]。超参数$\gamma=2$，$\alpha=0.25$，$\lambda_{cls}=2$。所有实验在4个GeForce RTX 3090 GPU上以批量大小为4，并进行500个周期的训练。

2.3 实验结果

本文在多种分辨率以及多个骨干网络报告了检测障碍物的平均精度（Average precision，AP），平均平移误差（Absolute Trajectory Error，ATE）以及平均比例尺误差（Average Scale Error，ASE）。其中平均精度类似于二维目标检测中测量精度和召回率，计算方式是基于二维中心距离在真实框的匹配。平均精度的计算公式为

$$AP=\sum_{i=1}^{n-1}(r_{i+1}-r_i)\cdot p(r_{i+1}) \tag{21}$$

其中：r_i为第 i 个召回率值；$p(r_{i+1})$为对应的精度值；n为预先定义的离散召回率点的数量

平均平移误差通过计算目标中心点的偏差来衡量目标检测模型在位置预测的误差。对于每个检测目标，其中心点位置与真实值的误差公式为

$$TE=\sqrt{(x_{\text{pred}}-x_{\text{gt}})^2+(y_{\text{pred}}-y_{\text{gt}})^2+(z_{\text{pred}}-z_{\text{gt}})^2} \tag{22}$$

其中：$(x_{\text{pred}},y_{\text{pred}},z_{\text{pred}})$为预测的目标中心点坐标；$(x_{\text{gt}},y_{\text{gt}},z_{\text{gt}})$为真实的目标中心点坐标。

平均平移误差的计算公式为

$$ATE=\frac{1}{M}\sum_{j=1}^{M}TE_j \tag{23}$$

其中：M为检测目标的数量。

平均比例尺误差衡量模型在物体尺寸预测的误差。对于每个检测目标，其尺寸预测误差为

$$SE=\frac{1}{3}\left(\frac{|l_{\text{pred}}-l_{\text{gt}}|}{l_{\text{gt}}}+\frac{|w_{\text{pred}}-w_{\text{gt}}|}{w_{\text{gt}}}+\frac{|h_{\text{pred}}-h_{\text{gt}}|}{h_{\text{gt}}}\right) \tag{24}$$

其中：l_{pred}、w_{pred}、h_{pred}分别为预测的物体长度、宽度和高度；l_{gt}、w_{gt}、h_{gt}分别为真实的物体长度、宽度和高度。

平均比例尺误差的计算公式为

$$ASE=\frac{1}{M}\sum_{j=1}^{M}SE_j \tag{25}$$

其中：M为检测目标的数量。

本文所提方法的实验结果如表2所示。可以看出，在图像分辨率1 024 × 1 024并采用ResNet101为骨干网络的时候，平均检测精度可达到最优的85.5%。当分辨率下降至256 × 256时，障碍物检测的平均精度大约有4.6%的下降。本方法的检测结果如图5所示。较大且距离较近的障碍物可被很好地检测出来，一些较小的障碍物不会阻碍巡视器的前进，因此无论在标注还是检测过程中均被忽略。此外，红色的虚线圈内展示了检测错误的情况，在同一个检测框内包含了多个较小的障碍物。

表 2 障碍物检测结果

Table 2 Results of obstacle detection

骨干网络	分辨率	AP/%↑	ATE/m↓	ASE/m↓
ResNet50	256 × 256	78.1	33.8	33.5
ResNet50	1 024 × 1 024	84.9	22.6	19.9
ResNet101	256 × 256	80.9	26.7	29.8
ResNet101	1 024 × 1 024	85.5	21.2	20.9

