基于 DeepLabV3+ 的甲骨拓片图像文字分割方法

史小松 [1,2,3] 代记圆 [1]
SHI Xiaosong DAI Jiyuan

摘 要 为将甲骨拓片中的文字和背景有效分割，提出一种基于深度可分离卷积网络 DeepLabV3+ 的甲骨文拓片图像文字分割方法，实验使用了包含 9800 张甲骨文拓片图像的数据集进行训练和测试。结果表明，提出的方法在进行甲骨文拓片文字分割时具有较高的准确率，能为甲骨文字的识别、甲骨图片的缀合等提供一定的基础数据。

关键词 甲骨文拓片；图像文字分割；深度学习；DeepLabV3+；数据集

doi：10.3969/j.issn.1672-9528.2024.04.050

0 引言

甲骨文拓片图像是甲骨文的重要载体，通过对甲骨文拓片图像的研究，可以更加深入地了解古代商代社会的历史演变、社会制度、政治文化、宗教信仰等方面的信息，对于推动中国古代史研究和文化遗产保护具有重要意义。

然而，在迄今已发现的 4500 多个甲骨文单字中，仍然有 3000 个左右的单字未获识别。甲骨文拓片研究仍然面临着相当大的挑战，主要体现在以下三方面。第一，拓片背景噪声严重。由于甲骨常年埋在地下，造成大面积损坏，致使相当一部分拓片都有着比较严重的噪声干扰，上面的很多文字模糊不清，难以辨识。第二，拓片上残缺字较多。甲骨出土时很容易断裂，断裂发生在甲骨字附近就会形成很多残缺字，难以准确定位。第三，拓片上文字大小方向不规则。同一片甲骨上锲刻的甲骨文字大小不一，间距不同，方向不定，以及由此形成多处重叠的文字区域都给研究带来了困难。因此，甲骨拓片图像存在着许多需要攻克的难点。

随着计算机技术及人工智能技术的蓬勃发展和甲骨文数据的大规模化，甲骨文信息处理方面的研究成为一种新的趋势。如果能将甲骨拓片这些原始数据数字化，然后利用人工智能和模式识别技术从甲骨拓片中分割出每一个字符，进行基于拓片图像的单字定位，就可以为后续的字形特征提取、字的识别、拓片缀合等提供基础的数字化数据，为进一步的研究奠定基础。2019 年，集甲骨文字库、著录库、文献库于一体的“殷契文渊”甲骨文大数据平台的发布，为甲骨学研究提供了大数据支持，更是标志着甲骨学研究进入智能化时代。近年来，许多内外学者针对拓片文字的定位、检测和识别展开了一些研究，取得了一定的进步，然而其检测识别率仍然较低，其性能远不能满足甲骨文研究的需求。

1 深度学习在甲骨文信息处理中的研究现状

传统的甲骨文拓片图像文字分割方法主要基于图像处理和计算机视觉技术，通常采用二值化、形态学操作、边缘检测等方法对图像进行预处理，然后通过阈值分割、连通域分析等方法实现文字和背景的分离。前期基于拓片的单字分割相关研究较少，2016 年，文献 [1] 采用基于连通区域的方法定位甲骨文，以期通过检测拓片内甲骨文单字的位置和轮廓，作为后续文字识别和图像理解的基础。经过实验，初步实现了拓片甲骨文字的自动定位。这些方法在处理简单的图像场景时表现良好，但在处理复杂的甲骨文拓片图像时存在着一定的局限性。

近年来，深度学习技术的快速发展促使研究人员逐渐开始利用深度学习技术进行甲骨文字检测的尝试，并提出了一些新的方法 [2-5]，与深度学习处理其他文字检测和识别的方法类似，基于深度学习的甲骨文拓片图像检测和识别一般都需要先在拓片上标注好文字的位置信息，建立相关的数据集，然后对相关数据进行训练。2019 年，华南理工大学黄双萍教授团队 [6] 标注了 5838 幅甲骨拓片图像，构造了一个甲骨文字检测数据集 OBCD，在此基础上，黄双萍教授团队结合使用基于区域的全卷积神经网络 R-FCN（region based fully

