中图法分类号:TP391.41　　**文献标识码**:A　　**文章编号**:1006-8961(XXXX)XX-0001-16

论文引用格式:Xu Guangzhu, Wu Yunqi, Ke Zhi, Lei Bangjun. Lightweight wide-range license plate recognition network with parallel encoder-decoder architecture[J/OL]. Journal of Image and Graphics, XXXX:1-16. DOI: 10.11834/jig.260305.(徐光柱, 吴昀其, 柯志, 雷帮军. 并行编解码架构的轻量级广域车牌识别网络[J/OL]. 中国图象图形学报, XXXX:1-16. DOI: 10.11834/jig.260305.)[DOI:10.11834/jig.260305]

并行编解码架构的轻量级广域车牌识别网络

徐光柱[1,2],吴昀其[2],柯志[2],雷帮军[3,4]

1. 湖北省水电工程智能视觉监测重点实验室(三峡大学)　宜昌　443002; 2. 三峡大学计算机与信息学院　宜昌　443002; 3. 数字金融创新湖北重点实验室　武汉　430205; 4. 湖北经济学院信息工程学院　武汉　430205

摘　要:**目的**　开放场景中,待识别车牌版式多样、字符位数不统一,现有车牌识别算法在通用性、识别精度与推理速度方面通常难以兼顾。另外,双行及复杂背景车牌数据集的匮乏,限制了多类型车牌识别技术的发展。针对上述问题本文提出基于并行编解码架构的广域车牌识别网络,同时提出基于冗余回收策略的多类型车牌数据生成方案。**方法**　首先通过专用轻量级卷积网络,实现车牌字符的多尺度特征高效提取。随后,将骨干网特征送入局部-全局交互模块,通过并行分组Transformer实现局部与全局特征的高效交互建模。接着,经过模态适配器聚合后送入并行查询式非自回归双层解码器,实现先粗后精的递进式匹配识别。数据生成方案通过回收单行测试数据集中的冗余样本,利用其背景及车牌区域的位置参数,结合程序化生成的虚拟目标车牌进行几何贴合,构建虚实结合的多类型车牌实验数据。**结果**　实验在CCPD数据集及通过冗余回收策略构建的混合数据集上开展,并与18种有代表性的开源及闭源算法进行对比,同时在两个公开数据集上进行跨数据集测试。结果表明,所提算法在通用性、识别精度和推理速度上均显著优于对比算法,在CCPD数据集上平均车牌识别准确率为99.05%,模型参数量为1.44 M,推理帧率为208.6 FPS。**结论**　基于冗余回收的数据生成方案有效缓解了多类型车牌样本不足的问题;基于并行编解码架构的轻量级车牌识别网络实现了通用性、准确性与实时性的协同优化,具有良好的理论研究与工程应用价值。

关键词:通用车牌识别;开放场景;Transformer;并行编解码架构;双行车牌合成

Lightweight wide-range license plate recognition network with parallel encoder-decoder architecture

Xu Guangzhu[1,2], Wu Yunqi[2], Ke Zhi[2], Lei Bangjun[3,4]

1. Hubei Key Laboratory of Intelligent Visual Monitoring for Hydropower Engineering (China Three Gorges University), Yichang 443002, China; 2. College of Computer and Information Technology, China Three Gorges University, Yichang 443002, China; 3. Hubei Key Laboratory of Digital Financial Innovation, Wuhan 430205, China; 4. School of Information Engineering, Hubei University of Economics, Wuhan 430205, China

Abstract: **Objective** In complex open scenarios with diverse license plate layouts, backgrounds and character lengths, existing license plate recognition models generally have insufficient general adaptability and cannot balance recognition accuracy and real-time speed. Severe shortage of samples for two-row structured, complex-background and variable-length character license plates further hinders the development of general-purpose license plate recognition research. To solve these problems, this paper proposes a wide-range license plate recognition network PL-TransLPRNet based on a lightweight parallel encoder-decoder architecture, and a multi-type license plate data generation scheme with a redundancy recycling

收稿日期:2026-05-27;**修回日期**:2026-09-03

基金项目:数字金融创新湖北省重点实验室开放基金(DFIKD2025D08)

strategy. **Method** TransLPRNet is composed of a lightweight visual encoder, a modal adapter and a parallel decoder based on learnable query tokens. In the lightweight visual encoder, two-stage cascaded high-channel downsampling-free inverted residual modules combined with full residual connections are adopted. Features from initial downsampling convolution are stacked layer-by-layer with two-stage enhanced features to achieve efficient extraction and aggregation of three-level progressive multi-scale features of license plate characters. On this basis, a third-stage downsampling inverted residual module is introduced, and a residual branch based on max-pooling is embedded to make up for the loss of key stroke information caused by downsampling. The extracted features are then input into a refined lightweight MobileViT module. Parallel grouped Transformer layers realize parallel interactive modeling of local and global features, which are aggregated by an adapter and then sent to a parallel Transformer decoder. This decoder abandons traditional autoregressive sequence dependency and adopts a one-step parallel decoding strategy with two cascaded decoding layers. In the first layer, learnable pattern query vectors and position query vectors are used to construct query tokens; through self-attention and cross-attention, coarse matching results are obtained. In the second layer, the coarse results are refined via self-attention for error correction and then cross-attention for fine matching, followed by a feed-forward network and a linear classifier to output class probability distributions for each character position in a batch decoding manner. Thus, the decoder adaptively identifies various license plates with variable layouts and character lengths. Aiming at the scarcity of two-row license plate data, a redundancy recycling strategy is proposed. Easily recognizable license plates from single-row test datasets are recycled as redundant samples. Virtual target license plates are programmatically generated and geometrically fitted by using their real backgrounds and geometric parameters of license plate regions. Without expanding dataset scale or generating extra collection costs, this strategy provides reliable data support for the training and verification of general-purpose license plate recognition models, and further constructs a hybrid dataset with both real and virtual samples covering diverse plate types. **Results** Based on the widely used CCPD dataset, two hybrid real-virtual datasets are constructed with real backgrounds and geometric parameters of license plate regions of redundant samples from its test set. One dataset contains 7-digit and 8-digit single- and two-row Chinese license plates, and the other covers 6-digit and 7-digit New York State license plates as well as 7-digit Chinese license plates. Models trained on the three datasets are comprehensively compared with 10 open-source and 8 closed-source license plate recognition algorithms. To compare fairly with closed-source algorithms, four license plate region extraction methods are applied to reduce interference brought by different license plate localization algorithms. Cross-dataset evaluation is carried out with the CLPD and PKUdata datasets to test the cross-dataset recognition performance of the proposed model. Experimental results show that the proposed model significantly outperforms the comparison algorithms in terms of generalization performance, recognition accuracy and inference efficiency. It achieves an average license plate recognition accuracy of 99.05% on the CCPD dataset. The proposed network has only 1.44 M parameters and yields an inference frame rate of 208.6 FPS, demonstrating strong cross-dataset generalization ability. **Conclusion** The data augmentation strategy based on redundancy recycling proposed in this paper effectively relieves the shortage of multi-type license plate experimental data. The proposed model solves the problem of unbalanced generalization, accuracy and real-time performance in license plate recognition from the architectural perspective, and has great practical value and deployment potential.

Key words: general license plate recognition; open environment; Transformer; parallel encoder-decoder architecture; double-row license plate synthesis

论文引用格式:DOI:10.11834/jig.260305

0 引 言

开放场景下的车牌识别技术被广泛应用于智能交通管控、城市停车管理、卡口安防监测等实际场景,其优势在于无需对车辆运行状态进行额外约束即可完成自动识别。然而,在实际应用环境中,复杂光照、恶劣天气、车牌遮挡或污损等因素会导致图像质量下降;同时,拍摄视角与车辆姿态的变化也会进一步加大识别难度,影响车牌识别准确率(Xu等,2026;Mu和Xu,2023)。

在车牌识别中,卷积神经网络(convolutional neural networks, CNN)和循环神经网络(recurrent

neural network，RNN）相结合的方案被广泛采用。为解决CNN中卷积感受野重叠导致相邻字符特征相互干扰、序列识别困难的问题，基于CNN与RNN的CRNN类车牌识别方法通常引入长短期记忆网络（long short-term memory，LSTM）（Li等，2018）或双向LSTM（Zou等，2020；Hua等，2024），利用其序列建模能力融合多局部特征来提升识别精度，但推理速度慢。

基于纯CNN的方法（如LPRNet）（Zherzdev和Gruzdev，2018）通过横向卷积增强字符间关系建模，在保持较好识别性能的同时有效提升了推理速度。为解决不定长字符序列的对齐问题，这类方法通常结合注意力机制（Mu和Xu，2023；Zou等，2020）或连接时序分类（CTC）损失函数（Shi等，2016）及动态解码策略。

然而，当车牌图像因拍摄角度产生透视畸变，导致字符尺度变化或形变时，上述方法易出现字符误识别或漏识别问题，在基于矩形框定位方案中则更为明显。相比之下，基于顶点的车牌定位方法通过预测到的车牌四个顶点坐标，可实现更精确的几何校正（Wang等，2022；Fan等，2026），有效提升识别准确率。但该类方法对数据标注精度要求高，且公开可用的顶点标注样本有限。

