应用光学
Journal of Applied Optics
SCOPUS,等来源期刊
中文核心期刊、中国科技核心期刊

基于仿人眼视网膜特性的三维目标识别技术

金鑫 蒋振刚

Three-dimensional target recognition technology based on the retinal characteristics of human eye imitation

JIN Xin, JIANG Zhengang

引用本文:
金鑫, 蒋振刚. 基于仿人眼视网膜特性的三维目标识别技术[J]. 应用光学, 2026, 47(1): 139-146. DOI: 10.5768/JAO202647.0102005
JIN Xin, JIANG Zhengang. Three-dimensional target recognition technology based on the retinal characteristics of human eye imitation[J]. Journal of Applied Optics, 2026, 47(1): 139-146. DOI: 10.5768/JAO202647.0102005

在线阅读 View online: https://doi.org/10.5768/JAO202647.0102005

文章编号:1002-2082 (2026) 01-0139-08
引用格式:金鑫, 蒋振刚. 基于仿人眼视网膜特性的三维目标识别技术 [J]. 应用光学, 2026, 47(1): 139-146.
JIN Xin, JIANG Zhengang. Three-dimensional target recognition technology based on the retinal characteristics of human eye imitation[J]. Journal of Applied Optics, 2026, 47(1): 139-146.

在线阅读

基于仿人眼视网膜特性的三维目标识别技术

金 鑫, 蒋振刚

(长春理工大学 计算机科学技术学院, 吉林 长春 130022)

摘 要:随着无人车等智能系统的发展,对三维目标识别技术的需求日益增长。针对传统三维目标识别方法在实时性和数据量方面的局限性,提出了一种基于仿人眼视网膜特性的三维目标识别方法。通过建立仿人眼视网膜映射模型,实现了对目标的空间变分辨率成像,有效压缩了非感兴趣区域的数据,提高了对感兴趣目标的识别效率。最终在公开数据集 KITTI 与 Jetson Xavier Nx 开发板通过实验验证,证明了所提方法的有效性。

关键词:三维目标识别;仿人眼视网膜;变分辨率成像;实时性;数据压缩
中图分类号:TN201; TP751 **文献标志码**:A **DOI**: 10.5768/JAO202647.0102005

Three-dimensional target recognition technology based on the retinal characteristics of human eye imitation

JIN Xin, JIANG Zhengang

(School of Computer Science and Technology, Changchun University of Science and Technology, Changchun 130022, China)

Abstract: With the development of intelligent systems such as unmanned vehicles, there is a growing demand for 3D target recognition technology. In this work, for the limitations of traditional 3D target recognition methods in terms of real-time and data volume, a 3D target recognition method based on the bio-inspired retina-like characteristics was proposed. By establishing a bio-inspired retina-like mapping model, spatially variant resolution imaging of the target was realized, which could effectively compress the redundant information in the non-interested region and improve the recognition efficiency of the interested target. Finally, the effectiveness of the proposed method was proved by experimental verification in the Jetson Xavier Nx development board and public dataset KITTI.

Key words: three-dimensional target recognition; bio-inspired retina-like; variant resolution imaging; real-time ability; data compression

引言

三维目标识别技术是现代智能系统中不可或缺的关键技术之一,它在无人驾驶汽车、智能监控、机器人感知等领域发挥着重要作用。随着这些领域技术的快速发展,对三维目标识别技术的性能要求也越来越高,尤其是在实时性和准确性方面。然而,现有的三维目标识别方法,尤其是在激光雷达(LiDAR)技术应用中,面临着数据处理量大、实时性慢、识别精度受限等挑战[1]。

人眼作为自然界中最为精密的视觉系统之一,其独特的视网膜结构和功能为解决上述问题提供了新的视角。人眼视网膜的非均匀结构特性,即

收稿日期:2024-07-05; **修回日期**:2024-12-06
基金项目:国家自然科学基金(62275022)
第一作者:金鑫, E-mail: luna_xin163@163.com

中心凹区域具有高分辨率，而周边区域则负责广阔的视野，这种结构使得人类能够在保持视野广度的同时，对感兴趣区域进行高分辨率的观察[2]。这种视觉机制的高效性启发了本文的研究，即通过模拟人眼视网膜的成像机制，提出一种新的三维目标识别方法，以期提高识别系统的实时性和准确性。在三维目标识别领域，激光雷达作为一种主动三维传感器，因其能够在全天候工作且提供精确的距离信息而被广泛应用于无人驾驶汽车等系统中[3]。然而，激光雷达采集的数据量巨大，且数据结构复杂，给实时处理和识别带来了巨大挑战[4]。例如，文献 [5] 中提到，激光雷达在无人车感知系统中扮演着重要角色，但其数据量较大，导致识别速度较慢。为了解决这一问题，研究者们尝试了多种方法，如基于多视角识别的方法[6]、基于体素化的方法[7] 等，但这些方法在实际应用中仍存在局限性。

