

# 基于 CNN-Transformer 的自动泊车车位感知算法\*

王玉龙<sup>1,2</sup> 翁茂楠<sup>1</sup> 黄辉<sup>1</sup> 覃小艺<sup>1</sup>

(1. 广州汽车集团股份有限公司汽车工程研究院, 广州 510641; 2. 湖南大学, 汽车车身先进设计制造国家重点实验室, 长沙 410082)

**【摘要】**为提高自动泊车成功率及准确性, 首先基于卷积神经网络(CNN)模型对输入图像进行特征提取, 然后利用 Transformer 模型的“编码-解码”机制对 CNN 提取到的图像特征平铺后进行计算推理, 通过前馈神经网络得到目标预测结果, 最后基于 180° 广角鱼眼图像进行推理识别, 车位角中心点和空车位入口中心点均采用二维坐标表示, 降低了输出信息的冗余, 优化了模型结构。测试结果表明, 该算法能够较好地适应不同车位线划线方式和不同的自然环境, 目标感知的召回率达到 98%, 车位角中心点定位平均误差小于 3 cm, 满足泊车系统对车位感知的鲁棒性、实时性和准确性要求。

**关键词:** 自动泊车 车位检测 视觉增强 卷积神经网络 Transformer

中图分类号: U463.3 文献标志码: A DOI: 10.19620/j.cnki.1000-3703.20221186

## Algorithm for Parking Space Detection in Automatic Parking System Based on CNN-Transformer

Wang Yulong<sup>1,2</sup>, Weng Maonan<sup>1</sup>, Huang Hui<sup>1</sup>, Qin Xiaoyi<sup>1</sup>

(1. Auto Engineering Research Institute, Guangzhou Automobile Group Co., Ltd., Guangzhou 510641; 2. State Key Laboratory of Advanced Design and Manufacturing for Vehicle Body, Hunan University, Changsha 410082)

**【Abstract】**In order to improve success rate and accuracy of automatic parking, firstly, the input image features were extracted based on Convolutional Neural Network (CNN) model, and then the encoding-decoding mechanism of Transformer model was used to tile the image features extracted by CNN for calculation and inference. Finally, the target prediction results were obtained by feedforward neural network. In this paper, fisheye images were used to recognize the target. The center point of the parking angle and the center point of the empty parking entrance were expressed by two-dimensional coordinate points, which reduced the redundancy of the output information and optimized the model structure. The test results show that the algorithm can better adapt to different parking space line marking mode and different natural environment, with the recall rate of target perception reaches 98%, and the average error of parking space corner center location is less than 3 cm, which meets the requirements of real-time application for robustness, real-time and accuracy.

**Key words:** Automatic parking, Parking space detection, Visual enhancement, Convolutional Neural Network (CNN), Transformer

**【引用格式】**王玉龙, 翁茂楠, 黄辉, 等. 基于 CNN-Transformer 的自动泊车车位感知算法[J]. 汽车技术, 2024(8): 1-6.

WANG Y L, WENG M N, HUANG H, et al. Algorithm for Parking Space Detection in Automatic Parking System Based on CNN-Transformer[J]. Automobile Technology, 2024(8): 1-6.

## 1 前言

自动泊车系统通过环视摄像头获取车身周围图像信息, 利用传统计算机视觉算法或者深度学习模型推理进行车位感知, 并由规控算法基于感知结果进行车辆泊入。车位的感知结果决定了自动泊车的泊入成功率和最终泊入效果。在实际泊车场景中, 不同停车场的车位

线划线方式存在较大差异, 且车位线的颜色和地面的颜色、材质等存在多种组合形式。同时, 周围环境的变化也会对环视图像的成像效果产生影响, 如强光、反光、弱光及阴影遮挡等, 给车位的识别和跟踪造成困难。

Zhang<sup>[1]</sup> 等通过由 3 个自适应增强 (Adaptive Boosting, AdaBoost) 算法构建的弱分类器对车身周围其他车辆的特征进行检测, 利用周围车辆的姿态信息对空