另一方面，目前的车牌识别研究主要集中在单行车牌场景，对双行及复杂背景类车牌关注不足。这主要是由于传统CRNN架构在特征提取与建模过程中更侧重于水平方向的序列依赖关系，难以有效刻画双行车牌中字符的垂直空间结构（Mu和Xu，2023）。此外，多类型车牌数据集，特别是双行车牌样本的缺乏，进一步限制了相关研究的发展。

随着视觉Transformer的快速发展，基于Transformer编解码架构的模型凭借其全局上下文建模能力，在OCR领域得到广泛关注，如TrOCR（Li等，2023）也为车牌识别提供了新的思路。与CRNN的序列建模方式不同，Transformer可将车牌图像的全局信息编码至各图像词元中，解码器则通过注意力机制动态聚合词元特征，提升字符识别的精度与鲁棒性。但这类方法通常采用自回归的解码方式，推理速度慢，部署成本高，无法满足开放场景下车牌识别算法的实时性与轻量化要求。

为此，本文提出轻量级并行编解码架构的广域车牌识别网络，主要贡献如下：

1）所提出的PL-TransLPRNet通过渐进式多尺度特征提取和局部-全局并行交互建模以及非自回归并行查询式双层解码器，从架构层面上有效解决了现有车牌识别模型难以兼顾通用性、识别精度与推理速度的问题。

2）提出冗余回收数据增强策略。通过采样数据集中易识别样本作为冗余，并在其车牌区域贴合程序化生成的多类型车牌图像。无需额外采集真实样本，即可为多类型车牌识别补充训练与验证数据，有效缓解双行车牌样本稀缺的问题。

1 相关工作

1.1 车牌识别相关算法

Li等人（2018）采用双向长短期记忆网络（bidirectional long short-term memory，Bi-LSTM）建模车牌字符特征，经非线性变换和Softmax得到字符的概率分布，再利用CTC解码，较早实现了免字符分割的端到端车牌识别，但易受图像倾斜与形变影响。

Luo等人（2020）采用生成对抗网络与Bi-LSTM探索了单行模糊车牌图像的识别。Zou等人（2020）采用Bi-LSTM生成的1D注意力图强化字符区域特征，再经线性分类器逐字符识别，可减少字符间干扰并可适应一定的倾斜形变。但所用循环神经网络推理效率低，且1D注意力不能适用于双行车牌。

Mu和Xu（2023）借鉴DAN（Wang等，2020）的解耦式2D注意力，利用多尺度特征与轻量U-Net网络分割字符区域，实现单字符特征提取，同时采用多分类头完成车牌字符的并行识别。

相比于Zou等人（2020）采用的1D注意力机制，Mu和Xu（2023）提出的方法适用场景更广，也是少数可同时适配长短号牌的并行车牌识别方案。但这种基于字符分割的注意力方案受弱监督学习约束，对车牌中非字符空位区域建模能力弱，且缺失对车牌全局空间布局信息的建模能力。

为提升推理速度，Zherzdev等人（2018）提出了不依赖RNN与注意力机制的实时车牌识别网络LPRNet。该网络采用轻量级CNN提取多层特征，以宽卷积替代LSTM建模局部上下文，并结合CTC损失函数实现可变长度的序列识别。该网络具有较高的推理速度，在资源受限平台上具备应用潜力，但其识别精度有限，也无法直接适配双行车牌识别。

Qin等人(2020)在LPRNet的基础上提出Eulpr网络。该方法通过增加分支网络回归双行车牌中上下字符区域的划分比例,将车牌划分为上下两部分,并分别采用LPRNet解码器进行识别。但当车牌存在较强透视形变时,上下字符区域的相对位置与比例会发生明显偏移,导致识别性能下降。

上述车牌识别方案的长距离依赖建模能力有限,难以充分捕捉车牌字符间的布局上下文关系,在多类型车牌尤其是双行车牌识别中存在局限。

随着Transformer模型在视觉领域的快速发展,相关科研工作者开始关注基于Transformer的OCR及场景文本识别(scene text recognition,STR)方案。

通用OCR网络模型PaddleOCR(Li等,2022)自v3版本采用CNN与STR中广泛使用的SVTR(Du等,2022)融合架构,以专用骨干网PP-LCNet提取局部特征,用SVTR建模跨字符依赖,融合特征后由CTC完成解码。由于该模型面向单行文本任务设计,无法直接适配双行字符序列,需额外分行预处理;但SVTR模块本身具备双行文本建模能力,经专用数据集重训后可适配双行车牌识别。

Li等人(2024)将TrOCR(Li等人,2023)应用于车牌识别任务,并在FPGA平台上实现约20 FPS的推理速度,在RTX 3060 GPU上约为10 FPS。Ismail等人(2025)采用DeiT(Tan等,2022)作为图像编码器,并结合线性变换层实现可变长度车牌字符序列识别,在TITAN RTX GPU上的推理速度约为12 FPS。

Tao等人(2024)采用YOLOv5(Ultralytics,2020)的Focus模块与ResNet基础残差块组合提取车牌特征,再通过标准Transformer编解码架构以自回归方式实现车牌字符识别,其在RTX 2080Ti上的批次大小(batch size)为5时的推理速度为159 FPS,在实时性要求高的场景难以满足要求。

Chen等人(2023)等提出一种端到端的车牌检测与识别一体化网络。为提升推理效率,该方法采用骨干网络特征共享策略,通过二维门控空间自注意力进行全局特征建模,接着利用单层轻量级交叉注意力实现特征图中的字符对齐,实现并行解码。

上述并行多查询式交叉注意力对齐在STR中被广泛使用。SRN(Yu等,2020)首次提出基于加性注意力的并行视觉注意力PVAM。ABINet(Fang等,2021)将加性注意力替换为缩放点积注意力,并引入轻量U型模块优化注意力值。PARSeq(Bautista等,2022)通过在交叉注意力模块中内嵌掩码排列语言建模方式实现迭代式非自回归解码。

上述模型均只使用单层交叉注意力,解码端未设置查询词元交互建模模块,不同字符查询词元间易出现视觉特征抢占冲突,网络无原生分层纠错能力,只能依托外置语义分支或非自回归迭代推理修正对齐误差。并且要求待识别文本具有强语义相关性,对车牌识别适配度低。

纯视觉方案SVTRv1(Du等,2022)依靠单阶段无语言分支架构,采用局部-全局混合注意力实现多粒度视觉特征建模,在通用STR任务上性能优良。但模型采用三阶段多层Transformer堆叠方案,整体计算量大、推理时延偏高。SVTRv2(Du等,2025)新增语义引导模块SGM,默认字符存在连续上下文依赖,对车牌识别任务适配性弱。

1.2 车牌图像数据集

高质量车牌数据集是模型训练与性能评估的基础,但在现有真实车牌数据集中,除CCPD(Xu等,2018)与LSV-LP(Wang等,2022)外,大多数规模较小,且标注信息通常仅包含边界框,缺乏顶点级标注。尽管LSV-LP数据量约40万张,但存在较高冗余,其有效独立样本数量及场景多样性明显弱于CCPD。

合成数据集如Ukrainian LP(Shpir等,2024)多采用简化背景与理想视角,与真实复杂环境存在明显差异。此外,真实与合成数据中,双行车牌样本均极为稀缺,因此现有方法也普遍缺乏对双行车牌识别任务的系统性研究。为此,本文选择规模最大、有效样本丰富且标注质量高的CCPD作为主实验基准,依托其真实背景与拍摄视角,构建了两类补充数据:一类是双行中文车牌合成数据,另一类是背景及布局复杂的美国纽约州车牌合成数据。

2 本文算法

PL-TransLPRNet(图1,2)由视觉编码器,适配器和解码器组成。视觉编码器以多粒度字符特征的局部与全局联合建模为核心,以可学习位置编码为衔接桥梁,为解码器提供信息丰富的可区分性输