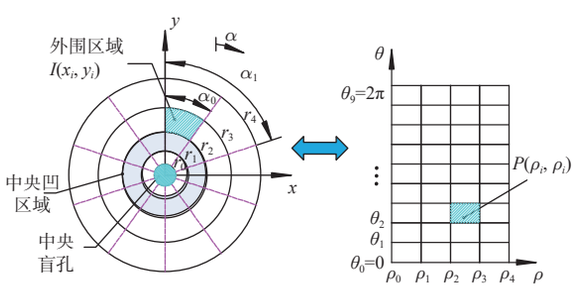
此外，三维目标识别的准确性也受到多种因素的影响。文献 [8] 指出，由于二维传感器无法直接获取三维信息，导致无人车距离感知能力缺失。而三维传感器虽然能够提供精确的距离信息，但其识别速度较慢，且对于复杂环境下的识别精度仍有待提高[9]。为了提高识别精度，一些研究者尝试将深度学习技术应用于三维目标识别[10]，但深度学习方法在实际环境中的性能还存在未知[11]。受到人眼视网膜特性的启发，本文提出了一种仿人眼视网膜的三维目标识别方法。通过建立仿人眼视网膜映射模型，实现了对目标的空间变分辨率成像，有效压缩了非感兴趣区域的数据，提高了对感兴趣目标的识别效率。这种方法不仅能够提高识别的实时性，还能够在一定程度上提高识别的准确性。硬件平台方面，本文设计了基于英伟达 Jetson Xavier NX 的硬件平台[12]，该平台能够实现对三维数据的有效采集和处理。Jetson Xavier NX 的高性能嵌入式计算能力，使得硬件平台能够满足实时性的需求。

1 仿人眼视网膜映射模型

1.1 映射模型的建立

本文首先对人眼视网膜的结构和功能进行了深入研究，发现视网膜中心凹区域具有高分辨率，而周边区域则负责广阔的视野。基于此，建立了仿人眼视网膜的正向和反向映射模型，实现了从笛卡尔坐标系到对数极坐标系的转换。正向映射模型用于将三维场景数据转换为视网膜模型的成像数据，而反向映射模型则用于将成像数据转换回三维场景，以便进行进一步的处理和分析。

仿人眼视网膜映射模型的核心是对数极坐标变换，是实现空间变分辨成像的主流思想[13]。仿人眼视网膜映射是一种几何图像变换，模拟人眼从视网膜到大脑视觉皮层的信息映射关系，拥有不同的数学转换模型[14]，综合考虑本文所使用的数据结构与编码的易用性，本文选择中间盲孔模型（central blind-spot model, CBM）[15]。仿人眼视网膜映射模型包括正向映射模型和反向映射模型，如图 1(a) 所示，半径分别为 r_0、r_1、r_2、r_3、r_4 的圆，转变为图 1(b) 的直线 ρ_0、ρ_1、ρ_2、ρ_3、ρ_4，角度 α_0、α_1、α_2、α_3、α_4 映射为 θ_0、θ_1、θ_2、θ_3、θ_4。此外，由于采集数据属于离散结构，而对数极坐标变换属于连续的非线性变换，因此对数极坐标与直角坐标系无法一一映射，需要考虑仿人眼视网膜离散化处理的方法。



(a) 视网膜空间 (b) 视皮层空间

图 1 视网膜-视皮层映射模型

Fig. 1 Retina-optic cortex mapping model

1.1.1 正向映射模型

正向映射模型是：视网膜空间图像经过对数极坐标转换后，形成视皮层空间图像。选择视网膜空间中心点 $o(x_0, y_0)$ 为注视点。如图 1(a) 视网膜空间坐标点 $I(x_i, y_i)$ 与图 1(b) 视皮层空间中极坐标点一一对应。ρ 为距离轴，θ 为角度轴，则变换关系如式（1）所示：

$$\begin{cases} r = \sqrt{(x-x_0)^2+(y-y_0)^2} \\ \rho = \log r \end{cases} \tag{1}$$

经过反正切运算，得到式（2）：

$$\begin{cases} \alpha = \arctan\left(\dfrac{y-y_0}{x-x_0}\right) \\ \theta = \alpha \end{cases} \tag{2}$$

用复数 z 表示，则可以表示成式（3）：

$$z = x + \mathrm{i}y = r(\cos\alpha + \mathrm{i}\sin\alpha) = rq^{\mathrm{i}\alpha} \tag{3}$$