\*基金项目: 湖南大学汽车车身先进设计制造国家重点实验室开放基金项目(31825011)。

车位进行推算,但此方法容易受到周围车辆停车姿态的影响。朱强<sup>[2]</sup>基于高度融合卷积网络(Highly Fused Convolutional Network, HFCN)模型对俯视全景图中的车位线进行分割,利用多尺度融合的损失函数训练模型,通过识别车位线信息推理出停车位的位置信息。杨科<sup>[3]</sup>等利用裁剪后的YOLO-v3模型对全景俯视图中的车位角进行检测。Wu<sup>[4]</sup>等基于停车位检测(Parking Slot Detection, PSDet)的体系结构,提出采用圆形表征方式对车位角中心点和车位线的方向进行描述,并利用该模型进行感知识别。上述文献多数采用拼接后的俯视图像作为算法处理的对象,往往存在图像视野减小,且立体信息丢失的问题,若车位内侧存在障碍物,易因图像信息的不完整造成空车位的误识别。同时,上述文献多采用矩形框表征车位角,以矩形框中心点作为车位边界的临界位置,造成识别车位存在偏差,车位角中心点识别结果的偏差同样会造成车位姿态计算的误差。

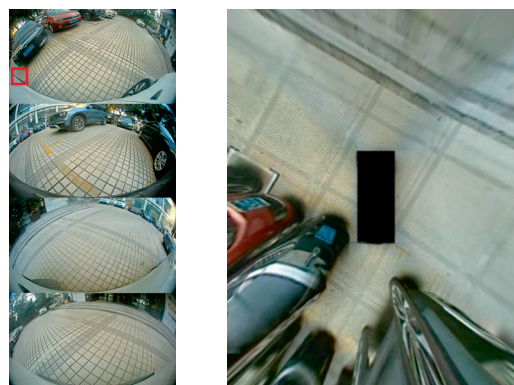
为解决泊车精度不足的问题<sup>[5-6]</sup>,本文利用CNN模型对输入图像进行特征提取,并对浅层细节特征与深层抽象特征进行拼接,然后利用Transformer模型<sup>[7]</sup>的“编码-解码”机制对CNN提取到的多层图像特征进行推理,并在不同序列特征上叠加位置信息,最后采用前馈神经网络(Feed-Forward neural Networks, FFN)对Transformer模型的解码结果进行推理,实现对空车位入口中心点、车位角中心点的感知,并通过实车测试验证该算法的有效性。

## 2 空车位精确感知算法

### 2.1 自动泊车车位精确感知算法架构

为构建完整的车身周围环境,本文共安装4路180°鱼眼摄像头,分别位于前、后侧保险杠面罩上方和左、右侧后视镜下方,安装过程中保证相邻摄像头之间存在一定的视野重叠。

部分场景下临近车位的车位角能够同时出现在2个摄像头的视野范围内,当一侧视野被遮挡时,临近摄像头的感知结果仍可采用,而全景图像拼接过程中对单幅图像的裁剪可能导致有效信息丢失,故本文以原始鱼眼图像作为模型的输入。为验证俯视拼接全景图像与原始图像在视野上的差异性,本文对摄像头的内参、外参进行了标定<sup>[8-9]</sup>。4路原始鱼眼图像和俯视全景拼接图像如图1所示。图1a中左视图图像在标注的红框处能够清楚地显示目标车位左侧的外侧车位角,但在俯视全景图像拼接过程中,单幅图像的裁剪导致该车位角丢失。



(a) 鱼眼图像

(b) 俯视全景图像

图1 鱼眼图像与俯视全景图像视野对比

对采集到的图像在保证原始图像比例的基础上进行尺度归一化处理作为深度学习模型的输入;模型的输出结果为图像上存在的车位角中心点的坐标和空车位入口位置对应2个车位角中心点中点的坐标,和对应坐标的类别及其可信度。在模型训练阶段,本文利用颜色空间转换、直方图均衡化、中值滤波等图像处理算法对数据集进行增强处理,以提高模型的泛化能力。最后,利用离线数据集对训练的模型进行量化评估,并进行实车验证。