入特征;双层级联解码器,在模式查询向量和位置查询向量的双匹配保障下,以并行查询解码方式,

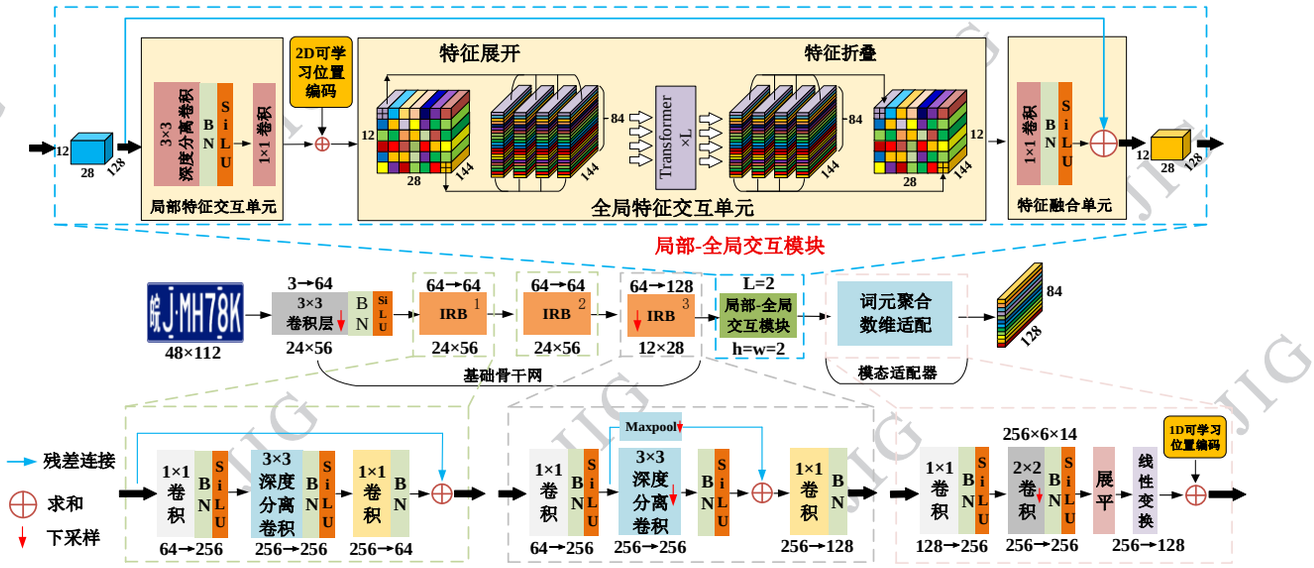


图 1　PL-TransLPRNet 编码器网络结构组成

Figure 1　Network architecture of the encoder for PL-TransLPRNet

实现首层粗匹配，二层纠错与微调的高效解码。

2.1　PL-TransLPRNet 视觉编码器与适配器

PL-TransLPRNet 编码器采用轻量化设计（图 1）。骨干网由 1 层 3×3 卷积与 3 个倒残差（inverted residual block，IRB）模块级联构成。常规图像分类网络首层卷积通道数常设为 16 并逐层递增，以适应大尺度图像的全局语义提取。但车牌字符尺寸小、排布紧凑，过深的卷积会使感受野超出字符范围，且缓慢升维易丢失细粒度的笔画细节。为此，本文骨干网将首层卷积通道设为 64，既能充分编码字符笔画特征，又为后续 IRB 模块提供恒定通道维度，无需中途调整；此外，采用恒定通道不会阻断残差路径且可使反向传播直达首层，强化源头特征学习。

无下采样且通道数不变的 IRB（图 1 中 $\mathrm{IRB}^{1,2}$），其主路径经 1×1 卷积解耦通道特征，再通过深度可分离卷积与 SiLU 激活函数强化筛选有效特征，最后由 1×1 卷积压缩冗余通道；残差分支直接传输细粒度特征，与主分支的输出逐元素相加，在保留车牌边缘、纹理细节的同时强化字符关键特征，提升特征表达的完整性与判别性，其过程为：

$$\boldsymbol{X}_1 = \sigma_S\left(\mathrm{BN}\left(\mathrm{Conv}_{1\times1}(\boldsymbol{X})\right)\right) \tag{1}$$

$$\boldsymbol{X}_2 = \sigma_S\left(\mathrm{BN}\left(\mathrm{DWConv}_{3\times3}(\boldsymbol{X}_1)\right)\right) \tag{2}$$

$$\mathcal{F}_m(\boldsymbol{X}) = \mathrm{BN}\left(\mathrm{Conv}_{1\times1}(\boldsymbol{X}_2)\right) \tag{3}$$

$$\boldsymbol{Y}_\mathrm{I} = \mathcal{F}_m(\boldsymbol{X}) + \boldsymbol{X} \tag{4}$$

式中 $\boldsymbol{X}\in\mathbb{R}^{H\times W\times C}$ 为输入特征；$\mathrm{Conv}_{1\times1}$ 表示 1×1 点卷积；BN 表示批归一化；$\mathrm{DWConv}_{3\times3}$ 表示 3×3 深度可分离卷积；σ_S 为 SiLU 激活函数；$\mathcal{F}_m(\boldsymbol{X})$ 表示主分支变换函数；$\boldsymbol{Y}_\mathrm{I}$ 则为 IRB 模块最终输出。

在需要下采样的 IRB 中（图 1 中 IRB^3），传统卷积降维易因加权求和而平滑字符区域的孤立强响应，造成笔画与边缘信息丢失。为此，本文在第三个 IRB 中引入最大池化残差分支。该设计可保留局部最强响应，其过程为：

$$\boldsymbol{X}_1 = \sigma_S\left(\mathrm{BN}\left(\mathrm{Conv}_{1\times1}(\boldsymbol{X})\right)\right) \tag{5}$$

$$\boldsymbol{X}_\mathrm{r} = \mathrm{MaxPool}_{2\times2,s=2}(\boldsymbol{X}_1) \tag{6}$$

$$\boldsymbol{X}_2 = \mathrm{DWConv}_{3\times3,s=2}(\boldsymbol{X}_1) \tag{7}$$

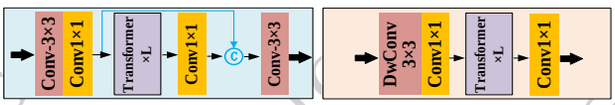
$$\boldsymbol{X}_3 = \sigma_S\left(\mathrm{BN}(\boldsymbol{X}_2)\right) + \boldsymbol{X}_\mathrm{r} \tag{8}$$

$$\boldsymbol{Y}_\mathrm{Id} = \mathrm{BN}\left(\mathrm{Conv}_{1\times1}(\boldsymbol{X}_3)\right) \tag{9}$$

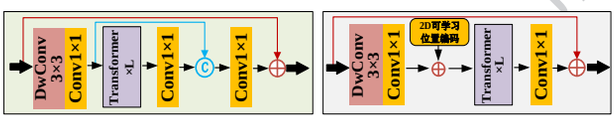
式中 $\boldsymbol{X}_\mathrm{r}$ 为残差分支特征；$\mathrm{MaxPool}_{2\times2,s=2}$ 为 2×2 最大池化；$\mathrm{DWConv}_{3\times3,s=2}$ 为步长为 2 的 3×3 深度可分离卷积；$\boldsymbol{Y}_\mathrm{Id}$ 为带下采样的 IRB 最终输出。

由于卷积网络难以有效建模字符序列依赖，纯 Transformer 计算开销又过高，而车牌识别需依托局部笔画与全局特征协同判断。为此本文引入 Mobile-ViT-v2（Mehta 等，2023）系列中窗口化的局部-全局交互模块（MobileViT 模块），并针对车牌字符识别任务，对原模块进行一定调整，在保留窗口化全局建模机制的同时，进一步降低计算开销，实现局部笔画与全局字符特征的协同优化。

如图 1 中蓝色虚线框所示，本文所用 MobileViT



（a） mobileViTv1方式　（b） mobileViTv2方式



（c） mobileViTv3方式　（d） 本文所用方式

图2　不同MobileViT模块组成结构对比

Figure 2　Comparison of the component structures of different MobileViT modules ((a) MobileViT-v1 scheme; (b) MobileViT-v2 scheme; (c) MobileViT-v3 scheme; (d) The proposed scheme)

模块由局部特征交互单元、全局特征交互单元和特征融合单元构成。与MobileViT-v3中的做法不同(Wadekar等,2022),本文所用的MobileViT模块仅在MobileViT-v2中的相应模块中增加了外层加性残差(如图2(d)中红线所示)。

局部特征交互单元通过深度可分离卷积完成局部信息交互,并经1×1卷积升维适配Transformer模块。MobileViT中窗口化Transformer仅提供隐式弱位置信息,而车牌识别对字符顺序与位置对齐的细粒度建模需求较高。因此,本文在全局交互前引入可学习2D位置编码,显式建模词元相对位置关系,强化跨窗口字符关联的空间约束,过程如下:

$$\boldsymbol{X}_L = \mathrm{Conv}_{1\times1}\left(\sigma_S\left(\mathrm{BN}\left(\mathrm{DWConv}_{3\times3}(\boldsymbol{X})\right)\right)\right)\oplus \boldsymbol{P}2 \tag{10}$$

式中$\boldsymbol{P}_2$为2D可学习位置编码,$\oplus$为加性融合操作。

在全局特征交互单元中,各2×2局部窗口中固定位置的特征被划分为四个词元分组,通过分组并行Transformer实现跨窗口的注意力建模,完成车牌字符的远程语义关联。经过全局交互建模后的图像词元,通过折叠操作恢复至原2D空间尺寸得到$\boldsymbol{X}_C$,并送入特征融合单元。然后经1×1卷积、BN与SiLU激活函数调整通道维度;接着与以模块输入特征为残差的$\boldsymbol{X}$逐元素相加,最终得到既保留车牌字符边缘、笔画等细粒度局部细节,又具备全局布局感知能力的互补特征,其过程为:

$$\boldsymbol{Y} = \sigma_S\left(\mathrm{BN}\left(\mathrm{Conv}_{1\times1}\left(\boldsymbol{X}_C\right)\right)\right)\oplus\boldsymbol{X} \tag{11}$$

为适配文本解码器语义对齐并提升计算效率,本文设计了专用模态适配器。适配器采用窗口聚合方式将编码器中2×2局部窗口特征进行聚合,并按顺序展开为一维词元序列。通过1×1卷积实现维度对齐后加入一维可学习位置编码,强化视觉词元中空间位置信息。

2.1.2　并行查询方式的非自回归解码器

并行查询方式的非自回归解码器(图3),由两层Transformer解码层堆叠而成,每层依次包含多头自注意力(multi-head self-attention,MHSA)、多头交叉注意力(multi-head cross-attention,MHCA)与前馈网络(feed-forward network,FFN)模块。模型以视觉词元和固定数量的可学习查询词元为输入,实现并行查询解码,无需自回归迭代,显著提升了推理效率。可学习查询词元由可学习位置查询向量与可学习模式查询向量相加构成。