1. 安阳师范学院计算机与信息工程学院 河南安阳 455000
2. 甲骨文信息处理教育部重点实验室 河南安阳 455000
3. 河南省甲骨文信息处理重点实验室 河南安阳 455000
[基金项目] 古文字与中华文明传承发展工程（项目编号：G2821，G1806，G3028，G1807）；河南省科技发展计划项目（项目编号：232102320162）；“甲骨文信息处理”教育部创新团队（项目编号：2017PT35）；河南省高等学校青年骨干教师培养计划（项目编号：2021GGJS129）

convolutional network）[7] 和特征金字塔模型 FPN（feature pyramid networks）[8] 进行了甲骨字检测研究。刘国英教授团队构造了一个包含 9500 张图像的甲骨文字检测数据集，通过分析近几年有代表性的通用目标检测框架（包括 Faster R-CNN、SSD、RefineDet、RFBnet 和 YOLOv3）在此数据集上的甲骨文字检测性能，找出其中性能最优的 YOLOv3，并对其进行了改进。日本立命馆大学 LIN Meng[9] 构造了一个包含 330 幅图像的甲骨文字检测数据集，改进了 SSD[10]，提高了较小字体甲骨字的检测准确率。

上述方法利用深度学习中不同的深度模型和框架进行文字检测和识别，大部分需通过对已有的网络模型进行改进来提升准确率。上述方法对于字体大小相似、噪声干扰较少的拓片文字有着不错的效果，然而并不能应用到大多数甲骨拓片中，这一方面与甲骨损坏严重、噪声多、样本量少、标注标准不统一等因素有关，另一方面，上述检测方法中大多基于回归方式，一般均需采用 anchor 机制产生文本候选框，而且每一个文字区域要用不同长宽比的多个 anchor，数量庞杂，对于重叠和距离很近的甲骨文字效果不佳。

谷歌团队提出的 DeepLab 系列语义分割网络，在很多语义分割数据集上均取得了良好的分割效果。其中，谷歌团队在 2018 年提出的深度可分离卷积网络 DeepLabV3+，利用空洞空间金字塔池化（ASPP）模块来提取多尺度特征，并使用 encoder-decoder 结构将高层次语义信息与低层次细节信息相结合，可提高模型的分割能力。同时，DeeplabV3+ 运用全连接条件随机场（CRF）来进行上下文信息的融合，将像素之间的相互作用考虑在内，从而消除图像中的噪声和不规则区域，提高分割的精度。本文利用深度学习较好的学习识别能力，针对甲骨拓片具体特征，开展基于 DeeplabV3+ 的甲骨拓片单字分割算法实验，通过基于 DeepLabV3+ 的分割以及一定的后处理操作，实现更准确的拓片文字分割效果。

2 网络结构

2.1 DeepLabV3+ 的网络结构

DeepLabV3+ 是一种基于 ASPP（atrous spatial pyramid pooling）模块的语义分割网络，通过引入 ASPP 模块，将深层特征图上采样到合适的空间尺度，并结合全局信息、边缘信息和局部信息进行语义分割，从而有效提升语义分割的精度。

DeepLabv3+ 模型的整体架构如图 1 所示，在编码器（encoder）中使用并行的空洞卷积（atrous convolution）对压缩四次的初步有效特征层进行特征提取，然后将特征进行合并和压缩。解码器（decoder）采用级联解码器，以空洞卷积的深度卷积神经网（DCNN）为主体，利用常用的分类网络如 ResNet（残差神经网络）和带有空洞卷积的空间金字塔池化模块（ASPP)），达到引入多尺度信息的目的；Decoder 模块的引入，将图像的底层特征与高层特征进一步融合，可以提升分割边界准确度。也就是说，DeepLabv3+ 在空洞卷积的全卷积网络（DilatedFCN）基础上引入了编码解码（encoder-decoder）的思路。在 Decoder 中使用 1*1 卷积调整通道数，然后将其与空洞卷积后的有效特征层上采样的结果进行堆叠，并进行两次深度可分离卷积块。这样，就可以获得一个最终的有效特征层，它是整张图片的特征浓缩。