### 2.2 基于CNN-Transformer的自动泊车车位感知模型

为提高模型推理的实时性,本文搭建主分支包含12层卷积层的CNN模型作为图像特征提取的基础网络,并对主分支的中间层多层特征进行拼接处理,然后利用Transformer模型结构对平铺后的图像特征进行推理计算,最后利用2个FFN分支对类别和坐标分别进行推理。

本文最终采用的模型结构如图2所示,主要包括CNN图像特征提取模块和Transformer感知推理模块,其中, $N_1$ 、 $N_2$ 分别为编码层和解码层的数量。鱼眼摄像头原始图像分辨率为1 280×720,为保证缩放后图像中目标不出现二次形变,本文在图像输入模型前先对样本的上、下边缘以黑色像素填充至1 280×1 280,在保证总体填充大小一致的情况下,上、下边缘填充大小分别随机生成;然后将图像尺寸缩放至模型的输入尺寸512×512;模型的输出为16×1×3的感知分类结果和16×1×2的感知坐标结果。感知分类结果代表每个预测点坐标所属的分类,为归一化至0~1范围内的浮点数,代表各类别的置信度;感知坐标结果代表在图像坐标系下各预测点的像素坐标。模型的CNN图像特征提取模块由3个分支构成,本文CNN模型的主干分支选取DarkNet-19模型结构前13个卷积层及卷积层后的13个激活函数层和4个池化层作为主干分支,依次将其中的卷积层

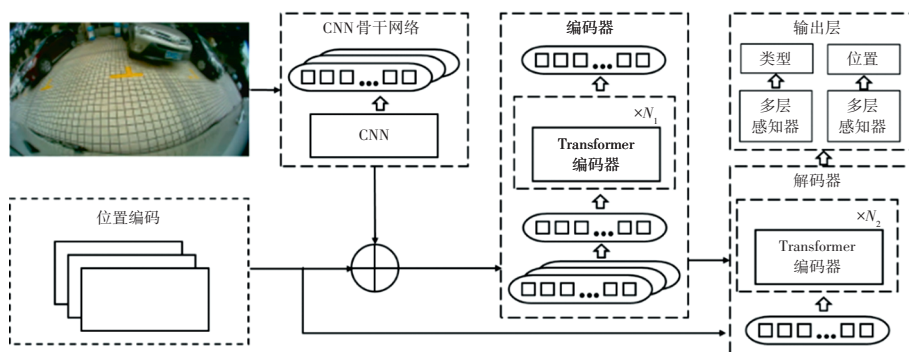


图2 基于 CNN-Transformer 的自动泊车车位感知模型结构

裁剪成 [16, 32, 64, 32, 64, 128, 64, 128, 256, 128, 256, 128, 256], 并在最后添加一层池化层对提取的特征进一步下采样。主干分支中均采用核大小为  $3 \times 3$  的卷积层, 激活函数选用修正线性单元 (Rectified Linear Unit, ReLU), 池化层采用核大小为  $2 \times 2$  的最大值池化。为了增加模型的感受野, 提取不同尺度的特征, 本文对 DarkNet-19 模型进行扩展, 在主干分支的第 6 层后引入 2 个不同尺度的分支。2 个分支结构与主干分支结构一致, 但分别采用空洞尺度为  $K=3$  和  $K=7$  的  $3 \times 3$  卷积核。本文 CNN 图像特征提取模块结构如图 3 所示, 为简化模型结构可视化效果, 图中不考虑各层提取特征的维度。

