

多尺度区域注意力 InfoGAN 车牌识别网络

徐胜军^{1,2}, 杜淼^{1,2}, 段中兴^{1,2}, 李明海^{1,2}, 韩九强^{2,3}

(1. 西安建筑科技大学信息与控制工程学院, 710055, 西安; 2. 西安市建筑制造智能化技术重点实验室, 710055, 西安; 3. 西安交通大学电子与信息学部, 710049, 西安)

摘要: 针对车牌图像倾斜、遮挡、失真、模糊导致车牌图像难以识别的问题, 提出一种多尺度区域注意力 InfoGAN 车牌识别网络。基于 InfoGAN 框架提出一种多尺度区域注意力车牌超分辨率模块, 通过引入颜色分布、字符结构特征的互信息约束, 提升网络不同特征维度上的判别能力, 重构失真、模糊的低分辨率车牌图像中的关键字符特征; 在无车牌字符位置标签信息的情况下, 使用多尺度区域注意力机制对车牌全局特征图中的车牌字符和背景解耦合, 渐进式地对不同尺度特征图进行不同区域加权, 提高网络对字符区域显著性的关注能力和背景噪声区域的抗干扰能力; 提出多尺度语义车牌字符特征提取模块, 对重构后车牌图像中字符特征解码并识别。在自建 XAUAT-Parking 数据集和公开 CCPD 数据集上进行车牌识别准确率实验, 实验结果表明: 所提网络在 CCPD 公开车牌数据集上的平均识别准确率为 99.3%, 在自建 XAUAT-Parking 数据集上的平均识别准确率为 99.2%。所提网络在复杂场景下具有准确的车牌识别效果和较强的鲁棒性。

关键词: 车牌识别; 注意力机制; 超分辨率; 生成对抗网络

中图分类号: TP391 **文献标志码:** A

DOI: 10.7652/xjtuxb202308020 **文章编号:** 0253-987X(2023)08-0206-03

Multi-Scale Regional Attention InfoGAN License Plate Recognition Network

XU Shengjun^{1,2}, DU Miao^{1,2}, DUAN Zhongxing^{1,2}, LI Minghai^{1,2}, HAN Jiuqiang^{2,3}

(1. School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China;

2. Xi'an Key Laboratory of Building Manufacturing Intelligent & Automation Technology, Xi'an 710055, China;

3. Faculty of Electronic and Information Engineering, Xi'an Jiaotong University, Xi'an 710049, China)

Abstract: A multi-scale regional attention InfoGAN license plate recognition network is proposed for difficult recognition of license plate images that are skewed, obscured, distorted, or blurred. First, a multi-scale regional attention license plate super-resolution module is proposed based on the InfoGAN framework. Mutual information constraints on color distribution and character structure features are introduced to improve the network's discriminant performance in different feature dimensions. Key character features in distorted and blurred low-resolution license plate images are reconstructed. Second, in the absence of the location label information of license plate characters, the license plate characters and background in the global feature map are decoupled by a multi-scale region attention mechanism to progressively weight different regions of the feature

map at different scales, improving the network's performance to focus on the saliency of character regions and resist the interference of background noise regions. Finally, a multi-scale semantic license plate character extraction module is proposed to decode and recognize the character features in the reconstructed license plate image. Both the self-built XAUAT-Parking dataset and the publicly available CCPD dataset are used for license plate recognition accuracy experiments. The experimental results show that the proposed network has an average recognition rate of 99.3% for the CCPD dataset and of 99.2% for the self-built dataset. The research results show that the network features accurate license plate recognition and great robustness in complex scenes.

Keywords: license plate recognition; attention mechanism; super-resolution; generative adversarial network

车牌识别技术是智能交通系统中的核心关键技术之一,其主要过程是利用模式识别、数字图像处理、计算机视觉等相关技术,从采集的车牌图像中提取车牌信息并实现车辆身份的认证。然而,实际场景中诸多不可抗力因素影响着车牌识别精度,主要包括:①浓雾、雾霾、雨、雪等强干扰信息导致车牌图像质量低、完整度差;②摄像机拍摄角度以及车辆高速行驶导致采样的车牌图像扭曲、倾斜、运动模糊。这些问题为车牌准确识别带来了巨大的挑战。

车牌识别方法主要分为基于字符分割^[1-5]和基于非字符分割^[6-11]识别两类。基于字符分割识别方法首先将定位后的车牌图像分割为单字符图像,然后通过文本识别网络对单字符图像进行识别。Gou等^[1]利用极端区域法检测车牌字符并细化车牌位置,提出一种混合判别受限波尔兹曼机非线性车牌字符分类器。Hou等^[2]提出一种基于笔画宽度变换的分段车牌字符处理方法,在分割前无需精定位整个车牌位置情况下,解决了车牌识别中旋转因素的影响。Hsu等^[3]将最大稳定极值区域应用于字符分割,使用线性判别分析分类器提取局部二元模式特征并进行字符分类识别。Navastara等^[4]提出基于视频数据 CCTV 的车牌识别系统,利用循环神经网络识别车牌字符。Tu等^[5]提出一个基于分层区域卷积神经网络的多车辆检测和车牌识别系统,利用 RCNN 网络对分割后的单个车牌字符进行识别,使得网络关注到单个车牌字符特征,增加了车牌字符的可识别性。然而,由于自然场景中如强光、阴暗、恶劣天气、运动模糊等环境干扰,难以准确提取分割车牌字符关键信息,因此常导致基于字符分割的方法效果不佳。

近年来,基于非字符分割的车牌识别受到广泛关注。Li等^[6]提出一种基于长短时记忆的递归神经网络车牌识别算法,有效识别了 CNN 网络提取

的序列车牌字符。Lee等^[7]提出一种基于软注意力建模的递归循环网络,通过隐性学习字符级语言模型隐形提取字符特征,提高了自然场景下车牌中字符信息的关注能力。Zherzdev等^[8]提出一种基于卷积神经网络端到端的轻量化车牌识别网络 LPR-Net,解决了多品类车牌中可变长字符的问题。Qin等^[9]针对无约束场景,采用无锚点方法检测车牌边界和 4 个角,将字符特征视为序列标注问题,提高了网络在倾斜车牌图像识别方面的鲁棒性。穆世义等^[10]提出一种基于单字符注意力的全品类鲁棒车牌识别网络,基于注意力机制对车牌全局特征图进行单字符级特征分割,有效解决了多品类车牌和倾斜车牌的二维字符布局问题。Xu等^[11]提出一种基于位置感知的二维注意力识别网络,有效识别了具有透视变形的单线和双线车牌。