解码器第一层中的多头自注意力模块(图3中部模块①)用于建模构成各查询词元的位置查询向量与模式查询向量的统计模式。双参数的设置配合多头自注意力模块中的K/Q/V矩阵,以数据驱动方式,可有效建模车牌图像中各字符的位置与特征分布,使得解码器第一层的查询词元,在交叉注意力模块(图3中部模块②)中就能有效锁定自己负责的字符对应的视觉词元,提取出对应的字符特征。

前馈网络(图3中部模块③)对子空间对齐特征执行通道域的非线性变换,强化车牌字符判别特征,提升序列识别区分度。

解码器第2层中的自注意力模块,负责纠错和自适应调整。当输入车牌图像出现特殊情况,如过度倾斜、扭曲时,第一层解码如果出现不同查询词元抢占了对方视觉词元的情况,第二层的自注意力模块会通过查询词元间交互,利用Q/K/V自相关运算剥离不属于各查询自身的视觉特征,实现纠错与精调,使其更聚焦关键视觉词元。训练阶段,解码器以可学习查询词元与视觉词元为输入,以车牌真实字符序列为监督信号,采用教师强制训练策略与交叉熵损失函数完成模型优化。推理阶段无需循环递归,直接利用训练好的固定查询词元实现并行解码,一次性输出完整车牌序列。

2.2　双行合成车牌数据集的构建

针对真实双行车牌数据匮乏问题,本文提出冗余回收策略。在冗余检测环节,首先将CCPD数据集中的基础子集随机均匀分为初始训练集(10万张)和初始测试集(10万张),与CCPD原论文一致。

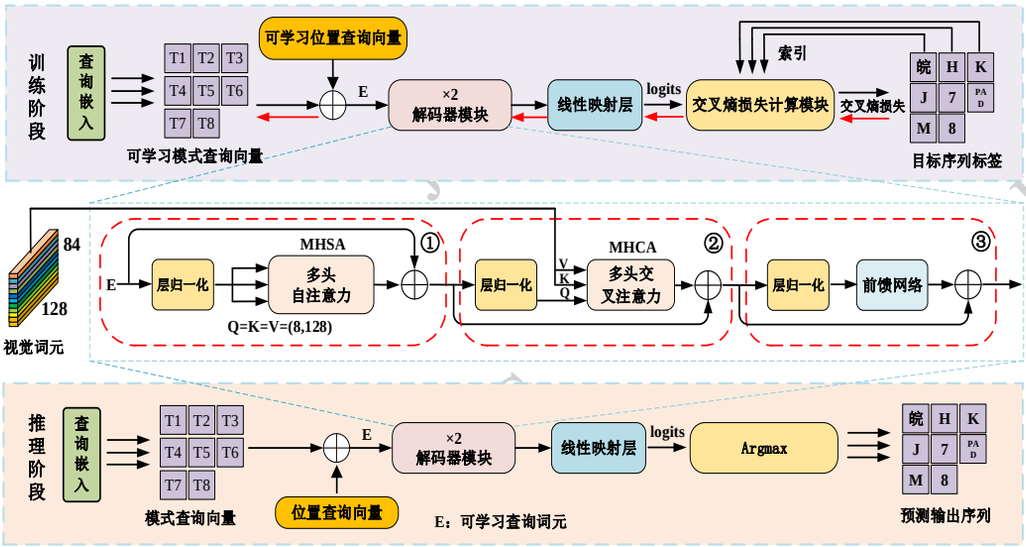


图3　并行查询式非自回归解码器网络结构组成图

Figure 3　Network architecture of the parallel-query-based non-autoregressive decoder

然后对CCPD论文所用模型RPNet(Xu等,2018)训练后,用其从测试集中筛选出识别正确的5万张车牌图像,作为冗余车牌样本。

随后,通过模板程序随机生成黄色与绿色双行车牌各2.5万张,经模糊化处理后,以回收冗余样本的真实背景为基础,并利用其具有的车牌顶点标签对生成的双行车牌进行透视变换后,覆盖原车牌区域。

接着,将生成的5万张双行车牌均匀分为两半,分别加入初始训练集和初始测试集;同时从训练集划分中以随机采样方式置换出2.5万张单行车牌图

像分配至测试集,以维持训练集与测试集数目不变,整个流程如图4所示。

最终,在原CCPD数据集基础上构建了包含单双行车牌的新基础训练集与测试集,总量保持200k不变。原100k的CCPD基础训练集保留60k张单行车牌,黄、绿双行合成车牌各10k;验证集包含15k单行和5k双行合成车牌;基础测试子集中,单行车牌占75k,双行车牌占25k,共100k图像。

上述做法在最大限度保留CCPD基础测试集有效性的同时,将训练集的影响降至最低。这种混合数据集既保留了CCPD独有的各类不同场景测试子集,同时扩充了双行车牌数据,且不增大数据规模。整个过程中,待评测模型不参与数据集构建,全部数据重分配操作均采用独立同分布随机采样方式,因此不会引起数据泄漏问题。

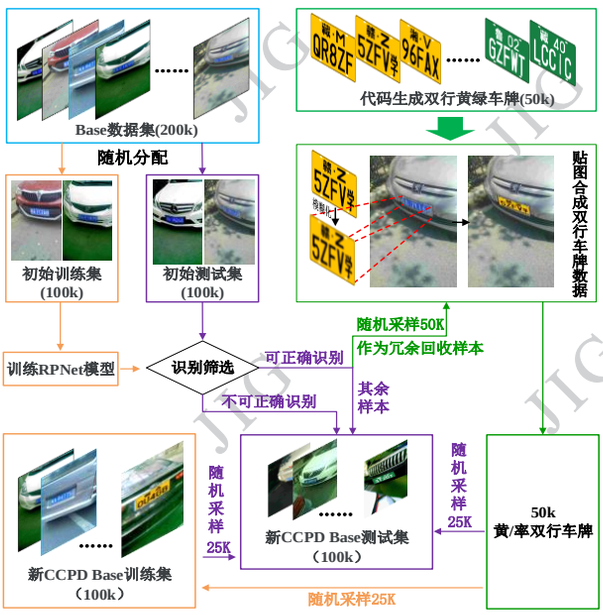


图4　混合CCPD Base数据集构建流程图

Figure 4　Flowchart for the construction of the hybrid CCPD-Base dataset

3　实验设置与结果分析

3.1　实验设置细节

训练机软硬件环境为:Ubuntu18.04,PyTorch 1.8,CUDA 11.7,cuDNN 8.5,GTX TITAN X (GPU)。测试机软硬件环境为:Ubuntu22.04,PyTorch2.3,CUDA 11.8,cuDNN 8.9,GTX 1080 Ti (GPU)。

训练数据为原CCPD(Xu等,2018)基础数据集

和混合CCPD基础数据集。混合CCPD基础数据集分两种，一种同时包含单双行中文车牌；另一种同时包含单行中文车牌和合成的纽约州车牌，均按8∶2的比例划分为训练集和验证集。测试集由原CCPD测试集和自建双行车牌测试子集及纽约州车牌测试子集构成。为测试模型跨数据集的识别性能，本文使用CLPD（Zhang等，2021）和PKUData（Yuan等，2017）作为跨域测试数据集。

训练所采用的数据增强包括：随机仿射变换、随机颜色抖动、高斯噪声扰动、随机区域遮挡。损失函数为交叉熵，优化器为Adam，初始学习率为1e-4。采用基于验证集准确率的自适应衰减策略：连续10轮性能未提升时，学习率衰减为原来的0.5倍，最小学习率约束为1e-7。总训练轮数为200，批大小为32。

本文输入图像尺寸为48 × 112。编码器输出视觉词元维度L = 84、通道维度d = 128；MobileViT模块词元维度为144；并行解码器为2层结构，注意力头数为4、头维为32，查询词元数为8，适配7/8位车牌识别。当6位/7位车牌同时存在时，查询词元数可设置为7，减少计算量。

3.2 算法评估指标

训练与测试所用评估指标主要包括车牌识别准确率R_{LP}和字符识别准确率R_{char}。R_{LP}定义为所有测试车牌图像中被正确识别的车牌数量N_C占测试总车牌量N_{LP}的比例$R_{LP} = N_C/N_{LP}$，单张车牌中所有字符均被正确识别则判定该车牌识别正确；R_{char}定义为模型预测正确的所有车牌字符数M_C占测试集中车牌字符总数M_{Char}的比例$R_{char} = M_C/M_{Char}$。

3.3 识别算法准确性分析

3.3.1 CCPD混合数据集上算法结果对比

为评估PL-TransLPRNet的性能，本文使用多种代表性开源算法开展对比测评。其中，LPRNet（Zherzdev和Gruzdev，2018）与Eulpr（Qin和Liu，2020）采用CNN与CTC架构；TrOCR（Li等，2023）和PARSeq（Bautista和Atienza，2022）均采用ViT与Transformer混合结构；PaddleOCRv3（Li等，2022）采用CNN、Transformer与CTC混合结构；ABINet（Fang等，2021）采用CNN与Transformer混合结构；lpr-transformer（sosopop，2024）为MobileNet与Transformer混合架构；SVTRv1-T（Du等，2022）与SVTRv2T（Du等，2025）均采用Transformer与CTC混合结构。