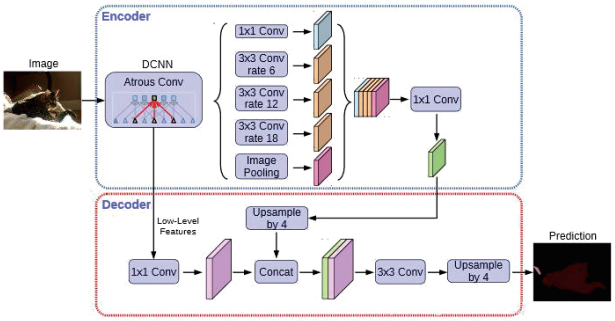

图 1 DeepLabV3+ 主干网络

2.2 Xception 网络

DeeplabV3+ 使用的主干网络是深度卷积神经网络 Xception，在 Xception 中，标准卷积操作被一种类似于深度可分离卷积的新型卷积操作所替代。首先，它对输入的每个通道都使用仅包含一个通道的卷积核进行卷积操作后，得到一组中间结果；然后，对这组中间结果进行 1*1 卷积操作，将各个通道之间的信息进行整合，得到最终的输出。Xception 还使用了通道混洗（channel shuffle）操作，将输入通道分成多个组，并将不同组的通道交错连接，以增加模型的表达能力和灵活性，进一步提高模型的准确率。这样一来，可以在保持高精度的同时，大幅度减少网络参数数量和计算量，从而提高模型的效率和准确率。其网络结构如图 2 所示。

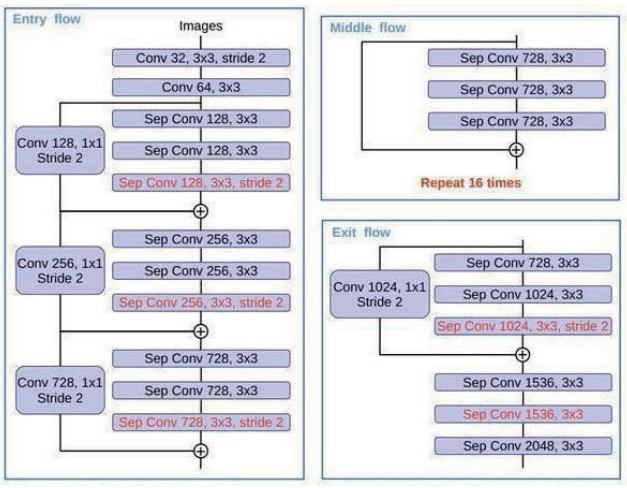

图 2 Xception 架构

3 基于 V3+ 的甲骨文拓片图像文字分割方法实现

3.1 数据集的构建和标注

实验采用了自行构建的甲骨文拓片图像数据集，如图 3 所示，其中包括《甲骨文合集》的 9800 张图像，所有图像划分为训练集、验证集和测试集。训练集、验证集按照 10:1 的比例进行划分，分别包含 8000 和 800 张图像，测试集包含了 1000 张图像。

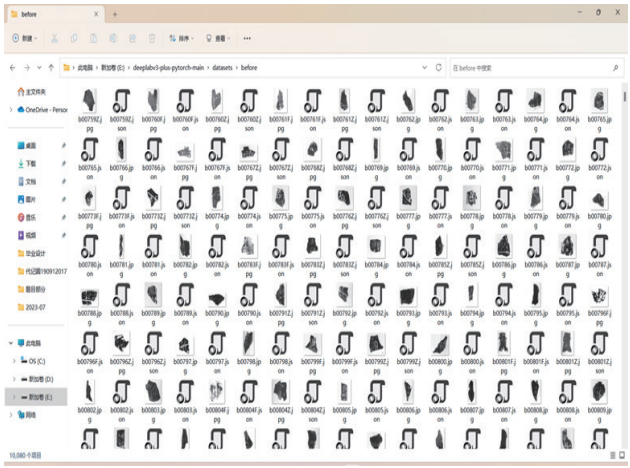

图 3 甲骨文拓片图像数据集

每张图像的分辨率为 512×512，图像中的文字信息需要用 labelme 工具标注出来（如图 4），并使用 json_to_dataset.py 将 json 文件转为 PNG 图片（如图 5）。