然而,对于这类严重扭曲、倾斜、运动模糊的车牌图像,上述直接采用端到端方式对车牌字符进行识别仍是一个挑战。基于超分辨率车牌图像重构再识别的方法为解决这类识别问题提供了一个有效的解决方案。Dong等^[12]提出了基于生成对抗网络(generative adversarial network, GAN)的超分辨率卷积神经网络,利用对抗学习策略和感知损失,通过恢复高级特征生成更逼真的超分辨率图像。Liu等^[13]提出了一种渐进式车牌搜索的 DP-GAN 框架,生成多种空间对应关系和高分辨率的图像,使得网络关注于像素级的细节恢复。Lee等^[14]提出一种基于字符感知损失的车牌图像超分辨率模型 CS-RGAN,解决了重构时车牌关键字符信息完整度的问题。Zhang等^[15]基于端到端的框架,提出了一种融合车牌超分辨率和识别的多任务生成对抗网络的车牌识别系统。Hamdi等^[16]提出一种双 GAN 联合对抗训练的 D_GAN_ESR 网络,提高了网络对高度失真和运动模糊条件下的车牌字符可识别性。

Min 等^[17]提出一种基于 Fast-SRGAN 的模型,通过去除车牌图像噪声提高整体车牌识别准确率。Huang 等^[18]提出一种基于生成对抗网络的超分辨率 DDA-SRGAN 框架,使用双维注意力机制学习关键区域容易丢失的特征细节。徐胜军等^[19]提出一种编解码结构的车牌图像超分辨率网络 SRRN,利用所提 KPRN 对倾斜车牌进行矫正,并自适应重构倾斜低分辨率车牌图像。上述基于 GAN 的车牌超分辨率方法利用车牌图像中像素级相似分布关系生成车牌超分辨率图像,取得了较好的效果,但是这些方法没有更多关注车牌图像中的关键车牌字符特征,容易在实际场景中导致车牌识别准确率低下。

综上所述可知,基于深度学习的车牌识别网络已经取得了较好的效果。但是,在实际场景中采样得到的原始车牌图像仍存在模糊、污损、扭曲、倾斜、过曝等问题,这些强干扰因素导致这些方法无法准确提取车牌关键字符特征,难以解耦低分辨率车牌图像中关键字符特征与车牌背景噪声。虽然基于传

统 GAN 的车牌识别方法通过学习车牌图像间像素级数据相似分布关系来关注关键字符特征,但这些方法忽略了生成车牌中字符特征的可识别性,背景噪声的干扰。针对上述问题,本文提出一种多尺度区域注意力 InfoGAN 车牌识别网络 (multi-scale regional attention InfoGAN license plate recognition network, MRI-LPR),所提网络利用多尺度区域注意力对车牌图像进行关键字符特征与背景噪声解耦,实现了无噪声和有精细纹理的高清车牌图像的超分辨率重构,建立了车牌字符多尺度语义特征提取模块,使得网络更多关注车牌字符关键信息,有效提高了车牌识别效果。

1 MRI-LPR 车牌图像识别网络

1.1 网络结构

针对复杂场景下模糊、污损、扭曲、倾斜、过曝等造成车牌关键字符特征难以提取的问题,提出一种基于多尺度区域注意力的 InfoGAN 车牌识别网络,其整体框架如图 1 所示。

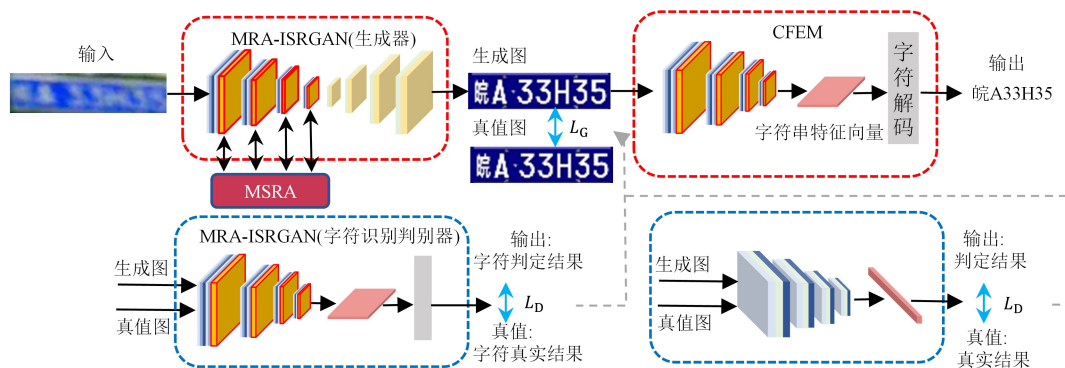


图 1 多尺度区域注意力车牌图像识别网络

Fig. 1 Multi-scale regional attention license plate image recognition network

所提网络主要包含 3 个部分:多尺度区域注意力 InfoGAN 车牌超分辨率模块 (multi-scale regional attention InfoGAN license plate super-resolution module, MRA-ISRGAN), 车牌字符多尺度语义特征提取模块 (license plate character multi-scale semantic feature extraction module, CFEM), 多尺度区域注意力模块 (multi-scale regional attention, MSRA)。MRA-ISRGAN 模块由生成器与两个判别器网络组成,其中生成器网络采用编解码结构,由多层卷积模块与上采样解码模块组成,通过解码输出高分辨率车牌图像。车牌字符判别器以分辨真假的方式衡量车牌生成图与真值图分布之间的非相似

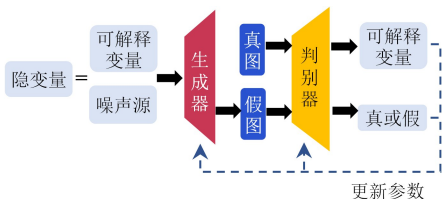
性。车牌字符识别判别器采用字符语义特征识别方式衡量车牌生成图与真值图之间的车牌字符可识别性。最终,将车牌字符判别器和识别判别器的损失融合,并采用反向传播方式优化车牌生成器网络。CFEM 模块主要由车牌超分辨模块输出的车牌生成图像作为车牌字符特征提取模块的输入,联合轻量级卷积神经网络组成的车牌字符特征提取模块和车牌超分辨模块进行端到端训练,提取输入不同字符长度车牌图像的关键字符特征,解码推理出字符概率序列,采用车牌字符概率最大准则对车牌字符类别进行预测。MSRA 模块通过对车牌特征图进行不同尺度编码得到不同区域的描述符,然后对相

同区域描述符经过空间映射获得车牌图像的通道注意力权重矩阵。因此,加权后得到的车牌特征图不但可以显著关注字符信息,还能有效区分车牌背景和无关图像背景。

1.1.1 MRA-ISRGAN 模块

GAN 网络采用对抗学习使生成器和判别器之间达到纳什均衡,从而使生成器生成以假乱真的假图。为解决 GAN 的输入噪声随机、无约束、生成图像缺乏可解释性的问题,InfoGAN^[20] 等被相继提出。InfoGAN 框架示意如图 2 所示。InfoGAN 引入隐变量编码,建立与生成数据的互信息约束,利用互信息约束对生成图像进行控制,其关键思想是将隐变量分解为一组可解释变量和一个不可解释噪声源,通过在原始 GAN 目标函数中添加一个额外的判定项,捕获隐变量中可解释变量和生成器输出之间的互信息,进而增加可解释性。

受此启发可知,原始的低分辨率车牌图像与生成车牌图像存在高度特征相关性,比如颜色分布、字符结构特征等,二者之间具有较强的互信息。首先,将无类别标签的低分辨率车牌图像作为生成器的输入,与真值标签图建立互信息约束,挖掘低分辨率车