所有开源算法均在统一软硬件环境下训练与测试：LPRNet采用PyTorch改进版（sirius-ai，2019）；Eulpr保持原生代码单双行分离训练模式；针对TrOCR，本文分别在有无预训练权重的条件下开展训练与评测；PP-OCRv3在混合数据集上重新训练以适配任务场景。

实验在3.2节构建的单双行混合CCPD数据集上进行，该数据集同时包含七位、八位字符的单双行车牌。

针对CCPD中车牌尺度跨度大、各算法输入规格不同的问题，本文设计了统一预处理流程。首先依据标签从原图中提取所有车牌区域，并将其归一化为36 × 94，然后再用双线性插值上采样至224 × 224存储，以减少下采样产生的锯齿效应。评测时按需缩放至各算法所需尺寸，保证在一致且低失真细节下公平对比。

所有算法推理速度评测均在同一机器上以单张图像（*batch size* = 1）的纯推理耗时为衡量指标。先通过50次推理预热GPU，然后再重复1000次推理统计平均耗时换算FPS，并重复6次求平均。

表1对比了混合CCPD数据集上的识别准确率。本文算法在天气、挑战子集上排名次优，其余子集均为最优。带预训练权重的TrOCR通过大规模文本预训练获得了强上下文推理能力，对遮挡鲁棒，因而在天气与挑战子集表现突出；但当在CCPD上从零训练时性能显著下降，挑战子集准确率仅为75.58%，为对比方法中最低。

纯CNN类方法识别准确率普遍偏低，Eulpr综合表现优于LPRNet，但在远近子集上精度不及后者。车牌图像统一缩放后，远距离样本细节少、高频特征不足，近距离样本则细节密集、存在冗余信息。Eulpr使用Inception多尺度结构，在网络浅层就引入大尺度卷积融合全局特征，容易把远距离样本中稀少的笔画高频信息做平均化处理，导致细节特征丢失；而LPRNet以小卷积为主，侧重提取局部细粒度特征，更适配稀疏模糊的远距离样本，但由于缺少全局上下文建模，整体特征学习能力不足。

在双行车牌测试集上，Transformer类模型识别准确率大多数都超过99%，本文模型最优，准确为99.99%；PARSeq为次优，其视觉编码器为12层ViT，无特别的下采样操作，因此对双行车牌适配性

好。SVTRv1-T同样取得了较高准确率。Eulpr在双行车牌上的识别率仅为93.52%。这是因为Eulpr采用区域分割方式分行识别，抗形变能力不足导致。

Transformer依托自注意力完成全局特征建模，可适应车牌字符形变并建立字符位置与上下文关联，优势显著。但SVTRv2-T例外，准确率只有61.63%，这是因为其面向强语义的文本场景设计，车牌字符间语义相关性弱，其网络中的语义引导模块无法发挥正面作用。另外，ABINet的准确率也低于其他Transformer方法，这是因为它的视觉模型只提供辅助，语言模型通过迭代提升精度，但车牌字符语义性弱，因此影响了准确率。

表1　开源车牌识别算法在混合CCPD数据集下的车牌识别率与效率指标对比

Table 1　Comparison of accuracy and efficiency for open-source ALPR algorithms on the hybrid CCPD dataset

算法	各测试子集下车牌识别准确率R_{LP}(%)									效率指标(M)	
	均值	基础-单	基础-双	明暗	远近	旋转	倾斜	挑战	天气	FPS	参数量
LPRNet(Zherzdev和Gruzdev,2018)	93.15	95.85	-	97.1	97.28	90.65	94.43	80.48	96.28	118	**0.65**
Eulpr(Qin和Liu,2020)	94.73	99.29	93.52	97.35	94.67	95.96	97.23	83.24	96.55	63	1.01
PaddleOCRv3(Li等,2022)	97.99	99.65	99.43	98.71	99.24	98.82	98.83	91.06	98.16	58	8.6
TrOCR(Li等,2023)(无预训练)	98.82	99.64	99.43	99.24	99.32	99.37	99.45	**94.68**	**99.43**	12	34.6
	92.11	97.25	95.63	93.18	95.18	93.09	93.09	75.58	93.87		
lpr-transformer(sosopop,2024)	96.64	99.30	99.34	97.86	98.40	97.17	98.08	86.02	96.93	53	1.34
PARSeq(AR)(Bautista等,2022)(NAR,iteration=3)	97.85	99.48	99.88	98.59	98.99	98.33	98.67	89.65	97.86	42	23.8
	97.66	99.48	99.88	98.57	98.97	98.31	98.65	89.57	97.81	72	
ABINet(Fang等,2021)	95.78	98.95	98.94	97.98	98.16	94.14	95.77	85.50	96.78	61	36.8
SVTRv1-T(Du等,2022)	97.05	99.56	99.76	98.63	99.02	97.14	97.50	86.83	97.92	98	4.2
SVTRv2-T(Du等,2025)	89.75	98.57	61.63	98.18	96.33	91.80	94.19	80.90	96.38	117	5.1
本文算法	**99.06**	**99.93**	**99.99**	**99.72**	**99.79**	**99.57**	**99.71**	94.41	99.39	**208.6**	1.44

注：黑色字体标识的为最优结果，下划线标识为次优结果。

在推理速度方面，本文模型的优势显著，推理帧率可达208.6 FPS，这主要得益于模型所采用的并行解码机制。

3.3.2　CCPD数据集上算法识别结果参照对比

为与现有未开源车牌识别算法进行对比，本文引用CCPD论文方法(基线)RPNet(Xu等,2018)及其他方法在CCPD数据集上的实验结果(表2)做对比。可看出，各算法在不同测试子集上的识别性能分布存在差异。根据CCPD(Xu等,2018)的划分方式，基础子集主要为清晰无畸变样本；旋转与倾斜子集包含明显几何形变；远近与明暗子集分别受拍摄距离和光照变化影响；挑战子集则同时叠加多种复杂干扰。整体看，多数模型在基础子集上取得最高精度，在挑战子集上表现最低，与难度分布一致。

不同子集间性能差异还受输入尺寸、识别网络结构以及训练数据规模的影响，其中旋转与倾斜子集由于几何畸变明显，在未引入专门矫正处理时识别率通常低于基础、明暗及远近子集。Wang等人(2022)及Fan和Zhao(2026)由于引入基于车牌顶点的几何校正，在旋转与倾斜子集(图像较清晰)上提升更为明显；而在明暗与远近子集(本身存在模糊退化)中，校正带来的提升相对有限。

Mu和Xu(2023)以及Gao等人(2024)采用相同识别算法，但所用检测算法不同，因此识别率有差异，Mu和Xu(2023)在旋转、倾斜及挑战子集上取得较高准确率。Gao等人(2024)使用相同识别算法，当与所用定位算法联合优化时，检测模块会忽略质量差且不易识别的车牌，实现检测与识别精度的均衡，识别率高。但当检测与识别训练分开后，检测模块不受识别约束，识别准确率下降。

表2中绝大多数算法都采用前置车牌检测算法对测试子集中车牌区域进行定位与提取，不同车牌

检测算法的召回率不同，会导致识别模块所要判读的各个子集中数据分布差异大，如检测算法过滤检测失败的困难样本，使得参与评测的测试集偏向简单样本，最终结果无法真实反映模型在复杂场景下的准确性能。

为实现客观对比，本文采用四种车牌区域提取方式。第一种是利用原始定位标签进行一定的外向扰动后（模拟定位误差），提取各个子集中的车牌图像区域进行识别。第二种是在第一种的基础上增加透视变换矫正。第三、四种使用YOLOv5s（Ultralytics，2020）和SSD（Liu等，2016）定位车牌并提取各测试子集中车牌区域。

表2　CCPD数据集下各车牌算法准确率与推理效率对比

Table 2　Accuracy and inference efficiency comparison of various ALPR algorithms on the CCPD dataset

算法	均值	各测试子集车牌识别准确率R_{LP}(%)							定位算法	识别帧率@硬件
		基础	明暗	远近	旋转	倾斜	挑战	天气		
RPnet(Xu等人，2018)	92.29	98.5	96.9	94.3	90.8	92.5	85.1	87.9	RPnet	61 FPS@Quadro P4000
Zou等人（2020）	95.89	99.3	98.5	98.6	92.5	96.4	86.6	99.3	YOLOv3	–
Wang等人（2022）	99.00	99.9	99.7	99.4	**99.9**	99.9	94.8	99.4	自建*	158.7 FPS@GTX 1080Ti
Chen等人（2022）	97.61	99.7	99.1	98.8	98.4	98.5	90.0	98.8	SSD	37 FPS@GTX TITAN Xp
Mu和Xu（2023）	98.97	99.89	99.57	99.56	99.68	99.8	94.92	99.38	–	204 FPS@GTX 1660s
Gao等人（2024）	98.67	99.9	99.4	99.4	99.6	99.6	93.8	99.0	自建	–
	97.39	99.7	98.8	98.8	98.1	98.3	89.5	98.5	仅识别	–
**Li等人（2024）	98.90	99.68	98.42	98.54	99.1	98.85	**98.34**	99.39	标签	10 FPS@RTX 3060
Tao等人2024	98.74	99.9	99.5	99.5	99.5	99.3	94.1	99.4	YOLOv5s	159.8 FPS@RTX 2080Ti
Fan和Zhao（2026）	97.67	99.7	99.1	98.4	99.2	99.1	90.2	98.0	自建*	72 FPS@GTX 3090
本文算法	99.09	**99.95**	**99.77**	99.83	99.64	99.78	95.24	99.43	SSD	208.6 FPS@GTX 1080Ti
	99.14	**99.95**	99.74	99.81	99.65	99.82	95.52	99.47	YOLOv5s	
	99.24	99.94	99.74	**99.84**	99.89	**99.91**	95.75	**99.58**	标签扰动*	
	99.05	99.93	**99.77**	99.82	99.67	99.74	94.97	99.44	标签扰动	

