

# 基于样本扩充的小样本车牌识别

尤鸣宇, 韩 煊

(同济大学电子与信息工程学院, 上海 201804)

[摘要] 车牌识别技术作为智能交通系统的核心课题之一, 一直受到广泛的关注. 近年来深度学习技术的迅速发展, 更是为其提供了一种良好的解决方案. 但实际场景下, 研究者有时很难收集到足够的数据以支持模型训练. 本文聚焦于小样本车牌识别问题, 提出了使用生成对抗网络生成车牌图像, 辅助后续模型训练的方法. 本文方法先使用 CycleWGAN 合成大量带标签车牌图像; 之后用合成图像对识别模型进行预训练; 最后使用原始真实数据微调模型, 进一步提高模型的准确率. 本文在多个数据集上验证此方法, 均获得了明显的效果增益, 特别是当真实数据相对有限时, 本文方法将准确率从已经较高的基线上又提升了 7.5%. 另外, 在较困难的双动态车牌图像上, 本文方法也取得了不俗的效果. 最后, 引入模型压缩技术, 在原方法的基础上设计并实现了 LightRCNN, 使识别速度提升近 1 倍.

[关键词] 小样本车牌识别, 生成对抗网络, 卷积神经网络, 双向循环神经网络

[中图分类号] TP391 [文献标志码] A [文章编号] 1001-4616(2019)03-0001-10

## Small Sample License Plate Recognition Based on Sample Expansion

You Mingyu, Han Xuan

(College of Electronics and Information Engineering, Tongji University, Shanghai 201804, China)

**Abstract:** As one of the significant topics of intelligent transportation system, license plate recognition technology has been widely studied. In recent years, the rapid development of deep learning has provided a good solution for it. However, in actual situations, it is difficult for researchers to collect enough data to support the training of the model sometimes. This paper focuses on the license plate recognition with small amount of data, and proposes a method to generate training data with adversarial generation network to assist the recognition model training. First, a large-scale image set is generated using the generator of GAN. Then, these images are fed to a deep convolutional neural network followed by a bidirectional recurrent neural network with long short-term memory, which performs the feature learning and sequence labelling. Finally, the pre-trained model is fine-tuned on real images. Our experimental results on a few data sets demonstrate the effectiveness of using GAN images; an improvement of 7.5 recognition accuracy percent points over a strong baseline with moderate-sized real data being available. The proposed framework achieves competitive recognition accuracy on challenging test datasets. This paper also leverages the depthwise separate convolution to construct a lightweight convolutional recurrent neural network, which is about half size and 2×faster on CPU. Combining this framework and the proposed pipeline, this paper makes progress in performing accurate recognition on mobile and embedded devices.

**Key words:** license plate recognition with small dataset, generative adversarial network, convolutional neural network, bidirectional recurrent neural network

车牌识别是智能交通系统(ITS)的核心课题之一, 众所周知, 车辆牌照与机动车一一对应, 识别车牌是鉴别机动车最直接最有效的手段. 交通管理, 电子支付以及刑侦等诸多领域都需要车牌识别技术的支持. 深度学习方法的迅速发展为其提供了一种良好的解决方案, 但在实际情况下, 研究者很难收集到足够的数据支持深度模型训练. 鉴于此, 本文将关注的焦点置于小样本车牌识别问题, 引入生成对抗网络(GAN)拓展训练数据, 辅助后续识别模型训练. 除此之外, 当前许多相关研究只停留在静态车牌识别的阶段, 所用图像大多在车辆静止时拍摄. 而实际中车辆普遍处于移动状态, 所以本文还探讨了运动车牌的识

收稿日期: 2019-07-05.

基金项目: 上海市自然科学基金(18ZR1442600).