解码阶段可表示为

$$F_g = f_j(F_m, \theta_j) \quad (2)$$

式中： F_g 表示生成器网络的输出； f_j 表示转置卷积、LeakyReLU、tanh 激活函数的组合； θ_j 是可学习参数。

基于语义监督的多维度车牌判别器网络用于优化车牌生成器网络。该网络由车牌字符判别器和字符识别判别器组成，车牌字符判别器利用输出标量判定概率结果分辨车牌图像真假，衡量车牌生成图与真值图分布之间的非相似性，在数据分布维度上逼近真值图分布，但是车牌字符判别器缺少识别数据特征的能力；车牌字符识别判别器网络对生成车牌图像中关键字符语义特征进行识别，衡量生成图中各个关键字符信息的可识别性。联合车牌字符判别器和字符识别判别器网络进行端到端训练时，使生成器网络可以更好地同时利用数据分布特征和字符语义特征，因而可以提高车牌图像的识别效果。

车牌判别器网络的输入分别为生成车牌图像 F_g 和真值图 F_t ，经过车牌判别器特征提取与解码阶段最终输出判别结果。车牌字符判别器的输出结果为输入车牌图像是真值图的概率，字符识别判别器的输出结果为生成车牌图像中关键字符预测值。车牌判别器网络结构如图 4 所示。图中，红色箭头代表两倍下采样最大池化操作。车牌判别器网络可表示为

$$D_1 = \mu[f_p(F_g, F_t)] + g(F_g) \quad (3)$$

式中： f_p 表示卷积、批归一化、LeakyReLU 激活函数的组合函数； μ 表示判别器中特征图的解码； g 表示字符识别分支的过程； D_1 表示为车牌字符判别器输出结果。字符识别判别器具体网络结构见图 5，车牌字符识别判别器网络可表示为

$$D_2 = d(F_g) \quad (4)$$

式中： $d(\cdot)$ 表示字符识别判别器网络； D_2 表示字符识别判别器网络输出关键字符信息预测结果。

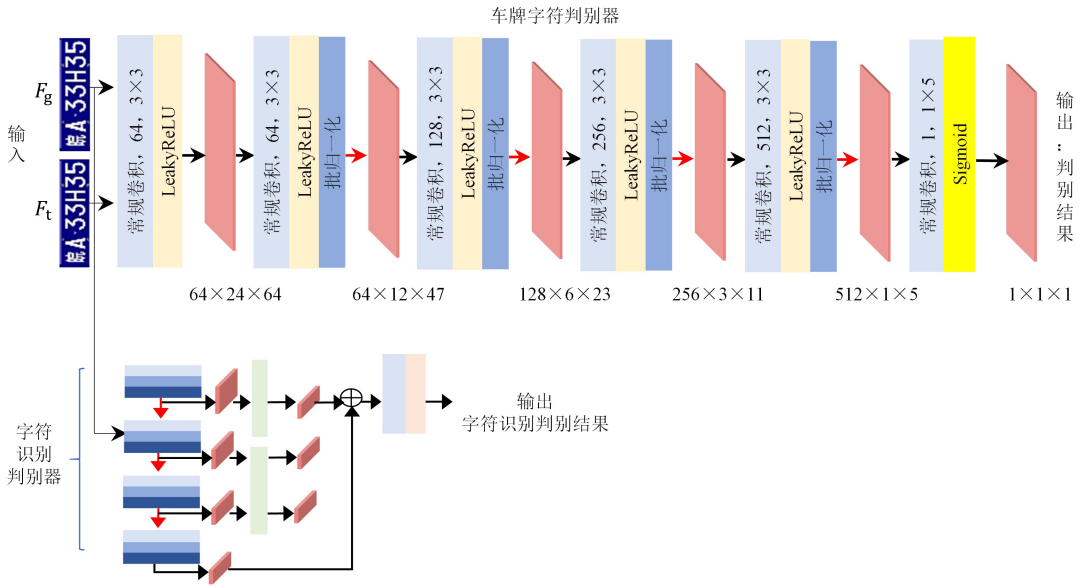


图 4 基于语义监督的多维度车牌判别器

Fig. 4 Semantic supervision-based multi-dimensional license plate discriminator

1.1.2 CFEM 模块

CFEM 模块的作用是提取车牌生成图像的不同长度字符信息，然后解码推理出字符概率序列，在序列每个位置取类概率的最大值作为最优字符预测结果。基于 VGG 网络架构，车牌字符多尺度语义特征提取模块采用 45 层卷积结构作为特征提取网络。基础单元由常规核大小为 3×3 的卷积、批归一化、ReLU 激活函数组成，网络降采样使用 3D 最大

池化。随着网络深度的增加，在不同深度网络层提取 4 种尺度的特征图，并利用不同核大小平均池化将不同尺度的特征图统一调整为 4×18 像素的特征图，然后将特征图在通道维度串联拼接融合，最后使用核大小为 1×1 的卷积将特征图的深度调整为字符类编号，并对每个位置取最大置信度作为最终预测结果。网络细节设置如图 5 所示，其中红色箭头代表 3D 最大池化降采样操作。

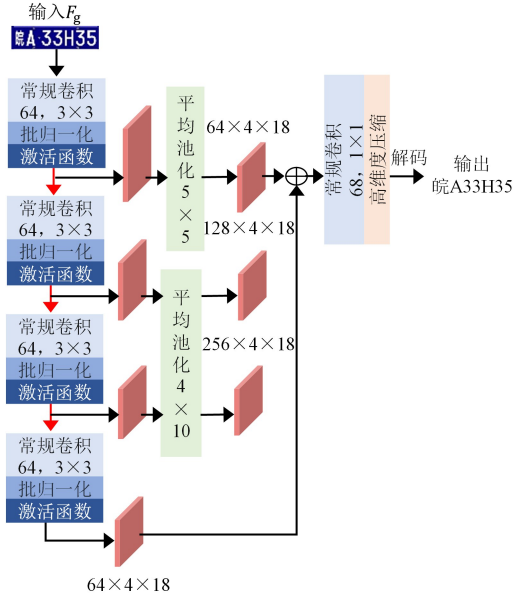


图 5 多尺度语义车牌字符特征提取模块

Fig. 5 Multi-scale semantic license plate character feature extraction module

超分辨率模块的输出生成图 F_g 作为字符特征提取模块的输入。具体地,字符特征模块利用 4 次基础单元与 3D 最大池化串联拼接组合得到 4 种尺度的特征图,表示为

$$S_n = g_n(F_{n-1}, \theta_n), \quad n = 1, 2, 3, 4 \quad (5)$$

式中: S_n 是第 n 个特征图; g_n 表示第 n 个尺度特征提取过程; θ_n 是可学习参数; F_{n-1} 是第 $n-1$ 个尺度特征图。然后,使用平均池化将 4 个尺度特征图统一调整为最小尺度 4×18 像素的特征图,在通道维度拼接得到融合多尺度特征信息的特征图 F_k

$$F_k = S_1 \oplus S_2 \oplus S_3 \oplus S_4 \quad (6)$$

式中 $\oplus$ 表示在通道维度拼接。

如图 5 所示,全局特征图包含尺度 $H \times W$ 分别为 22×92 、 20×90 、 18×44 、 4×18 视野域的特征信息,融合多尺度信息的特征图 F_k 经过解码将通道数调整为与分类字符数相等,最终得到车牌图像识别的预测结果,具体过程可表示为