注：黑色字体标识的为最优结果，下划线标识为次优结果。*表示采用了车牌矫正；**表示采用了预训练权重与FXP8推理权重量化；文献（Mu和Xu等人，2023）仅强调使用了定位算法，未明确定位算法，故以‘–’标识；文献（Tao等人，2024）中的FPS为*batch_size* = 5时的值。

对比同样采用基于标签提取车牌区域的算法Li等人（2024）的结果可看出，本文算法在除挑战子集外的其他子集上，识别准确率均优于Li等人（2024）的算法。虽然Li等人（2024）算法在难度最高的挑战子集上取得了所有对比算法中的最优结果，但其在其余子集上的性能均低于多数对比算法，整体性能分布不符合CCPD数据集的常规难度规律。这种反常现象的成因，与该算法文中所用推理量化技术所带来的正则化效应直接有关。Li等人（2024）为加速算法而采用FXP8精度存储权重，而本文采用PyTorch默认的FP32精度，因此结果差异明显。

和同样使用SSD、YOLOv5s以及采用车牌矫正的算法（Chen等，2022；Tao等，2024；Wang等，2022；Fan和Zhao等，2022）相比，本文算法准确率都具有明显优势。对比表2中本文算法使用标签扰动和定位算法两种方式提取车牌图像后的识别结果，还可看出在挑战子集上存在明显差别，这主要是因为采用YOLOv5，SSD检测车牌时，有些车牌被漏检，而这些漏检的车牌中也包含部分识别较为困难的车牌。当挑战子集中部分难样本被过滤后，其上的识别率就会偏高。

另外，采用YOLOv5定位时，在明暗和远近子集上的识别率分别比标签扰动情况低0.02%和0.01%。这是因为标签扰动下训练与测试均用标签提取，方式一致；而采用YOLOv5提取车牌区域时，识别模型所用训练集仍采用标签方式提取车牌区

域，测试集改用YOLOv5检测，数据提取方式不一致，增大了识别难度，虽过滤了部分难检样本，但提升效果被来源不一致抵消。这一点从SSD定位方式下的识别结果也可看出，除挑战子集识别率高于标签扰动方式外，其他子集上两种方式下识别率差别不大。原因是SSD定位的准确率和召回率均更低，过滤样本更多，识别率未明显变化。

3.3.3 CCPD数据集上算法推理速度参照对比

从表3中对比算法的硬件平台关键参数与推理FPS值可看出，基于Transformer自回归解码的算法如MP-LPR（Li等，2024）、PDLPR（Tao等，2024）推理速度普遍偏低；相比之下，纯CNN结构的网络普遍具有更高的推理效率，其中Mu和Xu（2023）在GTX 1660s平台上达到了204 FPS的推理速度。

本文编码器采用轻量级CNN和MobileViT模块的混合结构，结合Transformer并行解码方案，在强化多尺度特征提取能力的同时，规避了自回归解码带来的时序累积开销。模型总参数1.44 M，总计算复杂度为0.32 GMacs（详见3.6节），在GTX1080Ti

表3 混合CCPD数据集（含纽约州车牌）下各算法的车牌识别准确率R_{LP}（%）

Table 3 R_{LP}（%）of different algorithms on the hybrid CCPD dataset（with New-York-State license plates）

算法	均值	基础-单	基础-纽	明暗	远近	旋转	倾斜	挑战	天气
LPRnet（Zherzdev和Gruzdev，2018）	90.82	98.43	83.2	94.77	96.48	89.53	92.2	76.68	95.26
Eulpr（Qin和Liu，2020）	70.42	82.48	88.93	79.06	76.7	42.6	55.56	56.1	81.89
PaddleOCRv3（Li等，2022）	98.30	99.69	99.67	98.97	99.29	99.17	99.26	91.99	98.39
TrOCR（Li等，2023）	98.81	99.79	99.82	99.36	99.58	99.41	99.49	**93.86**	99.18
lpr-transformer（sosopop，2024）	96.74	99.34	99.39	97.69	98.34	97.27	98.1	86.42	97.4
PARSeq（AR）（Bautista等，2022）（NAR，iteration=3）	97.30	99.45	99.08	98.41	98.77	97.39	98.15	89.36	97.80
	97.29	99.45	99.08	98.41	98.77	97.39	98.16	89.28	97.78
ABINet（Fang等，2021）	94.02	98.17	97.93	96.11	97.02	93.05	94.89	80.31	94.70
SVTRv1-T（Du等，2022）	96.79	99.53	99.49	98.46	98.87	96.72	97.18	86.30	97.98
SVTRv2-T（Du等，2025）	96.36	99.35	98.73	98.12	98.44	95.63	96.92	85.90	97.81
本文算法	**98.92**	**99.90**	**99.86**	**99.62**	**99.77**	**99.51**	**99.59**	93.71	**99.41**

注：黑色字体标识的为最优结果，下划线标识为次优结果。

平台、*batch size* = 1的推理条件下，本文算法帧率可达208.6 FPS。

对比文献（Mu和Xu，2023）中的模型，其参数量为1.87M，计算复杂度为1.71 GMacs，推理帧率为204 FPS。尽管本文采用的CNN与Transformer混合模型其理论计算量显著低于该纯CNN算法，但二者推理帧率相近。这主要是因为纯CNN的卷积算子已在GPU上获得深度优化，硬件执行效率更高；而Transformer结构包含较多访存密集型算子，单步计算效率相对偏低。

受限于GPU型号、深度学习框架以及测试细节的差异，上述对比仅可作为定性参考，不宜视为严格的公平性能评比。更准确的推理速度对比需要在完全统一的硬件环境下复现后得出。

3.4 纽约州混合车牌数据集上的性能评测结果

为进一步验证所提模型对其他类型车牌的适应性，我们采用更具个性和布局多样性的美国纽约州车牌为实验对象，将混合训练集里的国内双行车牌替换为合成的美国纽约州车牌（如图5所示）。

由表3可看出，与表2类似，在混合数据集上，本文方法与TrOCR分别取得了最优与次优的识别准确率，TrOCR（使用预训练权重）在挑战集为最优。LPRNet虽面向单行车牌识别设计，但CNN结构对复杂背景与不规则字符间距的鲁棒性较弱，

在训练集引入具有装饰性图案、州标等干扰因素的纽约州车牌后，其特征建模稳定性下降。Eulpr在LPRNet基础上引入多分支Inception与多层上下文融合，使其对字符排布与间距更为敏感，在混合数

据集下易产生特征冲突，导致在CCPD子集上的识别性能明显下降，而在表2实验中，因其采用单双行车牌分开训练方式，因此该问题未明显体现。对比表2和表3中可看出，SVTRv2-T更适配单行，SVTRv1-T变化不大。

各算法在7个单行测试子集上准确率均有波动，通过计算可知，平均绝对误差最小的是本文算法（0.15%），其次是算法lpr-transformer（0.19%），再次是TrOCR和SVTRv2-T，均为0.24%。

图5　生成的美国纽约州车牌示例图

Figure 5　Examples of generated New-York-State (USA) license-plate images

3.5　模型跨数据集识别性能评测

为评测PL-TransLPRNet的跨数据集识别性能，本文将CCPD数据集下训练的本文模型与前述评测中性能次优的TrOCR用于PKUdata（Laroca等人，2018）和CLPD（Zhang等人，2021）数据集测试。由于PKUdata与CLPD中包含一些黄、绿、黑、白及特种车牌，颜色分布与CCPD差异显著，并且含有少量CCPD中未有的汉字字符。现有方法在评测时，常采用忽略汉字仅统计其他字符准确率的方式。这种做法存在一定局限性：既无法准确反映车牌整体识别精度，又造成评价上的双重标准。为此，本文采用蓝牌与其他车牌分开评测的方式，同时将少量未在CCPD数据集中出现的汉字字符屏蔽，确保逻辑自洽、结果客观，识别结果如表4所示。

表4　PKUdata与CLPD上的跨数据集识别准确率

Table 4　Cross-dataset recognition accuracy on PKUdata and CLPD datasets

测试集(车牌数量)		本文算法		TrOCR	
		R_{LP}(%)	R_{char}(%)	R_{LP}(%)	R_{char}(%)
PKUdata	G1 (743)	99.87	99.98	99.73	99.96
	G2 (643)	100	100	99.22	99.87
	G3 (641)	99.84	99.98	99.69	99.96
	G4 (247)	99.60	99.94	99.19	99.77
	G5 (1398)	98.36	99.64	98.00	99.60
	非蓝牌 (621)	0.16	35.39	11.27	63.45
CLPD	蓝牌 (1142)	88.88	98.01	72.85	95.66
	非蓝牌 (58)	1.72	26.12	6.9	52.67