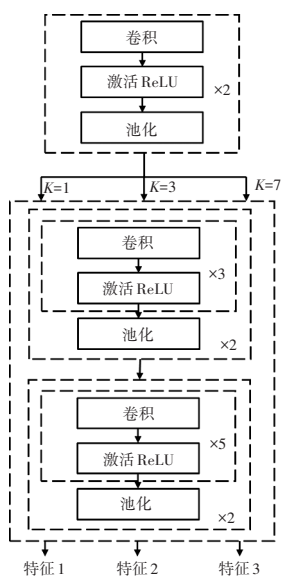


图3 CNN 图像特征提取层结构

CNN 图像特征提取模块最终输出 3 个分支提取尺度均为  $16 \times 16 \times 256$  的图像特征, 对特征进行拼接后构成  $16 \times 16 \times 768$  的特征, 前 2 个通道展开后构成 256 个  $1 \times 768$  的词向量, 作为后续 Transformer 感知推理模块的输入。

模型的 Transformer 感知推理模块采用 DETR (DEtection TRansformer) 模型<sup>[10]</sup> 中的编码器-解码器 (Encoder-Decoder) 结构, 如图 4 所示。模型中间层充分利用不同特征之间的相对位置关系叠加了特征位置

编码信息, 增强了特征对位置信息的表达, 将 CNN 图像特征提取模块得到的铺平特征叠加对应的位置编码信息后通过编码模块得到编码后的特征, 将其输入解码模块进行解码, 其中,  $N_1=N_2=6$  层, 最后, 分别通过 3 层 FFN<sup>[11]</sup> 得到维度为  $16 \times 1 \times 3$  的感知分类结果和维度为  $16 \times 1 \times 2$  的感知坐标结果。编码模块由多头注意力机制层和前馈神经网络层等基础层构成, 输入的特征拼接、平铺的图像特征首先叠加位置编码信息后经过多头注意力机制层, 编码层的输出与输入特征相加后经过前馈神经网络进行编码特征维度的变化。解码模块与编码模块结构组成类似, 不同的是解码模块中解码层的  $V$  值和  $K$  值来源于编码层模块的输出, 而  $Q$  值来源于上一时刻解码模块的输出。编码模块和解码模块的结构如图 4 所示。

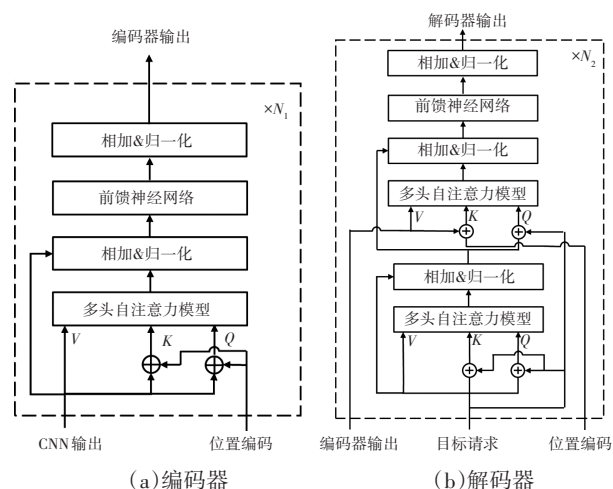


图4 DETR 模型中编码器和解码器结构

在解码模块后连接 2 个前馈神经网络模块, 分别用于目标类别的预测和目标坐标的预测。通过对摄像头的覆盖范围、实际应用需求等方面进行分析, 单幅图像最多能够涵盖目标车位和周围 4 个车位的相关信息, 本文最终预测的目标不多于 32 个, 则类别预测和坐标预测前馈神经网络的输出结果维度分别为  $32 \times 3$  和  $32 \times 2$ : 类别预测结果对应 32 个预测点属于各类别的可信度,

其中类别包括无目标点、车位角中心点和空车位入口中心点;坐标预测结果为图像坐标系下点的像素坐标,类别与类别预测结果相对应。类别预测输出结果经归一化指数函数(Softmax函数)处理,坐标预测输出结果经S型函数(Sigmoid函数)处理。