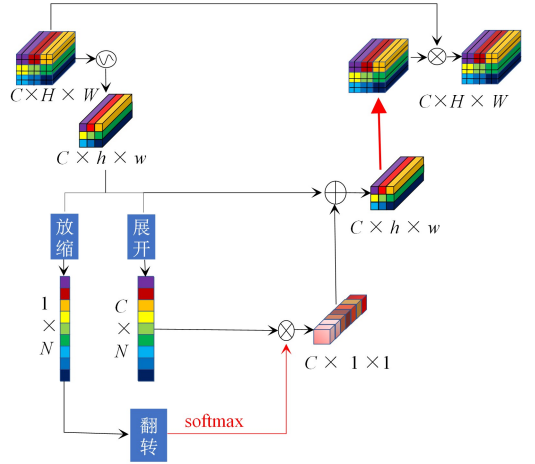
$$F_c = \delta[\text{Conv}(F_k)] \quad (7)$$

式中: $\text{Conv}(\cdot)$ 表示大小和步长都为 1 的卷积运算; $\delta(\cdot)$ 表示车牌预测过程; F_c 表示最终车牌识别结果。

1.1.3 MSRA 模块

基于常规 GAN 的超分辨率网络通过输入噪声与真值图之间的数据分布关系映射高清车牌生成

图,由于这种网络难以通过控制噪声的维度生成可解释的车牌语义特征,因而缺乏区分关键字符信息与背景噪声的能力。首先,由于车牌内字符信息具有横向分布特性,因此在不同尺度全局特征图中 MSRA 可以使用条状区域编码对车牌字符区域进行解耦合,使得网络更多关注到车牌局部关键字符信息;其次,根据编码形成的区域描述符相似性映射为每个通道的区域加权;然后,再根据每个通道区域加权后的全局特征聚合为基于区域的通道注意力,因此这种引入了上下文语义信息的 MSRA 注意力机制可以有效抑制区域间相似度低的通道,消除车牌背景噪声,增强网络提取有效局部字符区域关键信息的能力。多尺度区域注意力模块结构如图 6 所示。



⊗ 多尺度区域编码 → 自适应上采样 → 信息流

图 6 多尺度区域注意力模块

Fig. 6 Multi-scale regional attention module

多尺度区域注意力模块具体实现过程如下:首先,令 $\mathbf{X} \in \mathbf{R}^{C \times H \times W}$ 为输入的特征图,多尺度区域编码使用 $(H, 2H/3)$ 大小的卷积核和 $(1, 2)$ 步长的卷积对特征图进行编码,表示为

$$\mathbf{X}_1 = \tau(\mathbf{X}) \quad (8)$$

式中: $\tau(\cdot)$ 表示多尺度区域编码操作; $\mathbf{X}_1 \in \mathbf{R}^{C \times h \times w}$ 为多尺度区域编码特征图。

其次,对编码后的特征图 $\mathbf{X}_1$ 进行展开和放缩变换,可以得到不同区域描述符表示

$$\mathbf{X}_2 = \text{reshape}(\mathbf{X}_1) \quad (9)$$

$$\mathbf{X}_3 = \text{reshape}[\text{Conv}(\mathbf{X}_1)] \quad (10)$$

式中: $\mathbf{X}_2 \in \mathbf{R}^{C \times N}$ 表示展开后的区域描述符; $\mathbf{X}_3 \in \mathbf{R}^{1 \times N}$ 表示放缩后的描述符,其中 $N = h \times w$ 表示区域描述符总数。

然后,通过对放缩后区域描述符 $\mathbf{X}_3$ 进行 softmax 操作,得到区域间的相关性注意力矩阵,再将其和区域描述符 $\mathbf{X}_2$ 进行融合获得区域相关性映射,得到通道相关性的特征权重 $\mathbf{X}_4$

$$\mathbf{X}_4 = \gamma[\text{softmax}(\mathbf{X}_3)] \quad (11)$$

式中: $\mathbf{X}_4 \in \mathbf{R}^{C \times 1 \times 1}$ 为特征通道权重; $\gamma(\cdot)$ 表示翻转和对应元素相乘的变换。

最终,将多尺度区域编码后特征图 $\mathbf{X}_1$ 、通道相关性特征权重 $\mathbf{X}_4$ 和特征图 $\mathbf{X}$ 进行有效整合操作,得到多尺度区域注意力加权特征图

$$\mathbf{X}_c = \rho(\mathbf{X}, \mathbf{X}_1, \mathbf{X}_4) \quad (12)$$

式中: $\mathbf{X}_c \in \mathbf{R}^{C \times H \times W}$ 为加权特征图; $\rho(\cdot)$ 为逐元素相加、自适应上采样和矩阵对应元素相乘的组合操作。

1.2 基于对抗学习的联合损失函数

所提多尺度区域注意力 InfoGAN 车牌识别网络由多尺度区域注意力 InfoGAN 车牌超分辨率模块、车牌字符多尺度语义特征提取模块、多尺度区域注意力模块组成。根据不同模块端到端学习的特殊对抗学习关系应用不同的损失函数进行训练。多尺度区域注意力 InfoGAN 车牌超分辨率模块主要学习车牌生成图和真值图二者之间分布映射关系,使得车牌输出 F_g 生成图在数据分布维度逼近真值图 F_t ,对于生成器网络,对抗型损失函数表示为

$$L_G = L_{\text{smooth}} + L_D \quad (13)$$

生成器损失 L_G 包含 L_{smooth} 生成损失和 L_D 判别损失。 L_{smooth} 由生成器的输出生成图 F_g 与真值图 F_t 之间的均方误差和平均绝对误差组合构成

$$L_{\text{smooth}} =$$

$$\begin{cases} \sum_{i=1}^N \sum_{x=1}^W \sum_{y=1}^H 0.5(F_{g,i}(x, y) - F_{t,i}(x, y))^2, \\ \quad |F_{g,i}(x, y) - F_{t,i}(x, y)| < 1 \\ \sum_{i=1}^N \sum_{x=1}^W \sum_{y=1}^H |F_{g,i}(x, y) - F_{t,i}(x, y)| - 0.5, \\ \quad \text{其他} \end{cases} \quad (14)$$

式中: N 表示训练样本总数; i 表示当前训练样本。 L_D 是关于判别器的损失表示

$$L_D = L_{\text{BCE}} + L_{\text{CTC}} \quad (15)$$

$$L_{\text{BCE}} = \sum_{i=1}^N (-Y_{g,i} \log D(F_{g,i}) - Y_{t,i} \log D(F_{t,i})) \quad (16)$$