式中：r、α为直角坐标系下平面的极径与角度；q是相邻环之间的增长系数。将式(3)两边同时取对数得到式(4)：

$$\log_q(z)=\log_q(r)+\mathrm{i}\alpha \tag{4}$$

令$w=\log_q(z),\rho=\log_q(r),\theta=\alpha$，则式(4)可以改写成式(5)：

$$w=\rho+\mathrm{i}\theta \tag{5}$$

式中，ρ、θ为对数极坐标中的半径与角度，可见对数极坐标变换由z空间的圆形区域转换为w空间的矩形区域。根据以上分析与介绍，可以得出仿人眼视网膜传输具有两个显著的特性，即尺度不变性与旋转不变性。

1) 尺度不变性

将极坐标轴向尺度变化转换为对数极坐标图像的左右平移，公式推导如下：

$$\rho=\log(k\cdot r)=\log k+\log r \tag{6}$$

从式(6)可知，如果目标以凝视点为中心，放大k倍，则相当于映射变换图向右移动$\log k$个单位。

2) 旋转不变性

把极坐标的角度变化转变为对数极坐标图像的上下平移，公式推导如下：

$$\theta_1=\theta_0+L \tag{7}$$

从式(7)可知，如果目标围绕中心点旋转L弧度，则相当于映射变换图纵向移动L个单位。

1.1.2　逆向映射模型

逆向映射模型：将视皮层空间图像经对数极坐标逆变换后，形成视网膜空间图像。反向映射模型基本原理为：由视皮层空间图像对数极坐标出发，通过逆变换计算所对应的视网膜空间像素位置。视皮层空间映射后的图像阵列$P(\rho_i,\theta_i)$映射到视网膜空间图像阵列$I(x_i,y_i)$的公式如式(8)所示：

$$\begin{cases} r=\mathrm{e}^{\rho} \\ \alpha=\theta \end{cases} \tag{8}$$

则图像阵列在视网膜空间坐标为

$$\begin{cases} x=r\times\cos\alpha+x_0 \\ y=r\times\sin\alpha+y_0 \end{cases} \tag{9}$$

同时需要对逆变换后的坐标像素进行赋值：

$$I(x,y)=P(\rho,\theta) \tag{10}$$

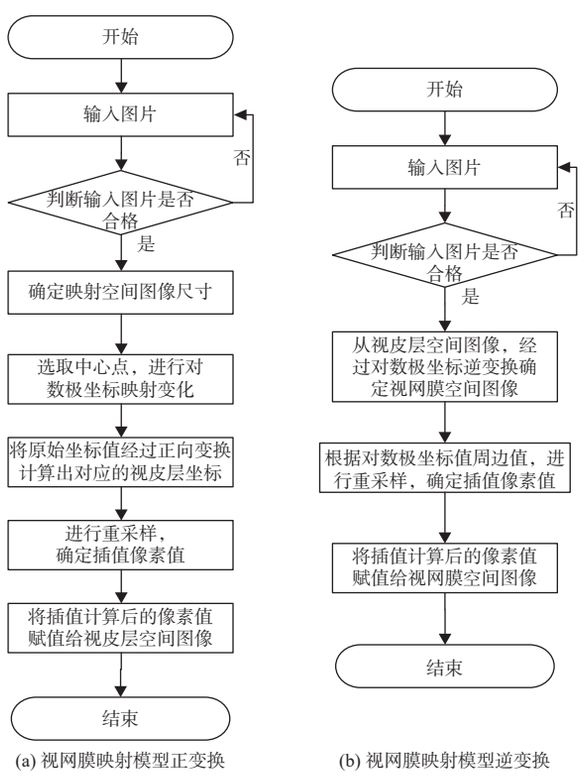
由于视网膜映射模型具有空间变分辨率的特性，导致部分信息丢失，因此逆变换图像并不能准确复原变换前的信息，表现为一个离散图像。在该离散图像中，随着距离中心点的距离增加，分辨率逐渐降低，而这一特性正好反映出人眼视觉的变分辨率特性，在凝视区域保持高分辨，而在周边区域分辨率逐渐降低。在目标识别相关算法中，这种像素排布使得算法体现出优越的实时性。

1.2　映射模型的验证

为了验证映射模型的有效性，本文设计了一系列仿真实验。实验结果表明，仿人眼映射模型能够有效压缩非感兴趣区域的数据，同时保持感兴趣区域的高分辨率。此外，本文还提出了亚像素级重采样方法，进一步提高了成像质量，为后续的目标识别提供了更清晰的图像数据。利用 Python 语言，结合 OpenCV 视觉库进行编写视网膜映射模型仿真验证程序。