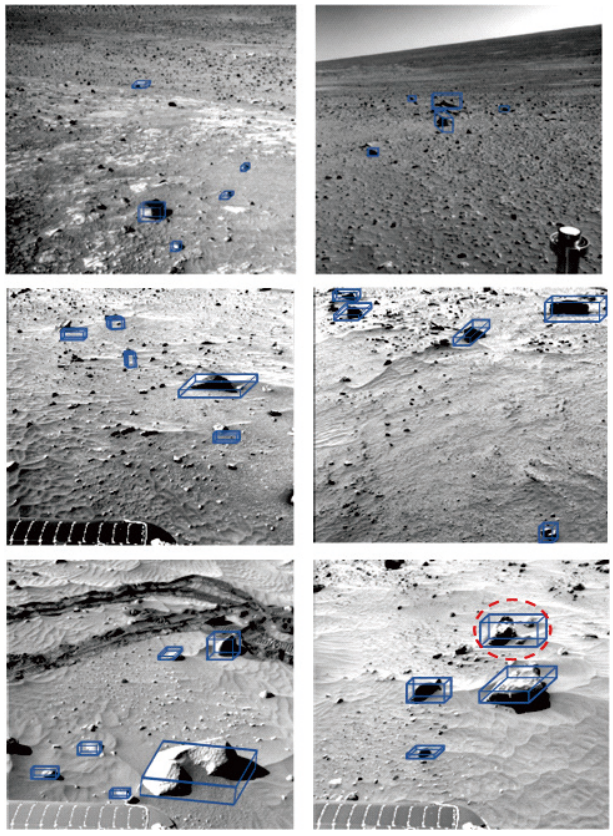

图 5 障碍物检测结果

Fig. 5 Results of obstacle detection

2.4　性能对比

在 3D 障碍物检测任务中，检测准确性至关重要，尤其是在复杂多变的未知环境中。高精度的检测可有效减少误检和漏检，确保巡视器能够准确识别和定位障碍物。同时，较高的推理速度使巡视器迅速地识别周围的障碍物，从而及时采取必要的行动，确保任务安全。因此，本文采用ResNet101作为特征提取网络，不同方法处理一对图像的推理时间、推理速度和检测性能如表3所示。其中，FPS （Frames Per Second） 表示模型每秒处理的图像帧数。本文提出的方法具有85.5%AP，相比DSGN（Deep Stereo Geometry Network）具有2.9%的性能提升。同时本文方法处理一对双目图像的时间仅为0.265 s。而基于双目立体匹配的方法DSGN则需要0.682 s，其计算瓶颈主要在于需要对构建的3D几何体积进行3D卷积。此外，Stereo RPN 基于卷积神经网络 （Convolutional Neural Networks，CNN）的两阶段目标检测器，而本文方法采用了端到端的Transformer 架构。尽管两者推理速度相近，本文方法的检测性能相比于Stereo RPN提升了 8.7%。

表 3　实验性能比较

Table 3　Comparison of experimental performance

对比方法	推理时间/s↓	推理速度/FPS↑	AP/%↑
Stereo RPN[19]	0.280	3.6	76.8
DSGN[20]	0.682	1.5	82.6
本文方法	0.265	3.7	85.5

2.5　3D位置编码的影响

为证明利用3D位置编码构建隐式3D表征方式的有效性，本文采用ResNet101作为特征提取网络，在表4评估了不同位置编码对障碍物检测结果的影响。当使用DETR标准的2D位置编码时，模型精度只能收敛到49.3%，在2D位置编码加入双目视角先验可得到6.6%的提升。本文采用的方法可直接达到85.5% 的AP，这表明3D位置编码为生成的隐式3D表征提供了强大的位置先验，从而感知3D场景。

表 4　位置编码的影响

Table 4　Impact of position embeddings

位置编码	AP/%↑	ATE/m↓	ASE/m↓
2D位置编码	49.3	71.8	40.4
多视角先验的2D位置编码	55.9	68.5	32.2
3D位置编码（所提方法）	85.5	21.2	20.9

3　结　论

本文针对巡视器在地外天体表面移动时的自主探测需求，提出了一种基于隐式3D表征学习的星表障碍物检测方法。该方法将三维坐标的位置信息编码为图像特征，生成的隐式三维空间特征能有效建立2D图像到3D空间的隐式转换，从而直接预测障碍物的空间位置。该方法同时兼顾高精度以及高推理速度，以“勇气号”采集的火星地表图像检验可达85.5%的平均精度，且处理一对双目图像的时间仅为0.265s。