式中: L_{BCE} 为车牌字符判别器的损失; Y_g 、 Y_t 分别是

生成图和真值图的标签置信值; $D(\cdot)$ 是经判别器解码预测后输出为预测真值图的概率; L_{CTC} 为字符识别判别器的损失,用于提高生成图中字符可识别性,定义为

$$L_{\text{CTC}} = \sum_{i=1}^N f_{\text{ctc}}(P_{g,i}, P_{t,i}) \quad (17)$$

式中: $f_{\text{ctc}}(\cdot)$ 表示 CTC 损失; P_g 、 P_t 分别是生成图预测字符概率序列和字符标签序列。

针对输入车牌图像字符长度可变情况, L_{CTC} 损失可以驱动可变长度序列的解码,防止车牌生成器生成畸形字符。在每次迭代训练过程中,首先将低分辨率含噪声原始车牌图像 F_i 输入车牌生成器中,通过前向传播来计算 L_G ,然后分别将真值为 0 的生成图 F_g 和真值为 1 的 F_t 输入到车牌字符判别器和字符识别判别器来计算 L_D ,二者通过对抗学习相互优化,最终以交替训练方式优化生成器与判别器,使得车牌生成器不仅可以通过 L_G 生成向真值图分布无限逼近的图像,而且利用 L_D 来约束生成器生成难以识别的字符。

2 实验结果与分析

2.1 训练设置

实验平台搭载 Intel Xeon E5 2650 处理器,376 GB 内存,4 个 NVIDIA 2080Ti 12 GB 显卡。深度学习框架采用 pytorch-1.8、NVIDIA 公司 CUDA11.2 的 GPU 运行平台以及 cuDNN8.0 深度学习 GPU 加速库。网络训练过程中,所提网络设置训练参数批大小为 256、epoch 为 131、初始学习率为 0.001,使用 Adam 优化器进行优化。网络训练过程中,学习率调整策略采用每 10 个 epoch 指数衰减一次,衰减系数为 0.1。

2.2 中国城市停车数据集

中国城市停车数据集 (Chinese city parking dataset, CCPD)^[21] 由 Android POS 机所拍摄的车牌照照片组成,涉及多种复杂环境,包括模糊、倾斜、雨天、雪天等。CCPD 数据集是目前最大的车牌数据集,提供超过 29 万张带有详细注释的中文车牌图像。该数据集根据识别难度、车牌区域的光照、拍摄时的车牌距离、水平倾斜和垂直倾斜程度以及天气 (雨、雪或雾) 分为不同的组。每个类别包括 10 000~20 000 张图像。CCPD 数据集由 Base、Blur、Challenge、DB、FN、NP、Rotate、Tilt、Weather 等子数据集组成。各子集特征如表 1 所示。

表 1 CCPD 数据集
Table 1 Detail of CCPD dataset

子数据集	特征	图像数
Base	不同视角、距离	199 996
Blur	较为模糊	20 611
Challenge	最难的基准	50 003
DB	车牌区域较明或暗	10 132
FN	位置较远或较近	20 967
NP	无车牌的新车	3 036
Rotate	有较大水平倾斜	8 685
Tilt	有水平、垂直倾斜	30 216
Weather	较差天气情况拍摄	9 999

2.3 XAUAT-Parking 数据集

XAUAT-Parking 数据集属于自建的混合车牌数据集,共有 40 795 张车牌车辆数据,车牌图像尺寸为 $1\,536\times 960$ 像素。该数据集不仅涵盖了各类实际场景下停车场进出卡口的蓝绿色混合车牌图像数据,而且具有大量较大差异的蓝绿色混合车牌数据,其中蓝色车牌共 30 158 张,新能源绿色车牌 10 637 张。XAUAT-Parking 数据集可以按进出口分为进场子集与出场子集,由于当前新能源车的迅速普及,因此自建数据集更贴近当前实际场景中车牌数据特征,可有效验证实际场景中基于不同深度学习网络的车牌识别算法的鲁棒性。

2.4 实验量化指标及分析

车牌识别准确率指标定义为所有测试样本中预测结果正确的样本数占测试样本总数的比例。其中,车牌所有字符全部预测正确则判定为该车牌识

别正确。车牌识别准确率定义为

$$R_{\text{LR}} = \frac{N_{\text{t}}}{N_{\text{LP}}} \tag{18}$$

式中: R_{LR} 为该样本车牌的识别准确率; N_{t} 为预测结果正确的样本数; N_{LP} 为测试样本总数。

为验证车牌图像超分辨率生成效果,采用峰值信噪(PSNR)、结构相似(SSIM)指标进行量化评估。PSNR 指标定义为

$$I_{\text{PSNR}} = 10\lg\left(\frac{C_{\text{MAX}}}{\sqrt{E_{\text{MS}}}}\right) \tag{19}$$

$$E_{\text{MS}} = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (L(i, j) - P(i, j))^2 \tag{20}$$

式中: C_{MAX} 代表图像颜色的最大值; E_{MS} 表示均方误差; $L(i, j)$ 表示坐标 (i, j) 的标签车牌图像像素强度值; $P(i, j)$ 表示坐标 (i, j) 的重构车牌图像像素强度值; M 、 N 分别代表车牌图像宽和高。

结构相似性指标分别从亮度、对比度、结构这 3 方面度量图像相似性。结构相似性指标定义为

$$I_{\text{SSIM}} = l(P, L)c(P, L)s(P, L) \tag{21}$$

式中: $l(P, L)$ 、 $c(P, L)$ 、 $s(P, L)$ 分别为重构车牌图像 P 和标签车牌图像 L 的亮度相似性、对比度相似性和结构相似性。

2.5 消融实验

为验证所提网络各模块的有效性,在 XAUAT-Parking 数据集上开展消融对比实验。实验中,将 XAUAT-Parking 数据集按字符长度分为 7 位蓝色车牌、8 位绿色车牌和混合车牌共 3 个子集,在 3 个子集中测试不同类别车牌识别准确率。消融实验结果如表 2 所示。

表 2 各模块在车牌识别网络中的定量效果分析

Table 2 Quantitative effect analysis of each module in the license plate recognition network

网络	识别准确率/%		
	7 位蓝色车牌	8 位绿色车牌	混合车牌
CFEM	93.6	89.2	91.0
CFEM+MRA-ISRGAN	98.5	96.1	97.3
CFEM+MRA-ISRGAN+MSRA1	99.0	97.2	98.1
CFEM+MRA-ISRGAN+MSRA1+MSRA2	99.4	97.9	98.6
CFEM+MRA-ISRGAN+MSRA1+⋯+MSRA3	99.6	98.3	99.0
CFEM+MRA-ISRGAN+MSRA1+⋯+MSRA4	99.7	98.6	99.2

注:表中加粗数据为最优值。

表 3 展示了使用 CFEM、MRA-ISRGAN、MSRA 这 3 个模块串联在一起完成车牌识别任务

的计算复杂度、参数量、单个车牌图像识别耗时。表中,计算复杂度用乘加操作数度量。

表 3 各模块计算复杂度和参数量消融实验
Table 3 Computational complexity of each module and parametric number of ablation experiments