1.2.1　正变化

视网膜映射模型正变换具体包括：1) 根据视网膜空间图像的尺寸和放大倍数确定视皮层空间映射图像尺寸；2) 根据图像长宽计算图像中心点$o(x_0,y_0)$作为对数极坐标映射原点；3) 将视网膜空间下的图像坐标$I(x_i,y_i)$，经过正向映射变换计算出对应的视皮层空间映射像素坐标$P(\rho_i,\theta_i)$；4) 根据$I(x_i,y_i)$周围像素值，进行重采样，确定像素值；5) 将像素值根据变换规则，赋值视皮层空间映射图像坐标$P(\rho_i,\theta_i)$。具体流程如图 2 所示。



(a) 视网膜映射模型正变换　　(b) 视网膜映射模型逆变换

图 2　视网膜映射模型正变换与逆变换流程图

Fig. 2　Flow charts of positive and inverse transformations of retinal mapping model

1.2.2 亚像素级重采样方法

沿半径方向进行二维插值，能在保证分辨率的同时降低计算复杂度。平面二维插值法方式如图 3 所示，圆形图像上的采样点在角度方向与半径方向上均按固定间隔均匀分布。4 个样本点分别为$p_{i,j}$、$p_{i+1,j}$、$p_{i,j+1}$、$p_{i+1,j+1}$，其中$p_{i,j}$、$p_{i,j+1}$为角度方向上相邻的两个点，$p_{i,j}$、$p_{i+1,j}$为半径方向上相邻的两个点，最终目标插值的点z位于上述 4 个点的中间位置。

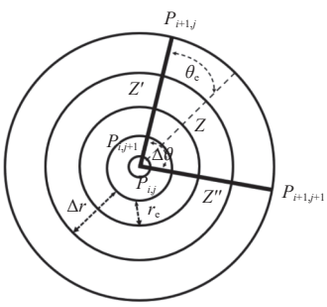


图 3 平面二维插值法

Fig. 3 Planar two-dimensional interpolation

平面二维插值主要分为两个步骤实现：1）在角度方向进行插值，得到角度插值和；2）在半径方向按照如式(11)进行插值，得到半径插值和。

$$\begin{cases} Z' = P_{i,j}\left(1-\dfrac{r_e}{\Delta r}\right)+P_{i+1,j}\dfrac{r_e}{\Delta r} \\ Z'' = P_{i,j+1}\left(1-\dfrac{r_e}{\Delta r}\right)+P_{i+1,j+1}\dfrac{r_e}{\Delta r} \\ Z = Z'\left(1-\dfrac{\theta_e}{\Delta\theta}\right)+Z''\dfrac{\theta_e}{\Delta\theta} \end{cases} \tag{11}$$

通过对 4 个样本点进行角度方向与半径方向插值，得到新的采样点坐标，实现了亚像素级重采样，使图像更加平滑，进一步提高了成像质量，为后续的目标识别提供了更清晰的图像数据。

1.2.3 逆变换

视网膜映射模型逆变换具体包括：1）从视皮层空间图像坐标$P(\rho_i,\theta_i)$，经过对数极坐标反向变换，计算出对应的视网膜空间映射坐标值$I(x_i,y_i)$；2）根据极坐标$P(\rho_i,\theta_i)$周围像素的像素值，进行重采样，确定像素值；3）将计算的像素值赋值给视网膜空间的映射图像坐标$I(x_i,y_i)$对应的像素值，如流程图 2(b) 所示。图像变换效果如图 4 所示，图 4(a) 为原图像，图 4(b) 为仿人眼视网膜正变换图像，图 4(c) 和图 4(d) 是仿人眼视网膜逆变换图像，区别在于图 4(d) 进行了重采样，而图 4 (c) 没有应用重采样算法。可以看出，原图像为笛卡尔坐标系下的图像，经过仿人眼视网膜映射模型正变换后，数据被压缩。再经过逆变换后可以看出，图 4(c) 在中间红圈处区域呈现高分辨率，与原始图像无差别，而周围区域随着距离中心越远，逐渐变得模糊，此现象符合仿人眼视网膜映射模型只对凝视区域高分辨、对非凝视区域压缩冗余数据的特性，经过重采样后，图 4(d) 周边区域的模糊效果被平滑，细节效果更佳。

(a) 原图像 (b) 仿人眼正变换图像 (c) 仿人眼逆变换图像(非重采样) (d) 仿人眼逆变换图像(重采样)

图 4 仿人眼正变换与逆变换图像对比图

Fig. 4 Comparison of bio-inspired retina-like positive and inverse transformed images

2 实验测试与分析

本研究以基于仿人眼深度学习的三维目标识别为核心，通过一系列实验与分析，充分验证了仿人眼成像技术在目标识别任务中的优势。

2.1 实验设计与实施

为了验证本章的研究方法，采用 Pytorch 深度学习框架建立点云三维目标识别模型，使用 KITTI 数据集的测试集进行网络性能的测试，验证其性能，实验环境如下文。