### 3 数据集采集与处理

数据集的采集与标注直接影响模型训练的准确性。本文数据集采集设备为4路180°广角鱼眼摄像头,对所采集图像中车位角中心点的坐标和车位入口中心点的坐标分别进行标注。车位角中心点坐标为2条车位线中线交叉点坐标,车位入口中心点坐标为车位入口对应2个车位角中心点中点的坐标。数据集标注示意如图5所示。“T”型车位角通常为2个相邻车位共用的车位角,本文的标注方式下单幅图像样本中单个车位角只存在唯一的车位角中心点标注结果。

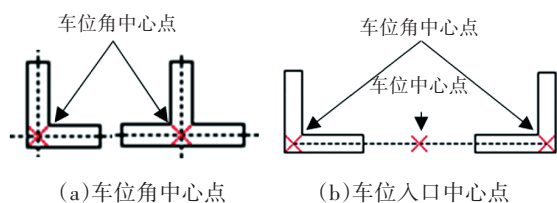


图5 数据集标注方式示意

为匹配后续的模型训练与测试,在保存坐标值时对标注信息进行归一化处理,使所有坐标值均映射至0~1范围内:

$$x' = x/W \quad (1)$$

$$y' = y/H \quad (2)$$

式中: $x$ 、 $y$ 分别为图像坐标系下的横、纵坐标, $W$ 、 $H$ 分别为图像的长度、宽度, $x'$ 、 $y'$ 分别为经过归一化处理后的横、纵坐标。

本文数据集分别采集了晴天、雨天、室内等不同场景、不同车位线划线方式下的5 530张原始鱼眼图像,包括车位线磨损以及车位角遮挡严重的场景图像,其中,晴天与雨天的数据量比例为3:1,清晰车位与非清晰车位比例为4:1,按场景类别选取其中20%的数据构成后续离线测试的测试集。在模型训练过程中按一定概率对训练图像随机采用颜色空间转换、直方图均衡化、随机尺度裁剪等传统图像处理方法实现模型输入图像的增强<sup>[12]</sup>,如图6所示,以达到扩充数据集和增强模型泛化能力的目的。在模型训练阶段,多种图像增强算法的引入能够较好地提高模型训练的泛化性能。另外,在灰度图转换中,会将单通道的灰度图重新转化为三通道图像,以适配模型的输入尺寸。

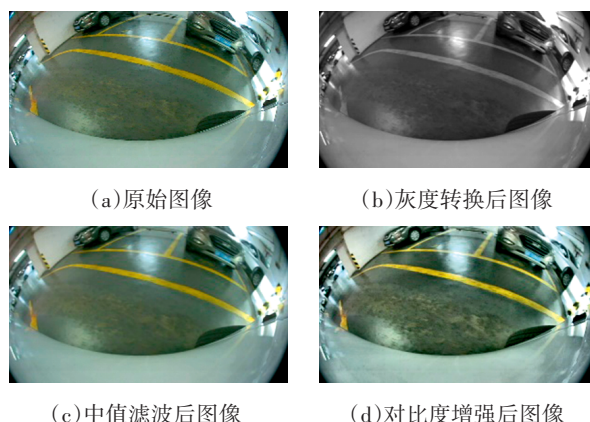


图6 数据增强可视化效果

### 4 模型训练与结果

#### 4.1 自动泊车车位感知模型损失函数

本文模型的输出序列表示为 $y'_i = (c'_i, p'_i)$ ,其中 $p'_i$ 为预测点 $i$ 的坐标, $c'_i$ 为预测点 $i$ 对应类别的置信度, $i \in [1, N]$ , $N$ 为预测的目标数量,本文取 $N=32$ 个;对应真值序列表示为 $y_i = (c_i, l_i)$ ,其中 $l_i$ 为点 $i$ 的真实坐标, $c_i$ 为点 $i$ 对应类别的置信度, $i \in [1, N]$ 。使用二元匹配保证在32个预测点与真实标签点之间实现最大匹配,标签点不足32个时用空值按对应类别进行补齐操作,以保证训练过程中数据能够绝对对齐。定义二元匹配的匹配策略 $\delta'$ 为:

$$\delta' = \arg \min \sum_{i=1}^N \left[ -\mathbf{1}_{\{c_i \neq \emptyset\}} p'_{\delta(i)}(c_i) + L_p(l_i, p'_{\delta(i)}) \right] \quad (3)$$