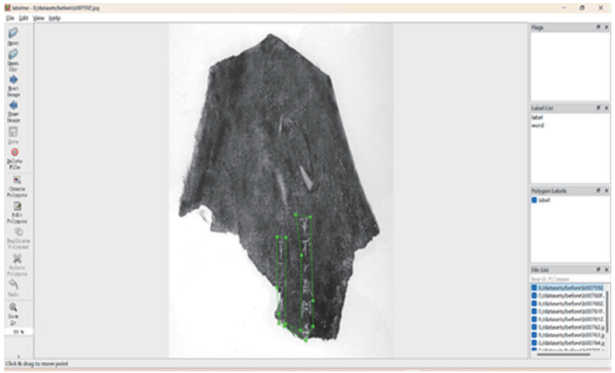

图 4 labelme 运行界面

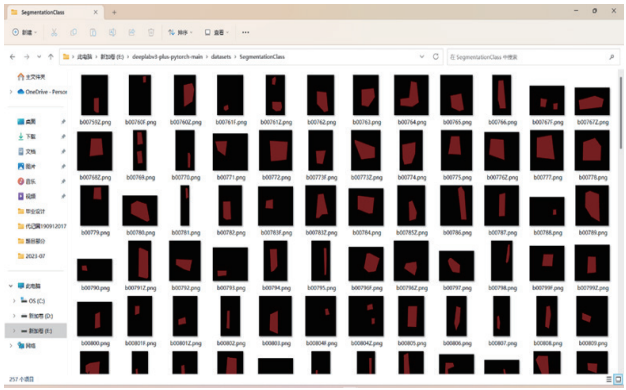

图 5 标注后的 SegmentationClass 文件

然后使用 voc_annotation.py 划分训练集与验证集，同时将文件名制作为 txt 文件用来训练。

3.2 系统环境

系统开发的硬件环境：CPU 为 Intel(R) Core(TM) i7-9750 H，内存为 8 GB，操作系统为 Windows 10。使用了 Python 语言和 PyTorch 深度学习框架。

3.3 训练策略和参数设置

在模型训练的过程中，需要合理设置学习率和优化器等参数。实验使用 adam 优化器和交叉熵损失函数，并将图像大小缩放到 512×512 进行训练，防止模型陷入局部最优解。同时，采用正则化技术来缓解过拟合问题，如 L1/L2 正则化、dropout 等。运行大约需要 40 ～ 60 min（如图 6）。最后生成 .pth 的权值文件，利用权值文件对图片进行预测。

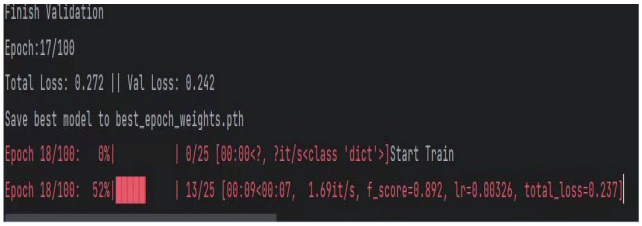

图 6 训练过程

为进一步提高基于深度可分离卷积网络 V3+ 的甲骨文拓片图像文字分割方法的性能和稳定性，实验中引入了注意力机制和多任务学习方法，注意力机制可帮助模型更加聚焦于图像中与文字相关的区域，进而提高模型对文字的识别能力。同时，采用多任务学习方法来提高模型的泛化能力。多任务学习可以使模型不只学习文字分割任务，还可以学习其他与文字相关的任务，如字符识别、文本识别等。这样可以增加模型的泛化能力，提高模型对于未知数据的适应性。

自适应学习率伪代码如下：

```
#-----------------------------------#
#  判断当前 batch_size，自适应调整学习率
#-----------------------------------#
nbs = 16
lr_limit_max = 5e-4 if optimizer_type == 'adam' else 1e-1
lr_limit_min = 3e-4 if optimizer_type == 'adam' else 5e-4
if backbone == "xception":
    lr_limit_max = 1e-4 if optimizer_type == 'adam' else 1e-1
    lr_limit_min  = 1e-4 if optimizer_type == 'adam' else 5e-4
Init_lr_fit  = min(max(batch_size / nbs * Init_lr, lr_limit_min), lr_limit_max)
Min_lr_fit  = min(max(batch_size / nbs * Min_lr, lr_limit_min * 1e-2), lr_limit_max * 1e-2)
#-----------------------------------#
#  获得学习率下降的公式
```

```
#----------------------------------#
lr_scheduler_func = get_lr_scheduler(lr_decay_type, Init_lr_fit, Min_lr_fit, UnFreeze_Epoch)
```

优化器参数伪代码如下：

```
#----------------------------------#
#  根据 optimizer_type 选择优化器
#----------------------------------#
optimizer = {
    'adam': optim.Adam(model.parameters(), Init_lr_fit, betas = (momentum, 0.999), weight_decay = weight_decay),
    'sgd': optim.SGD(model.parameters(), Init_lr_fit, momentum = momentum, nesterov=True, weight_decay = weight_decay)
}[optimizer_type]
```

3.4 测试及分析

实验使用测试集中的图像进行测试，并计算了平均交并比、平均像素准确度、平均精确度、平均召回率等指标。如图 7 部分图片测试结果所示，通过不同字体、不同大小和不同复杂度的文字的分割结果，可以直观地了解模型的文字分割效果。