参考文献

[1] 王立，刘洋，华宝成，等. 嫦娥五号探测器自主着陆视觉避障方法与评价[J]. 宇航学报，2021，42(8)：975-981.
WANG L，LIU Y，HUA B C，et al. Evaluation for Chang’e-5 visual autonomous hazard avoidance landing method[J]. Journal of Astronautics，2021，42(8)：975-981.

[2] DI K C，XU F L，WANG J，et al. Photogrammetric processing of rover imagery of the 2003 Mars exploration Rover mission[J]. ISPRS Journal of Photogrammetry and Remote Sensing，2008，63(2)：181-201.

[3] 邢琰，魏春岭，汤亮，等. 地外巡视探测无人系统自主感知与操控技术发展综述[J]. 空间控制技术与应用，2021，47(6)：1-8.
XING Y，WEI C L，TANG L，et al. Development of autonomous sensing and control technology for extraterrestrial mobile exploration unmanned systems[J]. Aerospace Control and Application，2021，47(6)：1-8

[4] 王荣本，顾柏园，郭烈，等. 月球环境感知中的石块识别方法研究[J]. 计算机工程，2006，32(15)：174-175.
WANG R B，GU B Y，GUO L，et al. Study on method of stone detection in field of Moon environment exploration[J]. Computer Engineering，2006，32(15)：174-175.

[5] 贾阳，申振荣，庞彧，等. 月面巡视探测器地面试验方法与技术综述[J]. 航天器环境工程，2014，31(5)：464-469.
JIA Y，SHEN Z R，PANG Y. A review of field test methods and technologies for lunar rover[J]. Spacecraft Environment Engineering，2014，31(5)：464-469.

[6] COLON F F，RUBIO-ESPINO E，AZUELA J H S，et al. Rock detection in a Mars-like environment using a CNN[C]//Proceedings of Pattern Recognition：11th Mexican Conference，MCPR 2019. Querétaro，Mexico：Springer International Publishing，2019.

[7] LIU H Q，YAO M B，XIAO X M，et al. RockFormer：a u-shaped transformer network for martian rock segmentation[J]. IEEE Transactions on Geoscience and Remote Sensing，2023，61：1-16.

[8] XU G W，CHENG J D，GUO P，et al. Attention concatenation volume for accurate and efficient stereo matching[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. New Orleans，LA，USA：IEEE，2022.

[9] 刘少创，贾阳，马友青，等. 嫦娥三号月面巡视探测器高精度定位[J]. 科学通报，2015，60(4)：372-378.
LIU S C，JIA Y，MA Y Q，et al. High precision localization of the Chang’E-3 lunar rover[J]. Chinese Science Bulletin，2015，60(4)：372-378.

[10] 陈建新，邢琰，滕宝毅，等. 嫦娥三号巡视器 GNC 及地面试验技术[J]. 中国科学：技术科学，2014(5)：461-469.
CHEN J X，XING Y，TENG B Y，et al. Guidance，navigation and control technologies of Chang’E-3 lunar rover[J]. Scientia Sinica Technologica，2014(5)：461-469.

[11] 吴伟仁，王大轶，邢琰，等. 月球车巡视探测的双目视觉里程算法与

实验研究[J]. 中国科学：信息科学，2011，41(12)：1415-1422.
WU W R，WANG D Y，XING Y，et al. Binocular visual odometry algorithm and experimentation research for the lunar rover[J]. Science in China(Information Sciences)，2011，41(12)：1415-1422.

[12] WANG T，ZHU X G，PANG J M，et al. Fcos3D：fully convolutional one-stage monocular 3D object detection[EB/OL]. (2021-4-22)[2024-7-4]. https://arxiv.org/abs/2104.10956.