与原始二元匹配算法不同,本文使用欧式距离作为点与点之间的距离度量:

$$L_p(l_i, p'_{\delta(i)}) = \mathbf{1}_{\{c_i \neq \emptyset\}} \|l_i, p'_{\delta(i)}\|_2 \quad (4)$$

式中: $\delta(i)$ 为与第 $i$ 个目标匹配的边界框的索引值, $p'_{\delta(i)}(c_i)$ 为目标被预测为对应类别的概率, $l_i$ 为样本对应坐标真值集合中的二维向量, $p'_{\delta(i)}$ 为通过二元匹配算法与二维真值向量相匹配的二维预测值向量, $L_p(l_i, p'_{\delta(i)})$ 为坐标点 $l_i$ 与坐标点 $p'_{\delta(i)}$ 间的欧式距离, $\mathbf{1}_{\{c_i \neq \emptyset\}}$ 表示 $c_i$ 不为空时为1。

本文算法的损失函数 $S$ 为:

$$S = \sum_{i=1}^N \left[ -\log p'_{\delta(i)}(c_i) + \lambda_i L_p(l_i, p'_{\delta(i)}) \right] + \gamma \|w\|_2 \quad (5)$$

式中: $\lambda_i$ 为权重系数,预测目标为车位角中心点、车位入口中心点时取值分别为1.0、0.2; $\gamma$ 为正则化系数,本文取为 $10^{-5}$ ;  $w$ 为训练网络权重。

式(5)等号右侧的前半部分为分类损失,后半部分为坐标预测损失。本文预测对象分为车位角中心点和车位入口中心点两类,两类对象对精度的要求存在差异。车位角中心点的偏差会直接导致车位边界的识别出现误差和车位的姿态出现畸形,而车位入口中心点的

预测仅用于车位的辅助判别,其识别结果只需在目标车位两个外侧车位角点中间区域即可,故两类点在损失函数中的权重系数差异较大。

#### 4.2 自动泊车车位感知模型训练结果评价

为保证模型训练阶段数据能够对齐,针对单张图片输入,本文模型恒定得到32个点的坐标预测结果和每个坐标预测结果对应的类别可信度。为了评估模型的预测性能,通过准确率、召回率、平均误差对模型进行评价。其中,准确率为预测正确的结果占总样本的比例,召回率为数据集中标注目标被预测正确的概率。目标预测判定为正确的判定依据为预测点与实际点之间的误差小于1%,两个坐标点之间的误差计算公式为:

$$E = \sqrt{\|p, p'\|} \quad (6)$$

式中: $p$ 为样本的真实坐标值, $p'$ 为归一化后本文算法的预测坐标值, $e$ 为预测值与真实值之间的绝对误差。

基于本文预测对象以点作为表征形式与传统目标以矩形框作为表征形式的差异性,本文首先对每类预测点通过二元匹配算法确定模型预测结果与真值之间的对应关系,即确定预测结果中坐标点 $p'_{\delta(i)}$ 为对真值中坐标点 $l_i$ 的预测结果,计算 $p'_{\delta(i)}$ 和 $l_i$ 之间坐标系归一化后的欧式距离 $\|l_i, p'_{\delta(i)}\|_2$ 作为误差的度量标准,通过误差是否小于阈值 $T$ 判定是否检测到目标。本文按照不同目标尺寸之间的比例关系,在计算车位角中心点之间距离时 $T$ 取值为8.0像素,在计算空车位入口中心点之间距离时 $T$ 取值为16.0像素,坐标点均为换算至实际图像坐标系下后的像素坐标值。