1）计算机硬件环境

INTEL i9-10900X CPU，主频 3.7 GHz，RTX 3080Ti GPU，内存 128 G，存储 1 T 固态硬盘，8 T 机械硬盘。

2）计算机软件环境

Ubuntu18.04 操作系统，Python 3.7，Pytorch 1.7，Open3D 0.1.2，OpenCV 4.5.1。

3）实验设置

训练时 batch_size 参数设置为 4，评估时设置为 1，训练次数为 240 epochs；只对 KITTI 数据集中汽车、行人、骑自行车人进行训练；训练点云范围为 [0, −39.68; −3, 69.12; 39.68, 1]，使用 7481 帧训练集进行训练，使用 7581 帧测试集进行算法性能评估；使用 Open3D 进行三维点云识别结果可视化，使用 OpenCV 进行二维图像可视化。

除此之外，本文还在实物方面进行实验，以验证算法的实际效果，设计主要包括以下几个步骤：1）系统搭建，配置硬件平台，安装软件环境，确保激光雷达与嵌入式开发板之间的通信正常；2）模型优化，将训练好的仿人眼三维目标识别算法模

型转化为 ONNX 格式, 并利用英伟达的 TensorRT 工具进行模型优化, 分别采用半精度(浮点 16)和单精度(浮点 32)进行加速。

实验平台包括硬件与软件的配置, 硬件平台主要包括: 1) 激光雷达, 作为主要的三维数据采集设备, 通过以太网与嵌入式平台相连, 提供高精度的三维点云数据; 2) 英伟达 Xavier NX 嵌入式开发板, 作为计算核心, 负责承载并执行优化后的仿人眼三维目标识别算法, 实现高速、高效的实时识别; 3) 上位机 GUI 交互软件, 提供友好的用户界面, 用于设置参数、选择激光雷达、显示识别结果等操作。实验原理图与实物图如图 5 所示。

视网膜凝视机制三维目标识别算法部署时与训练时不同, 不需要对数据进行扩充, 但是需要对激光雷达采集的点云进行数据预处理, 包括点云数据采样与滤波, 采样使用几何采样方式, 滤波使用统计滤波与快速双边滤波技术。视网膜凝视机制三维目标识别算法结构如图 6 所示。

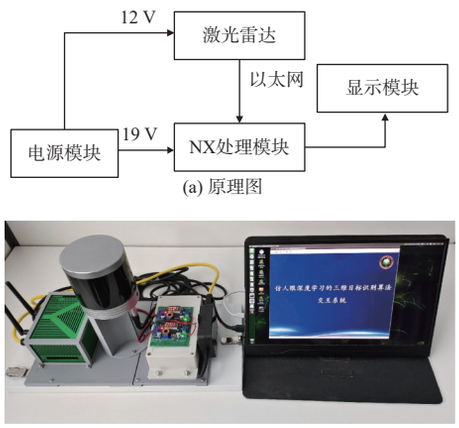


图 5　原理图与实物图

Fig. 5　Schematic and physical drawings

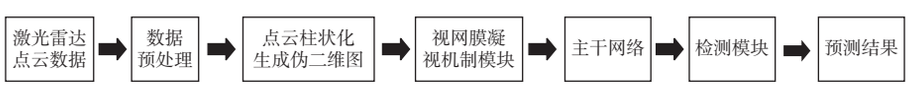


图 6　视网膜凝视机制三维目标识别算法部署时结构图

Fig. 6　Structural diagram of retinal gaze mechanism 3D target recognition algorithm when deployed

2.2　实验结果与分析

2.2.1　仿人眼模型性能对比

本小节分析评测在三维检测模式下, 通过识别简单、中等、困难 3 类的 KITTI 数据集难度, 将目前主流的三维目标识别算法的 3D 识别平均精度进行对比, 对比数据参考 KITTI 官方评测榜, 结果如下文。

表 1 将主流的三维目标识别算法和本文提出的仿生三维目标识别算法在精度方面进行对比, 可以看到通过视网膜凝视机制算法, 网络在识别汽车困难难度等级下、识别行人困难难度等级下、识别骑行人简单难度和困难难度等级下均取得了最好识别精度, 并且在困难难度等级下, 整体识别精度较好。其中, IOU 为交并比。表 2 中, 本文方法对于小目标行人和骑行人平均精度值达到了第一, 这是因为视网膜凝视算法更加关注局部特征细节, 使得对较小目标及遮挡物体有较好学习特性。值得一提的是, 该算法大幅度提升了运行速率, 速度达到了同类算法第一, 得益于算法压缩了数据, 实现了高效率的目标识别结果。

表 1　基于 KITTI 数据集 3D 检测的精度性能 (AP) 对比表

Table 1　Accuracy performance (AP) comparison table of 3D detection based on KITTI dataset