[13] LI Z Q，WANG W H，LI H Y，et al. Bevformer：learning bird's-eye-view representation from multi-camera images via spatiotemporal transformers[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence，2025，47(3)：2020-2036.

[14] CARION N，MASSA F，SYNNAEVE G，et al. End-to-end object detection with transformers[C]//Proceedings of European conference on Computer Vision. Switzerland：Springer International Publishing，2020.

[15] LI P L，CHEN X Z，SHEN S J. Stereo R-CNN based 3D object detection for autonomous driving[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Long Beach，CA，USA：IEEE，2019.

[16] DI K C，LI R X. CAHVOR camera model and its photogrammetric conversion for planetary applications[J]. Journal of Geophysical Research：Planets，2004，109(E4)：1-9.

[17] LOSHCHILOV I，HUTTER F. Decoupled weight decay regularization[EB/OL]. (2017-11-14)[2024-7-4]. https://arxiv.org/abs/1711.05101.

[18] KUHN H W. The Hungarian method for the assignment problem[J]. Naval Research Logistics Quarterly，1955，2(1-2)：83-97.

[19] HE K M，ZHANG X Y，REN S Q，et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas，NV，USA：IEEE，2016.

[20] CHEN Y L，LIU S，SHEN X Y，et al. DSGN：deep stereo geometry network for 3D object detection[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. Seattle. WA，USA：IEEE，2020.

作者简介：
杨文飞(1994－)，男，副研究员，主要研究方向：计算机视觉、深空人工智能理论与方法。
通信地址：合肥市中国科学技术大学高新校区(230031)
E-mail：yangwf@ustc.edu.cn
张天柱(1982－)，男，教授，博士生导师，主要研究方向：计算机视觉、深空人工智能理论与方法。**本文通信作者**。
通信地址：合肥市中国科学技术大学高新校区(230031)
E-mail：tzzhang@ustc.edu.cn

Implicit 3D Representation Learning for Extraterrestrial Obstacle Detection

YANG Wenfei[1]，JIANG Han[1]，PAN Xiaoyang[2]，LI Maodeng[3,4]，ZHOU Ye[1]，ZHANG Tianzhu[1,2]

（1. School of Information Science and Technology，University of Science and Technology of China，Hefei 230031，China；
2. Institute of Future Technology，Deep Space Exploration Laboratory，Hefei 230088，China；
3. Deep Space Exploration Lab，Beijing100097，China；
4. Beijing Institute of Control Engineering，Beijing 100094，China）

Abstract：Based on traditional image-based obstacle detection methods can only locate obstacles in 2D image plane，requiring additional measurement methods such as stereo matching to obtain depth information and then determine the 3D positions of obstacles. However，stereo matching faces challenges of high computational cost and decreased accuracy when dealing with complex environments. Therefore，we propose an implicit 3D representation learning method for extraterrestrial obstacle detection was proposed. It encodes the potential three-dimensional coordinates of each point into image features，and the generated features can effectively establish an implicit conversion from 2D images to 3D space，thereby enabling direct prediction of the 3D positions of obstacles. Experiments conducted on Mars surface images collected by the Spirit rover demonstrate that the proposed method can effectively identify locations and sizes of obstacles，achieving 85.5% average precision. The proposed method in this study presents an innovative framework for planetary surface obstacle detection，with substantial potential to advance autonomous navigation capabilities in lunar/Martian exploration rovers.

Keywords：extraterrestrial obstacle detection；3D position embedding；3D object detection.

Highlights：

- The proposed method introduces novel implicit 3D features to accurately forecast the locations and dimensions of obstacles on planetary surfaces.
- The proposed method achieves an impressive 85.5% average precision （AP） on Mars surface images captured by the Spirit rover.
- Compared to stereo matching based methods，the proposed method significantly improves inference speed，requiring only 0.265 seconds for processing a pair of stereo images.

[责任编辑：杨晓燕，英文审校：宋利辉]