为了与传统计算机视觉算法进行对比,本文基于文献[6]设计了相应的车位检测算法。测试数据集包含晴天、雨天、室内等场景下的不同车位划线方式下的1906张180°鱼眼图像,本文算法与传统视觉算法的测试结果如表1和表2所示。表1中测试指标为选取图像样本中视野范围15 m内的目标作为原始数据的统计结果,其中包含7287个车位角和675个空车位;表2的中测试指标为选取视野范围10 m内的目标作为原始数据的统计结果,其中包含3621个车位角和666个空车位。为保证评测算法的准确性,平均误差的计算均以图像输入尺寸512×512作为基准值。

表1 视野范围10 m内车位检测测试指标

项目		准确率	召回率	平均误差/像素
本文模型	车位角中心点	0.962	0.986	2.98
	车位入口中心点	0.932	0.985	7.32
传统视觉算法	车位角中心点	0.936	0.954	3.46
	车位入口中心点	0.919	0.952	7.95

表2 视野范围15 m内车位检测测试指标

项目		准确率	召回率	平均误差/像素
本文模型	车位角中心点	0.921	0.983	3.28
	车位入口中心点	0.911	0.976	8.56
传统视觉算法	车位角中心点	0.867	0.939	4.35
	车位入口中心点	0.858	0.935	9.60

由表1和表2可知,与传统视觉算法相比,对于近视野范围(10 m)内的车位角中心点,本文模型平均误差减小了13.9%,对于远视野范围(15 m)内的车位角中心点,本文模型平均误差减小了24.6%。从对比结果可以看出,本文模型的准确率和召回率均有所提高,平均误差较大幅度降低,且在远视野范围具有更好的效果。针对雨天、车位线磨损以及车位角遮挡严重的场景,本文模型也具有更高的适应性。

为验证本文算法的有效性,对各摄像头进行内参和外参的标定,通过标定得到的参数将所有预测点映射至车辆坐标系下,统计可得:近距离车位角中心点预测结果平均误差控制在3 cm以内,能够满足实际应用中对准度的要求;不同视野范围的空车位入口中心点准确度均可达到90%以上。

鱼镜头成像过程中引入了畸变,图像单个像素点代表的实尺度可能存在差异,导致像素坐标系下的偏差无法真实反映无畸变条件下的偏差。为充分验证本文算法在实际应用中的识别精度,对样车上4路摄像头的内、外参数分别进行标定,将模型的识别结果映射至车辆坐标系下,图像对应的标注结果按同样的方式映射至车辆坐标系下。通过坐标转换换算至车辆坐标系下,车位入口中心点的预测结果小于10 cm,远小于车位的实际宽度,能够满足实际应用对预测精度的要求。

不同场景、不同车位划线方式下实车测试效果如图7所示,其中红色点表示车位角中心点,绿色点表示车位中心点。

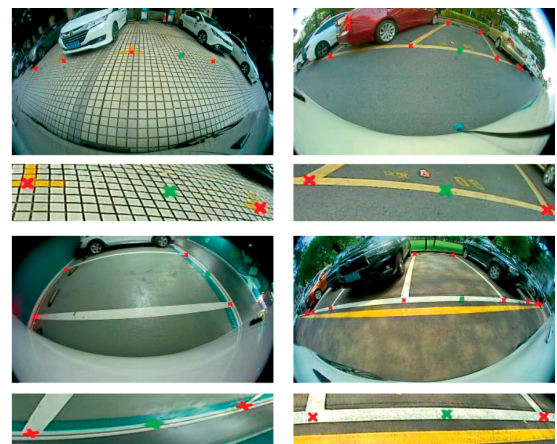


图7 不同场景下模型识别结果可视化

## 5 结束语

本文基于CNN模型对输入图像进行特征提取,利用Transformer模型的“编码-解码”机制对CNN提取到的多层图像特征平铺后进行计算推理,通过前馈神经网络推理得到车位入口中心点、车位角中心点的预测结果。通过包含不同场景、不同车位类型数据集的测试,相比于传统视觉算法,视野范围10 m内的车位角中心点识别平均误差降低了13.9%,在3 cm以内,视野范围15 m内的车位角中心点识别平均误差降低了24.6%,针对远视野、雨天、车位线磨损以及车位角遮挡严重的场景,CNN-Transformer模型也具有更好的效果。