图 7 部分测试结果

在本次实验中，使用了甲骨文数字数据库和甲骨文图像与文字数据库中的 9800 张甲骨文拓片图像进行模型训练和测试。在本次文字分割任务中，使用像素准确度（pixel accuracy，PA）、平均交并比（mean intersection over union，mIoU）和召回率（recall）来进行整体性能分析。

像素准确度（pixel accuracy，PA）是语义分割最基本的指标，指预测正确的像素占总像素的比例，本实验像素准确度达 87.67%，如图 8 所示，公式不再详述。

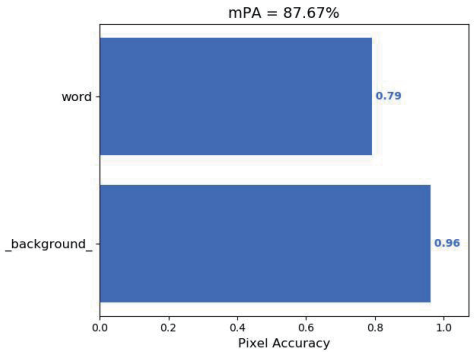

图 8 像素准确度

平均交并比（mean intersection over union，mIoU）是衡量语义分割模型分割效果的重要指标。通过计算预测分割结果与真实分割结果的交集与并集之间的比例，来描述分割结果与真实情况的重叠程度。mIoU 越高，说明模型的分割精度越好。实验求出的 mIoU 为 76.16%，如图 9 所示。计算公式为：

$$mIOU = \frac{1}{k+1}\sum_{i=0}^{k}\frac{P_{ii}}{\sum_{j=0}^{k}P_{ij}+\sum_{j=0}^{k}P_{ji}-P_{ii}} \quad (1)$$

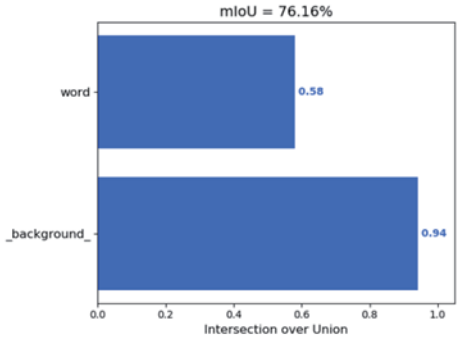

图 9 平均交并比

召回率是衡量语义分割模型真实标签覆盖率的指标，即模型能够正确预测出真实标签的比例，召回率越高，说明模型对真实标签的覆盖能力越强。实验中，召回率（recall）为 87.67%，如图 10 所示。

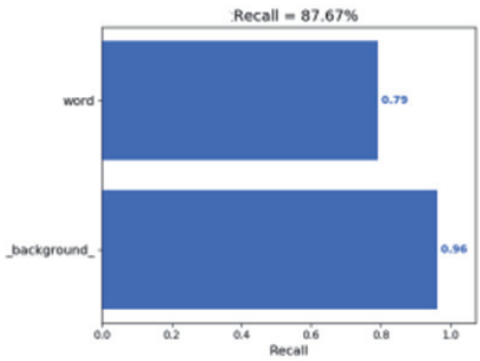

图 10 召回率

4 结论

本文利用深度可分离卷积网络 V3+ 对甲骨文拓片图像进行了分割，将已知甲骨文拓片图像数据集中的文字和背景进行有效分离，具有较高的准确率，减少了图像分割的人工操作，为甲骨文字的识别、甲骨图片的缀合等提供一定的基础数据，在处理甲骨图片的实际应用中具有一定的应用前景。然而，该方法仍存在不足之处，如在某些复杂的甲骨文拓片图像中，存在识别错误的情况，这可能是由于模型的学习能力不足或者训练数据不足造成的，需要进一步研究和改进。

参考文献：

[1] 史小松，黄勇杰，刘永革．基于阈值分割和形态学的甲骨拓片文字定位方法 [J]. 北京信息科技大学学报 ,2014,29(6):7-10+24.

[2] 李金洪．深度学习之 TensorFlow：入门、原理与进阶实战 [M]. 北京：机械工业出版社 ,2018.

[3] 张颐康，张恒，刘永革，等．基于跨模态深度度量学习的甲骨文字识别术 [J]. 自动化学报 ,2021,47(4):791-800.