%

方法	速率/Hz	汽车(IOU=0.7)			行人(IOU=0.5)			骑行人(IOU=0.5)		
		简单	中等	困难	简单	中等	困难	简单	中等	困难
MV3D	2.8	71.09	62.35	55.12	N/A	N/A	N/A	N/A	N/A	N/A
F-PointNet	5.9	81.2	70.39	62.19	51.21	44.89	40.23	71.96	56.77	50.39
AVOD-FPN	10	81.94	71.88	66.38	50.8	42.81	40.88	64.00	52.18	46.61
VoxelNet	5.9	77.47	65.11	57.73	39.48	33.69	31.5	61.22	48.36	44.37
Second	4.4	83.13	73.66	66.20	51.07	42.56	37.29	70.51	53.85	46.9
Pointpillars	62	79.05	74.99	68.30	52.08	43.53	41.49	75.78	59.07	52.92
PointRCNN	10	86.96	75.64	70.70	57.89	50.59	48.38	74.96	58.82	52.53
Part-A2-Net	12.5	87.81	78.49	73.51	53.10	43.35	40.35	79.17	63.52	56.93
PV-RCNN	12.5	90.25	81.43	74.82	52.17	43.29	40.29	78.60	63.71	57.65
本文	63.2	83.26	75.31	76.38	56.55	50.16	49.06	80.41	62.01	58.40

表 2 基于 KITTI 数据集 3D 检测的平均精度 (mAP) 性能对比表

Table 2 Mean accuracy performance (mAP) comparison table of 3D detection based on KITTI dataset %

方法	模态	汽车	行人	骑行人
MV3D	RGB + LiDAR	62.85	N/A	N/A
F-PointNet	RGB + LiDAR	71.26	45.44	59.71
AVOD-FPN	RGB + LiDAR	73.40	44.83	54.26
VoxelNet	LiDAR	66.77	34.89	51.32
Second	LiDAR	74.33	43.64	57.09
Pointpillars	LiDAR	74.11	45.7	62.59
PointRCNN	LiDAR	77.76	50.28	62.10
Part-A2-Net	LiDAR	79.93	45.6	66.54
PV-RCNN	LiDAR	82.17	45.25	66.65
本文	LiDAR	78.29	51.92	66.94

通过可视化的方式验证本文所提方法的实际效果。如图 7、图 8 所示，分别选取街道和高速路进行分析，每幅图中：上半部分图 (a) 为摄像头采集的 rgb 图像，左下部分图 (b) 为点云图像俯视图视角，右下部分图 (c) 为点云图像主视图视角；其中当识别目标为汽车时用红色矩形框框选，识别为行人用绿色框进行框选，识别为骑自行车的人用蓝色框进行框选；框的前部分交叉线代表预测目标的方向。

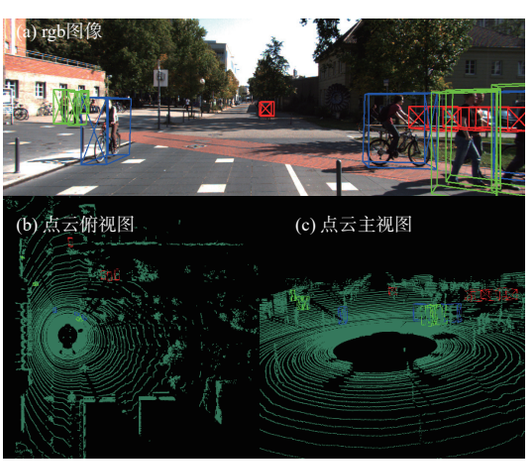


图 7 街道 3D 检测效果图

Fig. 7 3D detection rendering of the street

可以看到在街道场景下，算法对汽车、行人、骑自行车的人均有较好的识别效果，回归框与预测物体的方向也较为准确，此外，在图像右边汽车存在部分遮挡的情况下也可以较为准确地进行识别。而在高速路场景下，对于距离较远的汽车也能进行较好的识别。