在G1-G5、CLPD蓝牌子集上，本文算法准确率都明显优于TrOCR。但在PKUdata和CLPD非蓝牌子集上，TrOCR有明显优势，TrOCR能够识别一定比例的分布外的车牌图像，是因为它采用了预训练权重，有一定的先验知识，但准确率绝对值也比较低。因为这些车牌的布局、字体颜色及背景色与CCPD差异显著，属于分布外样本，超出模型任务定义与训练覆盖范围，目前本文算法还无法应对。

3.6　算法消融实验与分析

3.6.1　局部-全局交互模块消融实验

为对比内外残差和位置编码的作用，我们选择3个不同随机种子（表5所示），实现不同的参数随机初始化。相比而言，基于CCPD的车牌识别实验结果显示，内外残差均使用时并无明显优势，使用单残差配合2D可学习位置编码性能更好，参数量适中。只使用外残差，准确率居中，参数量最少。因此本文采用外残差加2D位置编码的方式。

3.6.2　骨干网络消融实验

表6对比了本文采用的骨干网通道升维方式与MobileNet-v1（Howard等，2017）采用的方式1以及MobileViT-v3（Wadekar等，2023）采用的方式2。IRB模块源于MobileNet-v2（Sandler等，2018），本文仅借用MobileNet-v1的升维序列作参照（MobileNet-v2方式不适用）。同时本文还分别使用方式4、5和6评测骨干网特征感受野，特征通道数目对网络识别性能的影响。

表5　局部–全局交互模块有效性验证

Table 5　Ablation study on the effectiveness of the local-global interaction module

使用局部–全局交互模块			√	√	√
外残差+内残差				√	
仅用外残差			√		√
2D可学习位置编码			√	√	
编码器参数量(M)			0.515	0.55	0.501
编码器计算复杂度(GMacs)			0.277	0.288	0.277
平均R_{LP}(%)	随机种子	200	99.02	98.83	98.88
		120	99.06	98.78	98.91
		68	99.05	98.80	98.86

从表6结果可明显看出，基础骨干网输出特征的感受野越小，平均车牌识别准确率越高；方式6方式3接近，方式3的通道数多于方式6，并带有残差，弥补了感受野大带来的影响。方式5感受野最大，没有任何残差辅助，导致字符细节特征无法有效传至后层，因此准确率最低。采用方式4，感受野最小，准确率最高，但计算复杂度最高。方式6与方式7只是通道数不同，可以看出使用64通道的方式7准确率更高。综合速度与准确率考虑，本文选择了方式7。

3.7　查询词元交叉注意力可视化分析

多类型车牌字符长度与布局各异，识别模型要能在超出实际长度的位置输出代表无字符的特殊符号。例如，当7位与8位车牌共存时，第8个查询词元，需依据图像类型特征关注特定区域并输出占位符。为评估本文车牌识别模型在此类场景下的性能，本文提取解码器最后一层查询词元的交叉注意力值，插值至输入图像尺寸并转换为热力图，叠加于原图之上(图6)。热力图亮度表示查询词元与对应区域图像词元的注意力强度，亮度越高，归一化注意力值越接近1；亮度越低，则越接近0。由图6可见，各查询词元预测字符时，其注意力权重所覆盖的图像词元区域与输入图像中对应字符的位置高度一致。这说明模型能隐式对齐查询词元与图像空间位置，具备动态调整注意力的能力，从而有效应对多类型车牌的字符长度与布局差异。当7位与8位车牌共存时，对于7位蓝牌，第8个查询词元需判断当前车牌不存在第8个字符，因此会根据训练所学的查询模式，自适应选择相关图像词元进行判断。如图6第1-4列所示，第8个查询词元通过综合车牌倒数第2个字符附近及上下

表6　不同骨干网络升维方式及下采样位置的实验对比

Table 6　Experimental comparison of different backbone networks with respect to /channel-expansion strategies and down-sampling positions

骨干网升维方式	升维序列(↓下采样)	编码器参数量	编码器计算复杂度	感受野	平均R_{LP}(%)
方式1	3↓→16→32↓→64→128	0.459	0.175	19	98.80
方式2	3↓→32→64↓→128→128	0.595	0.243	19	98.71
方式3	3↓→64→64↓→64→128	0.515	0.240	19	98.96
方式4	3↓→64→64→128	0.479	0.885	11	99.09
方式5	3↓→64↓→64↓→64→128	0.515	0.076	31	98.39
方式6	3↓→32→32→32↓→128	0.431	0.182	15	98.95
方式7(本文)	3↓→64→64→64↓→128	0.515	0.277	15	99.05

注：参数量的单位为M；计算复杂度的单位为GMacs；感受野数值为正方形感受野区域边长(像素)。

图6　不同类型车牌情况下解码器查询词元交叉注意力可视化结果图例

Figure 6　Visualization results of cross-attention for decoder query tokens under different types of license plates

区域的图像词元特征，给出当前位置无字符的判断。而当车牌为8位时（图6第5列），对应查询词元会聚焦于字符所在区域的图像词元，实现正确输出。当6位与7位车牌共存时（图6第6-8列），情况也类似，查询词元会根据车牌类型的不同去关注具有相同模式的位置来给出最后一位字符的判断。

3.8　识别错误情况分析

为分析PL-TransLPRNet在CCPD各测试集上的识别错误情况，本文绘制了如图7所示的7位车牌字符错误分布热力图。如图可见，第0位（汉字位）未出现误识别为数字或字母的情况，第1至6位也未出现误识别为汉字的情况，说明模型通过自注意力和交叉注意力掌握了车牌字符的分布规律。

图7左下方数字区域的错误情况显示，字符“0”和“1”在第2至6位错误次数高于其他数字，原因是其笔画与L、D、B、Q等字母相似，在模糊、遮挡等复杂条件下易混淆。“4”错误最少，原因是受民间选号习俗及交管号牌投放规则影响，CCPD中带“4”的车牌样本偏少。数字“8”则相反，一般不会在位置2出现，更多出现在位置6。

从字母区域的错误情况可看出，位置2错误次数最多，这是因为国内小型汽车号牌序号区为扩容号段，普遍将英文字母优先配置在序号首位（即全局第3个字符，对应图中位置2），导致该位置字母占比最高。

对于汉字区域，由于数据集中‘皖’字占比高，而‘陕’和‘豫’字样本量相对其他异地车牌数量略高，当车牌较为模糊或受光照、拍摄角度等因素影响时，这三种字符更易混淆。另外，个别子集还存在标签错误，误将其他省份车牌标注为安徽车牌。

4　结束语

本文通过设计专用轻量视觉编码器与并行解码器，有效适配各类版式与字符长度可变的车牌识别任务，从架构层面改善现有面向开放场景的车牌识别方法速度慢、通用性弱、准确率低的问题。同时依托冗余回收策略缓解多类型车牌训练数据不足的现状。实验结果表明，所提模型在识别精度与推理速度上均优于同期主流算法。

模型存在的局限性主要是对分布外样本的泛化识别能力不足，缺乏针对遮挡、污损、反光等因素导致的字符歧义伪像的推理校验能力，缺少对特征的质疑、比较与合理性验证机制。

未来工作将考虑从两个层面开展来应对上述局限。针对跨域泛化，探索基于多LoRA专家路由的快速适配方法，以可插拔的轻量模块实现对不同车牌图像域的快速适配；针对歧义伪像的校验缺失问题，引入独立的质量评测与结果验证模块。同时探索模型在端侧设备上的实际部署与应用。

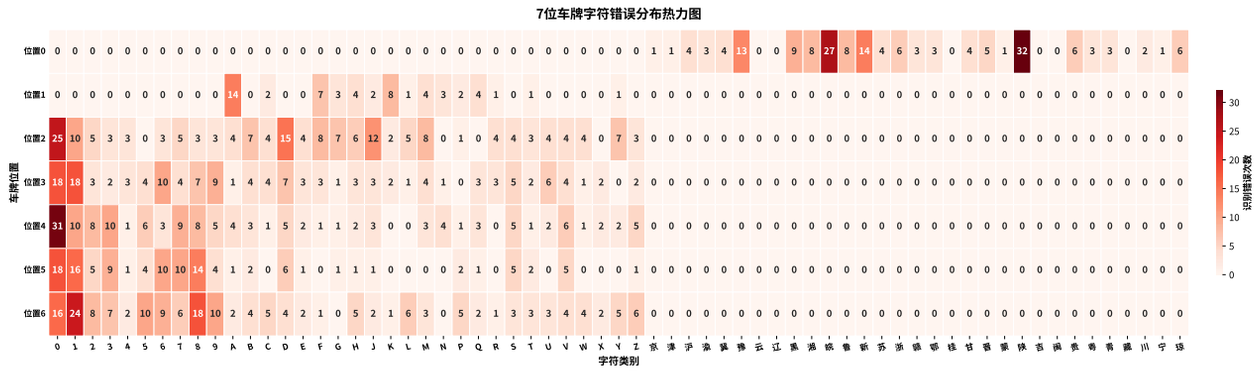


图7　七位车牌字符识别错误分布热力图

Figure 7　Heatmap of recognition-error distribution for seven-character license plates

参考文献(References)

Bautista D and Atienza R. 2022. Scene text recognition with permuted autoregressive sequence models//Lecture Notes in Computer Science. Cham: Springer Nature Switzerland, 13688: 178 - 196. [DOI:10.1007/978-3-031-19815-1_11]

Chen S L, Liu Q, Chen F and Yin X C. 2023. End-to-end multi-line license plate recognition with cascaded perception//Document Analysis and Recognition - ICDAR 2023. Cham: Springer Nature Switzerland:274-289 [DOI: 10.1007/978-3-031-41734-4_17]