网络	计算复杂度/ 10^9	参数量/ 10^6	单个车牌识别耗时/ms
CFEM	0.14	0.44	10
CFEM+MRA-ISRGAN	2.05	20.70	15
CFEM+MRA-ISRGAN+MSRA	2.23	21.20	19

为探究 MSRA 模块在不同网络深度时所关注的车牌区域特征,分别对网络中不同位置的 MSRA 模块进行定量消融实验和定性可视化实验。浅层网络引入 MSRA1 和 MSRA2 的可视化热力图如图 7 所示。

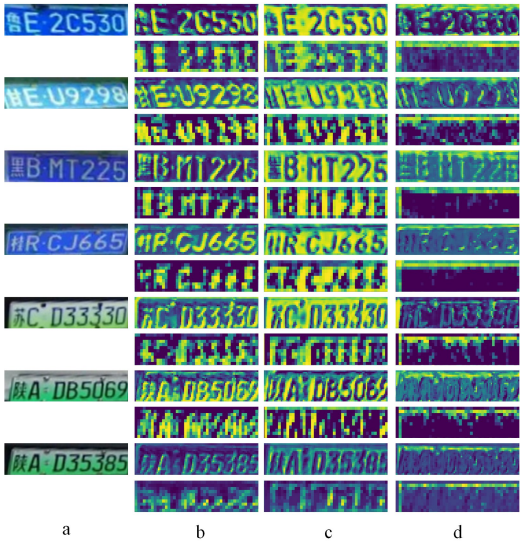


图 7 不同位置 MSRA 模块车牌图像可视化
Fig. 7 Visualization of MSRA modules at different locations in license plate images

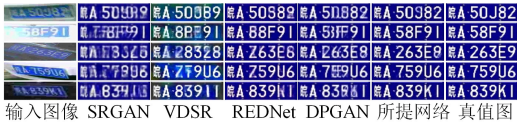
图中:奇数行表示嵌入 MSRA1 的特征图;偶数行表示嵌入 MSRA2 的特征图;a 列为车牌原图;b 列为车牌原图经 MSRA1 和 MSRA2 后的显著关注车牌字符信息特征图;c 列表示车牌原图经 MSRA1 和 MSRA2 后的除车牌字符信息外的车牌区域高亮特征图;d 列表示与车牌无关的背景区域。观察图 7 可知:前两行中的“鲁 E2C530”车牌特征图,从 b 列中可明显关注到与原图纵向对齐位置有高亮的“鲁 E2C530”字符信息,随着网络深度加深,经过 MSRA2

后也仍然可观察到字符纵向对齐区域获得了更多关注。在 c 列中,关键车牌字符外的蓝色、绿色车牌区域变为更受关注区域,尽管网络深层使得图像特征变得模糊,但与浅层特征纵向对应来看,字符外车牌区域仍然处于受关注区域。从 d 列中发现,车牌外背景噪声区域更加高亮,随着网络加深,图像噪声区域亮度变化不大。使用 MSRA 模块后的特征图在不同通道中主要关注到车牌图像的字符区域、除字符外车牌区域和背景噪声区域。可视化结果表明,MSRA 模块不仅能够有效区分这 3 类区域信息,而且能够在车牌实例中学习的关键字符结构信息。综上所述可知,MSRA 注意力模块不但可以区分常规蓝色和新能源车牌的关键字符信息与背景噪声,还能使生成车牌图像保留关键车牌字符信息,进而提高识别准确率。

2.6 CCPD 数据集测试评估

2.6.1 定性分析

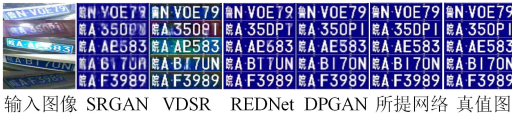
在 CCPD 数据集上,将所提网络分别与 SRGAN^[22]、VDSR^[23]、REDNet^[24]、DPGAN^[13] 网络进行车牌超分辨率重构效果的定性对比实验,部分实验结果如图 8 所示。可以看出,和 SRGAN、VDSR、REDNet、DPGAN 等对比网络相比,所提网络超分辨率模块可生成高质量车牌超分辨率图像。图 8(a)为最具挑战性的车牌图像样本重构,此类车牌图像具有严重模糊、倾斜、过曝、遮挡等现象。可以看出,SRGAN 网络重构图像常出现字符信息缺失,VDSR、REDNet 网络重构图像出现字符特征模糊,DPGAN 网络重构图像出现字符重影现象,这种较差的重构效果常导致车牌的误识别。图 8(b)和图 8(c)主要为倾斜和旋转的车牌图像重构,这种倾斜和旋转常导致车牌字符发生扭曲现象,难以正确提取关键字符特征信息。可以看出,VDSR 网络中重构后车牌字符图像均出现较多噪声。图 8(d)为极端天气影响下字符被遮挡的车牌图像重构,图 8(e)和图 8(f)为过曝、过暗情况下字符特征与其他噪声特征融为一体的车牌图像重构。可以看出:SRGAN 受以上干扰因素最为严重,重构的字符信息识别性不高;VDSR 重构中引入过多噪声信息;REDNet 重构车牌字符纹理细节缺失;DPGAN 在重构原始车牌中模糊的字符特征时,重构出的字符同样模糊。



(a)Challenge 子数据集对比效果



(b)Tile 子数据集对比效果



(c)Rotate 子数据集对比效果



(d)Weather 子数据集对比效果



(e)DB 子数据集对比效果



(f)FN 子数据集对比效果

图 8 CCPD 子集车牌超分辨率结果对比
Fig. 8 CCPD dataset license plate super-resolution results

综上所述可知,所提网络中的车牌生成器网络引入了多尺度区域注意力机制,提升了生成器网络区分车牌关键字符区域特征提取能力,能够有效区分字符信息、除字符外的车牌区域、背景噪声。同时,

所提网络联合基于语义监督的多维度车牌判别器与生成器对抗训练,使得生成车牌图像不仅具有传统 GAN 在像素级分布上相似于标签图像的特性,而且使生成字符信息更具可识别性。因此,所提网络车牌超分辨率模块的重构效果优于其他对比网络。

2.6.2 定量分析

以 CCPD 中的 Challenge 子数据集为例,超分辨率的评价指标对比如表 4 所示,对比指标为峰值信噪比和结构相似性。可以看出:所提网络在 Challenge 子数据的 PSNR 平均值为 43.2 dB, SSIM 平均值为 0.99;相较于 SRGAN、VDSR、REDNet、DPGAN 对比网络,PSNR 分别提高了 26.0、27.6、20.9、27.1 dB,SSIM 分别提高了 1.5、0.8、0.3、0.13 个百分点。

表 4 CCPD Challenge 子数据集超分辨率对比
Table 4 Comparison of super-resolution of CCPD Challenge sub-dataset

网络	I_{PSNR}/dB	I_{SSIM}
SRGAN ^[22]	17.2	0.84
VDSR ^[23]	15.6	0.91
REDNet ^[24]	22.3	0.96
DPGAN ^[13]	16.1	0.86
本文网络	43.2	0.99