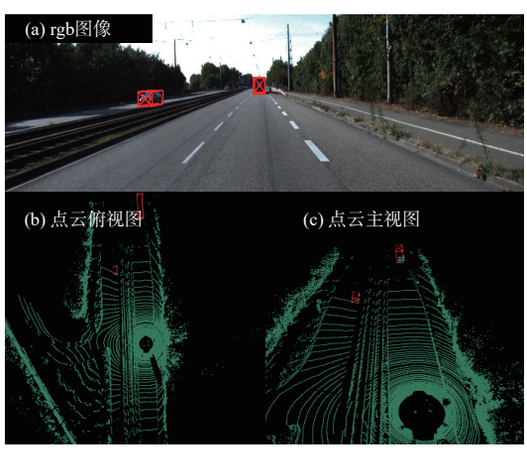


图 8 高速路 3D 检测效果图

Fig. 8 3D detection rendering of expressway

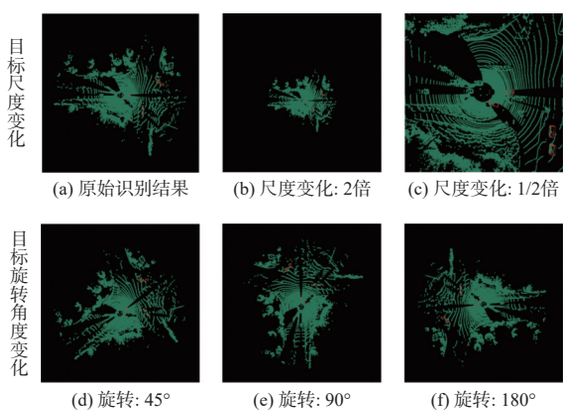
2.2.2 仿人眼目标尺寸和旋转角度实验验证

本实验验证不同条件下的目标识别结果，如目标尺度变化、旋转角度变化，以验证算法的鲁棒性。如图 9 所示，图 9(a) 为原始识别结果，图 9(b) 为将点云尺度扩大 2 倍输入给网络的识别结果，图 9(c) 为将点云尺度缩小 2 倍输入给网络的识别结果，可以看到在尺度扩大和缩小条件下，都能对点云目标进行良好的识别。图 9(d)～图 9(f) 为分别对图 9(a) 中点云逆时针旋转 45°、90°、180°的识别结果，可以看到均可对点云进行良好识别。因此无论目标尺寸和旋转角度如何变化，均对目标具有较好的效果，验证了算法的鲁棒性。

2.2.3 仿人眼嵌入式平台模型性能对比

在实际使用过程中，算法通常在嵌入式平台运行，因此本小节设计实验，验证该算法在嵌入式平台的性能，通过 TensorRT 进行加速后，结果如表 3

所示。未经优化的原始模型虽然具有较高的 mAP（86.32%），但识别帧频仅为 7.26 Hz，无法满足无人车实时感知的需求。经过 TensorRT 优化后，半精度模型识别帧频提升至 22.76 Hz，单精度模型提升至 15.88 Hz。尽管优化后模型的 mAP 略有下降，但考虑到大幅提升的识别速度以及对无人车应用来说可接受的精度损失，仿人眼三维目标识别算法在嵌入式平台上的优化部署效果显著。



(a) 原始识别结果　(b) 尺度变化: 2倍　(c) 尺度变化: 1/2倍

(d) 旋转: 45°　(e) 旋转: 90°　(f) 旋转: 180°

图 9　目标尺度变化、旋转角度变化测试结果

Fig. 9　Test results for changes in target scales and rotation angles

表 3　英伟达 Xavier NX 开发板上模型优化性能对比表

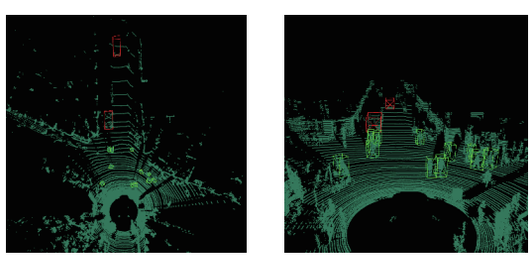
Table 3　Comparison table of model optimization performance on Nvidia Xavier NX development board

模型	mAP/%	速率/Hz
原始模型	86.32	7.26
TensorRT(浮点16)	84.69	22.76
TensorRT(浮点32)	85.25	15.88

2.2.4　实际场景识别效果

实验进一步通过实际场景识别验证了仿人眼三维目标识别算法在 NX 嵌入式平台上的有效性。如图 10 所示，系统成功识别出汽车（红色框）、行人（绿色框）和骑行人（蓝色框）。识别结果与精度测试结果表现一致，证明了该应用平台在实际场景下的可靠性和实用性。

通过系统的实验与分析，本研究有力地证明了仿人眼成像技术在目标识别任务中的显著优势。具体体现在以下方面：1）识别速度显著提升，仿人眼三维目标识别算法经过 TensorRT 优化部署后，识别帧频大幅提升至 22.76 Hz（半精度）和 15.88 Hz（单精度），满足无人车实时感知需求；2）精度损失可控，尽管优化后模型的 mAP 有所下降，但降幅较小（1.63% 至 1.07%），在保证较高识别精度的前提下，实现了识别速度的显著提升。通过实际场景验证：在英伟达 Xavier NX 嵌入式平台下进行的三维目标识别实验说明，系统能成功识别各类目标，可视化结果显示识别结果准确且稳定，验证了仿人眼成像技术在实际应用中的有效性。综上所述，仿人眼成像技术凭借其独特的非均匀成像、压缩冗余数据、旋转与尺度不变性等特点，有效地解决了传统深度学习三维目标识别中遇到的数据量大、识别速率慢等问题，显著提升了识别系统的实时性和准确性，为无人车等应用场景提供了高效、可靠的三维目标识别解决方案。