Du Y K, Chen Z N, Jia C Y, Yin X T, Zheng T L and Li C X, et al. 2022. SVTR: Scene text recognition with a single visual model//Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22. Vienna: International Joint Conferences on Artificial Intelligence Organization: 884-890 [DOI: 10.24963/ijcai.2022/124]

Du Y K, Chen Z N, Xie H T, Jia C Y and Jiang Y G. 2025. SVTRv2: CTC beats encoder-decoder models in scene text recognition//Proceedings of the IEEE/CVF International Conference on Computer Vision. Honolulu, HI, USA: IEEE: 20147 - 20156. [DOI: 10.1109/ICCV51701.2025.01874]

Duan B, Fu X, Jiang Y and Zeng J X. 2020. Lightweight blurred car plate recognition method combined with generated images. Journal of Image and Graphics, 25(9): 1813-1824 (段宾, 符祥, 江毅, 曾接贤. 2020. 结合GAN的轻量级模糊车牌识别算法. 中国图象图形学报, 25(9): 1813-1824) [DOI:0.11834/jig.190604]

Fan X D and Zhao W. 2026. A robust multi-oriented license plate detector and a derived end-to-end license plate recognizer. IET Image Processing, 20(1): e70272. [DOI:10.1049/ipr2.70272]

Fang S C, Xie H T, Wang Y X, Mao Z D and Zhang Y D. 2021. Read like humans: Autonomous, bidirectional and iterative language modeling for scene text recognition//2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Nashville: IEEE:7094-7103 [DOI: 10.1109/CVPR46437.2021.00702]

Gao Y, Mu S and Xu S. 2024. Toward Unified End-to-End License Plate Detection and Recognition for Variable Resolution Requirements. IEEE Transactions on Intelligent Transportation Systems, 25(9): 10689-10701 [DOI:10.1109/TITS.2024.3366314]

Howard A G, Zhu M, Chen B, Kalenichenko D, Wang W J, Weyand T, Andreetto M and Adam H. 2017. MobileNets: Efficient convolutional neural networks for mobile vision applications [EB/OL]. [2026-05-14]. https://arxiv.org/pdf/1704.04861.pdf

Hua L R, Ma X Y, Zhao C H, Zhang B L, Su Z J and Wu Y H. 2024. Recognition of vehicle license plates in highway scenes with deep fusion network and connectionist temporal classification. IET Image Processing, 18(13): 4066 - 4080. [DOI:10.1049/ipr2.13234]

Ismail A, Mehri M, Sahbani A, Ben Khalifa A and Hamam H. 2025. A dual-stage system for real-time license plate detection and recognition on mobile security robots. Robotica, 43(6): 1981-2002. [DOI: 10.1017/S0263574724001991]

Li C X, Liu W W, Guo R Y, Yin X T, Jiang K T and Du Y K, et al. 2022. PP-OCRv3: More attempts for the improvement of ultra lightweight OCR system[EB/OL].[2026-05-14]. https://arxiv.org/pdf/2206.03001.pdf

Li H, Wang P, You M Y and Shen C H. 2018. Reading car license plates using deep neural networks. Image and Vision Computing, 72: 14 - 23. [DOI:10.1016/j.imavis.2018.02.002]

Li J, Yan D J, He F Z, Dong Z C and Jiang M F. 2024. A mixed-precision transformer accelerator with vector tiling systolic array for license plate recognition in unconstrained scenarios. IEEE Transactions on Intelligent Transportation Systems, 25(12): 20280-20294 [DOI: 10.1109/TITS.2024.3457815]

Li M H, Lv T C, Chen J Y, Cui L, Lu Y J and Florencio D, et al. 2023. TrOCR: Transformer-based optical character recognition with pre-trained models//Proceedings of the AAAI Conference on Artificial Intelligence. Washington, DC: Association for the Advancement of Artificial Intelligence: 13094-13102 [DOI: 10.1609/aaai.v37i11.26538]

Liu W, Anguelov D, Erhan D, Szegedy C, Reed S, Fu C Yet al. 2016. SSD: Single shot multibox detector// European Conference on Computer Vision. Cham: Springer: 21 - 37. [DOI:10.1007/978-3-319-46448-0_2]

Mehta S and Rastegari M. 2022. Separable self-attention for mobile vision transformers[EB/OL].[2026-05-14]. https://arxiv.org/pdf/2206.02680.pdf

Mu S Y and Xu S G. 2023. Robust license plate recognition for full categories based on single character attention. Acta Automatica Sinica, 49(1): 122-134 (穆世义, 徐树公. 2023. 基于单字符注意力的全品类鲁棒车牌识别. 自动化学报, 49(1): 122-134) [DOI: 10.16383/j.aas.c211210]

Qin S X and Liu S J. 2020. Efficient and unified license plate recognition via lightweight deep neural network. IET Image Processing, 14(16): 4102 - 4109. [DOI:10.1049/iet-ipr.2020.1130]

Sandler M, Howard A, Zhu M, Zhmoginov A and Chen L C. 2018. MobileNetV2: Inverted residuals and linear bottlenecks//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 4510-4520. [DOI: 10.1109/CVPR.2018.00474]

Shi B G, Bai X and Yao C. 2017. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 39(11): 2298-2304 [DOI: 10.1109/TPAMI.2016.2646371]

Shpir M, Shvai N and Nakib A. 2025. License plate images generation with diffusion models[EB/OL].[2026-05-14]. https://arxiv.org/pdf/2501.03374.pdf

sirius-ai. 2019. LPRNet_Pytorch[EB/OL]. [2026-05-14] https://github.com/sirius-ai/LPRNet_Pytorch

sosopop. 2024. Chinese-lpr-transformer [EB/OL]. [2026-05-14] https://github.com/sosopop/chinese-lpr-transformer

Tan Y L, Kong A W K and Kim J J. 2022. Pure transformer with integrated experts for scene text recognition// European Conference on Computer Vision. Cham: Springer: 481 - 497. [DOI:10.1007/978-3-031-19815-1_28]

Tao L B, Hong S H, Lin Y X, Chen Y B, He P A and Tie Z X. 2024. A real-time license plate detection and recognition model in unconstrained scenarios. Sensors, 24 (9): 2791. [DOI: 10.3390/s24092791]

Ultralytics. 2020. YOLOv5[EB/OL]. [2026-05-14]. https://github.com/ultralytics/yolov5

Wadekar S N and Chaurasia A. 2022. MobileViTv3: Mobile-friendly vision transformer with simple and effective fusion of local, global and input features[EB/OL].[2026-05-14]. https://arxiv.org/pdf/2209.15159.pdf

Wang Q, Lu X C, Zhang C, Yuan Y and Li X L. 2023. LSV-LP: Large-scale video-based license plate detection and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 45(1): 752-767 [DOI: 10.1109/TPAMI.2022.3153691]

Wang T W, Zhu Y Z, Jin L W, Luo C J, Chen X X and Wu Y Q, et al. 2020. Decoupled attention network for text recognition//Proceedings of the AAAI Conference on Artificial Intelligence. Washington, DC: Association for the Advancement of Artificial Intelligence: 12216-12224 [DOI: 10.1609/aaai.v34i07.6903]

Wang Y, Bian Z P, Zhou Y H and Chau L P. 2022. Rethinking and designing a high-performing automatic license plate recognition approach. IEEE Transactions on Intelligent Transportation Systems, 23(7): 8868 - 8880. [DOI:10.1109/TITS.2021.3087158]

Xu L X, Shang Z H, Chen X J, Zeng T W, Dai W H and Zhao J M. 2026. Deep learning algorithms for license plate recognition: A review. Neural Networks, 200: 108842 [DOI: 10.1016/j.neunet.2026.108842]

Xu Z B, Yang W, Meng A J, Lu N X, Huang H and Ying C C, et al. 2018. Towards end-to-end license plate detection and recognition: A large dataset and baseline//Computer Vision- ECCV 2018. Cham: Springer International Publishing: 261-277 [DOI: 10.1007/978-3-030-01261-8_16]

Yuan Y, Zou W B, Zhao Y, Wang X N, Hu X F and Komodakis N. 2017. A robust and efficient approach to license plate detection. IEEE Transactions on Image Processing, 26(3): 1102-1114 [DOI: 10.1109/TIP.2016.2631901]

Yu D, Li X, Zhang C Q, Liu T, Han J Y and Liu J T. 2020. Towards accurate scene text recognition with semantic reasoning networks// 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle: IEEE: 12110-12119 [DOI: 10.1109/CVPR42600.2020.01213]

Zhang L J, Wang P, Li H, Li Z, Shen C H and Zhang Y N. 2021. A robust attentional framework for license plate recognition in the wild. IEEE Transactions on Intelligent Transportation Systems, 22(11): 6967-6976 [DOI: 10.1109/TITS.2020.3000072]

Zherzdev S and Gruzdev A. 2018. LPRNet: License plate recognition via deep neural networks [EB/OL]. [2026-05-14]. https://arxiv.org/pdf/1806.10447.pdf

Zou Y J, Zhang Y J, Yan J, Jiang X X, Huang T J and Fan H S. 2020. A robust license plate recognition model based on Bi-LSTM. IEEE Access, 8: 211630-211641 [DOI: 10.1109/ACCESS.2020.3040238]