但目前模型对于地下室光线昏暗以及车位磨损严重等场景识别率较低,后续需通过增加更多类似场景数据提高识别率。同时,模型对算力要求较高,后续可通过模型压缩和蒸馏等方式降低算力需求。

### 参 考 文 献

- [1] 张春洲,石晶. 基于ADAS实验平台的自动泊车系统研究[J]. 汽车实用技术, 2019(16): 41-42.  
ZHANG C Z, SHI J. Research on Automatic Parking System Based on ADAS Experimental Platform[J]. Automobile Applied Technology, 2019(16): 41-42.
- [2] 朱强. 基于全景相机的自动泊车系统设计与实现[D]. 成都: 电子科技大学, 2020.  
ZHU Q. Design and Implementation of an Automatic Parking System Based on a Panoramic Camera[D]. Chengdu: University of Electronic Science and Technology of China, 2020.
- [3] 杨科,王炜斌,徐维庆. 基于小型深度学习网络的车位检测方法[J]. 汽车实用技术, 2020(13): 155-158.  
YANG K, WANG W B, XU W Q. A Parking Spot Detection Method Based on Small Deep Learning Network[J]. Automobile Applied Technology, 2020(13): 155-158.
- [4] WU Z Z, SUN W W, WANG M, et al. PSDet: Efficient and Universal Parking Slot Detection[C]// 2020 IEEE Intelligent Vehicles Symposium (IV). Las Vegas, NV, USA: IEEE, 2020: 290-297.
- [5] 朱旺旺,黄宏成,马晋兴. 基于图像识别的泊车车位检测算法研究[J]. 汽车工程, 2019, 41(7): 744-749.  
ZHU W W, HUANG H C, MA J X. A Research on Parking Space Detection Algorithm Based on Image Recognition[J]. Automotive Engineering, 2019, 41(7): 744-749.
- [6] 黄晨,刘泽,罗禹贡,等. 基于背景光照去除和连通区域的车位检测[J]. 汽车工程, 2020, 42(1): 47-51.  
HUANG C, LIU Z, LUO Y G, et al. Parking Space Detection Based on Background Illumination Removal and Connected Regions[J]. Automotive Engineering, 2020, 42(1): 47-51.
- [7] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is All You Need[C]// Proceedings of the 31st International Conference on Neural Information Processing Systems. NY, USA: Curran Associates Inc., 2017: 6000-6010.
- [8] ZHANG Z Y. Flexible Camera Calibration by Viewing a Plane from Unknown Orientations[C]// Seventh IEEE International Conference on Computer Vision. Kerkyra, Greece: IEEE, 1999.
- [9] 范勇,周敬滨,王传珂,等. 一种实用的平板高精度摄像机标定方法[J]. 计算机测量与控制, 2010(1): 172-174.  
FAN Y, ZHOU J B, WANG C K, et al. Practical High Precision Camera Calibration Method from a Plane[J]. Computer Measurement and Control, 2010(1): 172-174.
- [10] CARION N, MASSA F, SYNNAEVE G, et al. End-to-End Object Detection with Transformers[C]// European Conference on Computer Vision. Cham, Switzerland: Springer, 2020.
- [11] ZHU X Z, SU W J, LU L W, et al. Deformable DETR: Deformable Transformers for End-to-End Object Detection [EB/OL]. (2021-03-18)[2023-03-01]. <https://arxiv.org/abs/2010.04159>.
- [12] SHORTEN C, KHOSHGOFTAAR T M. A Survey on Image Data Augmentation for Deep Learning[J]. Journal of Big Data, 2019, 6(1): 1-48.

(责任编辑 斛 畔)

修改稿收到日期为2024年2月20日。