(a) 点云俯视图　(b) 点云主视图

图 10　实际场景识别效果图

Fig. 10　Actual scene recognition effect diagram

3　结论

在实验验证方面，本文通过一系列实验对所提出的三维目标识别方法进行了验证。实验包括了不同条件下的目标识别测试，如 KITTI 公开数据集测试、目标尺寸旋转角度变化、实际场景测试等。实验结果表明，与传统的三维目标识别方法相比，基于仿人眼视网膜特性的方法在目标识别的准确性和实时性方面均有显著提升。综上所述，本文提出的基于仿人眼视网膜特性的三维目标识别方法，有效地解决了传统方法在数据量和实时性方面的局限，具有重要的理论和应用价值。未来的研究将进一步优化硬件设计，提高系统的集成度和稳定性，以满足更广泛的应用需求。同时，还将探索将该方法与其他先进的图像处理和机器学习技术相结合，以进一步提高三维目标识别的性能。

参考文献：

[1] NIE C, JU Z Y, SUN Z F, et al. 3D object detection and tracking based on lidar-camera fusion and IMM-UKF algorithm towards highway driving[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2023,

7(4): 1-11.

[2] LEE B, JEONG S, LEE J, et al. Wide-field three-dimensional depth-invariant cellular-resolution imaging of the human retina[J]. Small, 2023, 19(11): 2203357.

[3] JIN X J, YANG H, HE X K. Robust lidar-based vehicle detection for on-road autonomous driving[J]. Remote Sensing, 2023, 15(12): 3160.

[4] ANAND B, KAMBHAMPATY HR, RAJALAKSHMI P. A novel real-time lidar data streaming framework[J]. IEEE Sensors Journal, 2022, 22(23): 23476.

[5] SRIVASTAVA A K, SINGHAL A, SHARMA A. Analysis of lidar-based autonomous vehicle detection technologies for recognizing objects[C]// 2023 6th International Conference on Information Systems and Computer Networks (ISCON). Mathura, India: IEEE, 2023: 1-5.

[6] LI G F; ZOU C J, JIANG G Z, et al. Multi-view fusion network-based gesture recognition using sEMG data[J]. IEEE Journal of Biomedical and Health Informatics, 2023,28(8): 4432-4443.

[7] 范希明. 基于图神经网络的体素化室内点云场景识别方法研究[D]. 南京: 南京邮电大学, 2023.
FAN Ximing. Research on voxelized point cloud recognition ofIndoor scene based on graph neural network [D]. Nanjing: Nanjing University of Posts and Telecommunications, 2023.

[8] 张新钰, 高洪波, 赵建辉, 等. 基于深度学习的自动驾驶技术综述[J]. 清华大学学报 (自然科学版), 2018, 58(4): 438-444.
ZHANG Xinyu, GAO Hongbo, ZHAO Jianhui, et al. Overview of deep learning intelligent driving methods[J]. Journal of Tsinghua University (Natural Science Edition), 2018, 58(4): 438-444.

[9] ANDREOPOULOS A, TSOTSOS J K. 50 years of object recognition directions forward[J]. Computer Vision and Image Understanding, 2013, 117(8): 827-891.

[10] QI C R, SU H, MO K, et al. PointNet: deep learning on point sets for 3D classification and segmentation[C]// 2017 IEEE Conference on Computer Vision and Pattern Recognition. Honolulu, HI, USA: IEEE, 2017: 652-660.

[11] LI Y, BU R, SUN M, et al. PointCNN: convolution on x-transformed points[C]// 32nd Conference on Neural Information Processing Systems (NIPS). Montreal, Canada: IEEE, 2018: 31.

[12] NVIDIA. NVIDIA Jetson Xavier NX[EB/OL]. (2020-03-18)[2024-04-24]. https://www.nvidia.cn/autonomous-machines/embedded-systems/jetson-xavier-nx.

[13] SCHWARTZ E L, GREVE D N , BONMASSAR G . Space-variant active vision: definition, overview and examples[J]. Neural Networks, 1995, 8(7/8): 1297-1308.

[14] BOLDUC M, LEVINE M D. A review of biologically motivated space-variant data reduction models for robotic vision[J]. Computer Vision & Image Understanding, 1998, 69(2): 170-184.

[15] MOLONI A B, VALSESIA D, FRACASTORO G, et al. Towards deep unsupervised sar despeckling with blind-spot convolutional neural networks[C]// 2020 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Waikoloa, HI, USA. IEEE, 2020: 2507-2510.