[4] 林小渝，陈善雄，高未泽，等．基于深度学习的甲骨文偏旁与合体字的识别研究 [J]. 南京师范大学学报，2021, 44(2): 104-116.

[5]SHI X, HUAN Y, LIU Y. Text on oracle rubbing segmentation method based on connected domain[C]// 2016 IEEE Advanced information Communicates, Electronic and Automation Control Conference.Piscataway:IEEE,2016:459-463.

[6] 王浩彬．基于深度学习的甲骨文检测与识别研究 [D]. 广州：华南理工大学 ,2019.

[7]DAI J, LI Y, HE K, et al．R-FCN: object detection via region-based fully convolutional networks[C]//Advances in Neural Information Processing Systems 29,vol. 1.La Jolla, CA:Neural Information Processing Systems, 2016:379-387.

[8]TSUNG Y, PIOTR D, ROSS G, et al. Feature pyramid networks for object detection[C]//2017 IEEE Conference on Computer Vision and Pattern Recognition(CVPR). Piscataway: IEEE, 2017: 936–944.

[9]MENG L, LYU B，ZHANG Z, et al. Oracle bone inscription detector based on SSD[C]//New Trends in Image Analysis and Processing-ICIAP 2019.Cham:Springer Nature Switzerland AG, 2019:126-136.

[10]LIU W, ANGUELOV D, ERHAN D, et al. SSD: single shot multibox detector[C]//Computer Vision-ECCV 2016-14th European Conference,Amsterdam,The Netherlands,October 11-14,2016,Proceedings,Part I:Vol 9905.Cham:Springer, 2016: 21-37.

【作者简介】

史小松（1981—），女，河南林州人，硕士，副教授，研究方向：甲骨文信息处理。

代记圆（2000—），男，河南许昌人，本科，研究方向：计算机应用。

（收稿日期：2024-01-22）

（上接第 215 页）

参考文献：

[1] 交通运输部 .2022 年交通运输行业发展统计公报 [N]. 中国交通报 ,2023-06-16(002).

[2] 杨沛斌，张紫潇，赵永库．一种基于相控阵天线的前向机载防撞系统 [J]. 电讯技术 ,2023,63(8):1145-1150.

[3] 王士龙．浅谈直升机防撞雷达现状与技术发展 [J/OL]. 现代雷达 :1-11[2024-03-08].http://kns.cnki.net/kcms/detail/32.1353.TN.20240108.1026.002.html.

[4] 孔盼，蒲文杰．商用车自动驾驶单目测距设计与应用 [J]. 汽车工业研究 ,2023(3):19-21.

[5] 董兵，吴悦，郝宽公，等．基于位置误差模型的机场侧向跑道碰撞风险评估 [J]. 航空计算技术 ,2024,54(1):27-31.

[6] 周珂．基于双目视觉的目标测距和三维重建研究 [J]. 信息技术与信息化 ,2023(6):68-71.

[7] 朱亮，雷晓欣，李小鹏，等．加改装飞机局部结构载荷谱实测与数据处理方法研究 [J]. 航空科学技术 ,2022,33(6):46-52.

[8] 贾翔宇．基于小波变换的夜间低照度图像降噪与增强算法 [J]. 信息技术与信息化 ,2019(2):107-109.

[9] 华宇宁，李淼．基于双目视觉的人体头部尺寸测量方法研究 [J]. 信息技术与信息化 ,2023(2):42-45.

[10]SHU F, WANG J, PAGANI A, et al. Structure plp-slam: efficient sparse map** and localization using point, line and plane for monocular, rgb-d and stereo cameras[C]//2023 IEEE International Conference on Robotics and Automation (ICRA). Piscataway:IEEE, 2023: 2105-2112.

[11] 陈清华，张旭．基于双目立体匹配的巷道三维重建研究 [J]. 激光杂志 ,2022,43(10):208-212.

[12]ZHANG Y. Camera calibration[M]//3-D Computer Vision: Principles, Algorithms and Applications. Singapore: Springer Nature Singapore, 2023: 37-65.

[13] 孙艳坤，张威，杨雄伟，等．飞机牵引滑行技术研究进展与技术挑战 [J/OL]. 交通运输工程学报 :1-24[2024-03-08].http://kns.cnki.net/kcms/detail/61.1369.U.20230526.1331.002.html.

【作者简介】

孙豪杰（1997—），男，陕西扶风人，硕士研究生，助理工程师，研究方向：航电设备与系统。

（收稿日期：2024-03-20）