注:表中加粗数据为最优值。

不同网络在 CCPD 数据集上的车牌识别准确率如表 5 所示。

表 5 不同网络在 CCPD 数据集上的车牌识别准确率

Table 5 License plate recognition accuracy of different networks in CCPD

网络	识别准确率/%							
	平均值	Base	DB	FN	Rotate	Tilt	Weather	Challenge
文献[25]网络	87.7	97.2	94.4	90.9	82.9	87.3	85.5	76.3
文献[26]网络	91.6	98.3	96.6	95.9	88.4	91.5	87.3	83.8
文献[8]网络	86.9	97.8	92.2	91.9	79.4	85.8	92.0	69.8
文献[27]网络	98.3	99.5	98.1	98.6	98.1	98.6	97.6	86.5
文献[28]网络	95.8	99.3	98.5	98.6	92.5	96.4	99.3	86.6
文献[29]网络	93.6	98.9	96.1	96.4	91.9	93.7	95.4	83.1
文献[9]网络	97.6	99.5	93.3	93.7	98.2	95.9	98.9	92.9
文献[30]网络	97.0	99.8	99.2	99.1	96.4	97.6	98.5	88.9
文献[31]网络	87.4	99.8	97.5	88.3	94.1	83.1	73.2	75.3
文献[19]网络	97.9	99.7	93.8	94.5	98.9	97.8	99.4	90.2
无 MSRA 时的本文网络	98.6	99.7	94.2	95.6	99.0	98.5	99.4	92.6
本文网络	99.3	99.8	94.9	96.3	99.2	98.8	99.5	94.5

注:表中加粗数据为最优值。

从表 5 可以看出,所提车牌识别网络在 CCPD 数据集中有较高的识别准确率,在其子集 Base、DB、FN、Rotate、Tilt、Weather、Challenge 中均超越其他对比算法,识别准确率分别达到了 99.8%、94.9%、96.3%、99.2%、98.8%、99.5%、94.5%,各子集平均准确率达到 99.3%。对比网络中,整体效果最优的为文献[19]网络,其在 Base、DB、FN、Rotate、Tilt、Weather、Challenge 各子集中的识别准确率分别为 99.7%、93.8%、94.5%、98.9%、97.8%、99.4%、90.2%,所提网络较其分别高出 0.1、1.1、1.8、0.3、1.0、0.1、4.3 个百分点。对比网络中,虽然文献[30]网络在 DB、FN 两个子数据集达到 99.2%和 99.1%的识别准确率,但在 Rotate、Tilt、Weather、Challenge 子集中的识别准确率较差,分别为 96.4%、97.6%、98.5 和 88.9%,而所提网络在这 4 个子集中分别比文献[30]网络高出 2.8、1.2、1.0、5.6 个百分点,尤其对于最具挑战性的 Challenge 子集,所提网络识别准确率提高了 5.6 个百分点。综上所述可知,所提网络在各种复杂环境下具有较高识别准确率和鲁棒性。

2.7 XAUAT-Parking 数据集测试评估

为验证所提网络的泛化性能,在自建 XAUAT-Parking 数据集上进行了对比实验。由于 XAUAT-Parking 数据集中蓝色车牌图像与绿色新能源车牌图像的数量比为 3:1,所以按照此比例分别从蓝色车牌和绿色新能源车牌数据集中随机抽取 3/5 进行训练,其余的 2/5 进行测试。对比重构效果如图 9 所示。



图 9 XAUAT-Parking 数据集部分车牌重构对比结果
Fig. 9 Comparison results of partial license plate reconstruction for XAUAT-Parking dataset

两类车牌长度不一致且外观颜色相差较大。在训练样本不充足的情况下,常规深度学习网络难以有效提取绿色车牌特征,从而导致基于常规深度学习网络的车牌重构算法的泛化能力较弱。在蓝、绿混合车牌子集中,字符长度会随着车牌类型不同而改变,当不能准确判别车牌类型和错误识别车牌字符信息时,会导致重构车牌图像失真,并且直接影响

最后的车牌字符预测结果。由图 9 可以发现,SRGAN、VDSR、REDNet、DRGAN 在进行蓝、绿混合车牌子集测试时,重构 8 位新能源车牌效果较差,车牌字符特征失真。当使用所提网络进行重构时,不论是在常规 7 位蓝色车牌图像还是在字符长度改变的 8 位新能源绿色车牌中,都可以有效提取关键字特征,并重构为无噪声且纹理细节特征完整的车牌图像。

定量对比结果如表 6 所示。可以看出,所提网络表现较好的文献[19]网络,识别准确率分别分别提高了 1.4、1.4 和 1.1 个百分点。

表 6 XAUAT-Parking 数据集上的车牌识别准确率对比

Table 6 Comparison of recognition ratio for XAUAT-Parking dataset

网络	识别准确率/%		
	7 位蓝色 车牌	8 位绿色 车牌	蓝绿混合 车牌
文献[8]网络	93.6	89.2	91.0
文献[29] 网络	94.1	90.1	92.1
文献[30] 网络	92.3	89.3	90.8
文献[9] 网络	94.5	91.6	93.1
文献[19] 网络	98.3	97.2	98.1
本文网络	99.7	98.6	99.2

注:表中加粗数据为最优值。

进一步分析可知,对比网络对蓝色车牌的识别准确率显著高于绿色车牌。这是由于数据集采用了不同颜色车牌作为训练数据,而实际应用场景中,绿色车牌数量明显少于蓝色车牌,因此网络训练所采用数据集呈现长尾分布特征,导致对比网络对绿色车牌的识别准确率显著低于蓝色车牌。由于所提网络引入的隐变量具有多样性,建立了更多颜色分布、字符结构特征的互信息约束,提升了网络在颜色、字符等不同特征维度上的判别能力,因此所提网络对不同颜色车牌都均有较高的识别准确率。

3 结 论

针对复杂实际场景中的模糊、污损、扭曲、倾斜车牌图像识别问题,提出了一种多尺度区域注意力的车牌识别网络。所提网络设计了基于多尺度区域注意力 InfoGAN 的车牌超分辨率模块,将低分辨率车牌图像恢复到高分辨率且特征完整的车牌图像。在不需要单字符位置标签监督训练的情况下,利用

多尺度区域注意力的对车牌前景和背景区域进行特征提取,实现了并行化车牌字符特征恢复。基于多尺度语义车牌字符特征提取模块对生成的高分辨率车牌图像进行字符特征提取,准确预测车牌中的可变长字符。在公开数据集 CCPD 和自建数据集 XAUAT-Parking 上,通过定量实验与定性实验验证了所提网络的有效性和鲁棒性。消融研究的定量和定性分析证明了所提 CFEM、MRA-ISRGAN、MSRA 模块的有效性。在未来的工作中,可尝试提高所提模型效率,并将其应用到其他类型车牌的识别中。

参考文献:

- [1] GOU Chao, WANG Kunfeng, YAO Yanjie, et al. Vehicle license plate recognition based on extremal regions and restricted Boltzmann machines [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2016, 17(4): 1096-1107.
- [2] HOU Yongshun, QIN Xunhui, ZHOU Xiangdong, et al. License plate character segmentation based on stroke width transform [C]//2015 8th International Congress on Image and Signal Processing (CISP). Piscataway, NJ, USA: IEEE, 2015: 954-958.
- [3] HSU G S, CHEN J C, CHUNG Y Z. Application-oriented license plate recognition [J]. *IEEE Transactions on Vehicular Technology*, 2013, 62(2): 552-561.
- [4] NAVASTARA D A, MUSYAFIRA N, FATICHAH C, et al. Video-based license plate recognition using single shot detector and recurrent neural network [C]//2021 13th International Conference on Information & Communication Technology and System (ICTS). Piscataway, NJ, USA: IEEE, 2021: 151-154.
- [5] TU Chunling, DU Shengzhi. A hierarchical RCNN for vehicle and vehicle license plate detection and recognition [J]. *International Journal of Electrical and Computer Engineering*, 2022, 12(1): 731-737.
- [6] LI Hui, WANG Peng, YOU Mingyu, et al. Reading car license plates using deep neural networks [J]. *Image and Vision Computing*, 2018, 72: 14-23.
- [7] LEE C Y, OSINDERO S. Recursive recurrent nets with attention modeling for OCR in the wild [C]//2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Piscataway, NJ, USA: IEEE, 2016: 2231-2239.
- [8] ZHERZDEV S, GRUZDEV A. LPRNet: license plate recognition via deep neural networks [EB/OL]. [2022-10-01]. <https://arxiv.org/abs/1806.10447>.
- [9] QIN Shuxin, LIU Sijiang. Towards end-to-end car license plate location and recognition in unconstrained scenarios [J]. *Neural Computing and Applications*, 2022, 34(24): 21551-21566.
- [10] 穆世义, 徐树公. 基于单字符注意力的全品类鲁棒车牌识别 [J]. *自动化学报*, 2023, 49(1): 122-134. MU Shiyi, XU Shugong. Full-category robust license plate recognition based on character attention [J]. *Acta Automatica Sinica*, 2023, 49(1): 122-134.
- [11] XU Hui, GUO Zhaozhong, WANG Dahan, et al. 2D license plate recognition based on automatic perspective rectification [C]//2020 25th International Conference on Pattern Recognition (ICPR). Piscataway, NJ, USA: IEEE, 2021: 202-208.
- [12] DONG Chao, LOY C C, HE Kaiming, et al. Image super-resolution using deep convolutional networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, 38(2): 295-307.
- [13] LIU Wu, LIU Xinchun, MA H, et al. Beyond human-level license plate super-resolution with progressive vehicle search and domain priori GAN [C]//Proceedings of the 25th ACM international conference on Multimedia. New York, NY, USA: Association for Computing Machinery, 2017: 1618-1626.
- [14] LEE Seyun, KIM J H, HEO J P. Super-resolution of license plate images via character-based perceptual loss [C]//2020 IEEE International Conference on Big Data and Smart Computing (BigComp). Piscataway, NJ, USA: IEEE, 2020: 560-563.
- [15] ZHANG Minghui, LIU Wu, MA Huadong. Joint license plate super-resolution and recognition in one multi-task gan framework [C]//2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). Piscataway, NJ, USA: IEEE, 2018: 1443-1447.
- [16] HAMDI A, CHAN Y K, KOO V C. A new image enhancement and super resolution technique for license plate recognition [J]. *Heliyon*, 2021, 7(11): e08341.
- [17] MIN D, LIM H, GWAK J. Improved method of license plate detection and recognition facilitated by fast super-resolution GAN [J]. *Smart Media Journal*, 2020, 9(4): 134-143.
- [18] HUANG Chien, LI Y H, ASLAM M S, et al. Super-resolution generative adversarial network based on the dual dimension attention mechanism for biometric image super-resolution [J]. *Sensors*, 2021, 21(23): 7817.
- [19] 徐胜军, 邓博文, 史亚, 等. 一种编解码结构的车牌图像超分辨率网络 [J]. *西安交通大学学报*, 2022, 56(10): 101-110.

- XU Shengjun, DENG Bowen, SHI Ya, et al. An encoder-decoder-based super resolution network for license plate images [J]. *Journal of Xi'an Jiaotong University*, 2022, 56(10): 101-110.
- [20] CHEN Xi, DUAN Yan, HOUTHOOFT R, et al. InfoGAN: interpretable representation learning by information maximizing generative adversarial nets [C]// *Advances in Neural Information Processing Systems 29 (NIPS 2016)*. Vancouver, Canada: NIPS, 2016: 2180-2188.
- [21] XU Zhenbo, YANG Wei, MENG A, et al. Towards End-to-End license plate detection and recognition: a large dataset and baseline [C]// *ECCV 2018*. Cham, Germany: Springer International Publishing, 2018: 261-277.
- [22] LEDIG C, THEIS L, HUSZÁR F, et al. Photo-realistic single image Super-Resolution using a generative adversarial network [C]// *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ, USA: IEEE, 2017: 105-114.
- [23] KIM J, LEE J K, LEE K M. Accurate image super-resolution using very deep convolutional networks [C]// *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Piscataway, NJ, USA: IEEE, 2016: 1646-1654.
- [24] PENG Xi, FERIS R S, WANG Xiaoyu, et al. RED-net: a recurrent encoder-decoder network for video-based face alignment [J]. *International Journal of Computer Vision*, 2018, 126(10): 1103-1119.
- [25] ZHANG Linjiang, WANG Peng, LI Hui, et al. A robust attentional framework for license plate recognition in the wild [J]. *IEEE Transactions on Intelligent Transportation Systems*, 2021, 22(11): 6967-6976.
- [26] REN Shaoqing, HE Kaiming, GIRSHICK R, et al. Faster R-CNN: towards real-time object detection with region proposal networks [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2017, 39(6): 1137-1149.
- [27] LIU Wei, ANGUELOV D, ERHAN D, et al. SSD: single shot MultiBox detector [C]// *Computer Vision-ECCV 2016*. Cham, Germany: Springer International Publishing, 2016: 21-37.
- [28] LUO Canjie, JIN Lianwen, SUN Zenghui, MORAN: a multi-object rectified attention network for scene text recognition [J]. *Pattern Recognition*, 2019, 90: 109-118.
- [29] ZOU Yongjie, ZHANG Yongjun, YAN Jun, et al. A robust license plate recognition model based on Bi-LSTM [J]. *IEEE Access*, 2020, 8: 211630-211641.
- [30] WANG Tianwei, ZHU Yuanzhi, JIN Lianwen, et al. Decoupled attention network for text recognition [C]// *Proceedings of the AAAI Conference on Artificial Intelligence*. Palo Alto, CA, USA: AAAI Press, 2020: 12216-12224.
- [31] SEO T M, KANG D J. A robust layout-independent license plate detection and recognition model based on attention method [J]. *IEEE Access*, 2022, 10: 57427-57436.

(编辑 陶晴)