通讯联系人: 尤鸣宇, 副教授, 研究方向: 模式识别与智能系统. E-mail: myyou@tongji.edu.cn

别问题,由于数据的缺乏,目前对这一问题的研究并不广泛。

近年来,新的 LPR 方法不断出现,性能指标也逐步提升。Li 等人<sup>[1]</sup>使用级联网络与滑动窗口提取图像特征,实现了无分割车牌识别。受此类研究<sup>[1-2]</sup>的启发,我们将深度卷积网络(DCNN)与循环神经网络(RNN)结合,提出一种无分割的卷积-循环(C-RNN)车牌识别框架,该框架以整张车牌为对象捕获特征。

深度学习方法的成败与否依赖于数据量是否足够。而由于隐私保护等限制,我们很难收集到足量的带标签动态车牌图像。为了应对这一问题,我们设计并实现了一种基于 CycleGAN<sup>[3]</sup>的车牌图像生成算法,后续的实验也证明了使用此方法可以获得高质量的动态车牌图像。

除此之外,为了适应手机等平台的需求,本文对模型进行压缩,提高了前向运算的速度,同时精度损耗也被控制在较小的范围。

本文的主要工作包括以下几个方面:(1)提出了使用 GAN 拓展车牌图像,辅助后续识别模型训练的方法。借助该方法,即使在真实数据相对有限时,依然可以获得足量的训练数据。(2)将 Wasserstein 损失与 CycleGAN 相结合,优化了 CycleGAN 的训练过程。(3)考虑到车牌识别问题的特殊性,将 CNN 与 RNN 相结合,设计并实现了无分割的车牌识别 C-RNN 网络。(4)以实验的方式,探究并分析了 GAN 生成数据对后续模型训练的影响。(5)本文设计并实现了轻量级识别网络 LightC-RNN,参数量仅为原 C-RNN 的一半,前向运算速度较后者提高了近 1 倍。

## 1 相关研究

### 1.1 车牌识别

现有的车牌识别方法可以分为两类:有分割 LPR<sup>[4-6]</sup>与无分割 LPR<sup>[1]</sup>。前者需要先分割车牌中各个字符,再使用分类器逐一识别,用于字符分割的方法主要有基于投影的方法<sup>[5-6]</sup>和基于连通元的方法<sup>[7-8]</sup>,单字符分类则可以用模板匹配<sup>[9-10]</sup>或基于学习的方法<sup>[8,11-12]</sup>解决。就目前的研究而言,诸如支持向量机<sup>[12]</sup>、隐马尔可夫模型<sup>[11]</sup>与神经网络<sup>[8,13]</sup>等基于学习的方法,效果上较模板匹配法有着明显的优势。

反观无分割 LPR,Li 等<sup>[1]</sup>提出了一种基于 DCNN 与 LSTM 的级联车牌识别框架,该方法用滑窗提取图像的特征序列。本文方法也属于无分割 LPR,识别框架设计参考了 Shi 等人的研究<sup>[2]</sup>,该框架摒弃了滑窗的策略,直接用 CNN 从整张车牌图中提取特征,后续的字符序列识别则使用双向长短期记忆网络(LSTM)。

### 1.2 生成对抗网络

生成对抗网络<sup>[14]</sup>的核心是使用生成器和判别器进行对抗式训练。早在 2014 年,Conditional GAN 就证实了通过调节生成器与判别器可以实现由标签到图像的生成。之后,深度卷积对抗网络(DCGANs)<sup>[15]</sup>的提出又为 GAN 的训练提供了稳定的网络框架。时至今日,GAN 已经可以完成诸如文本到图像<sup>[16]</sup>、图像到图像<sup>[17]</sup>等更复杂的生成任务。

本文使用的 CycleGAN 允许我们在没有真假图像对的条件下获得真伪图像空间之间的映射关系。应用的关键在于循环一致性损失,其内在原理可以描述为:如果一个对象被从一个空间映射到了另一个空间,那么也可以反其道而行之,恢复到最初的起点。2017 年发布的 Wasserstein GAN(WGAN)<sup>[18]</sup>对 GAN 的训练过程进行了改进,在提升训练过程稳定性的同时,也避免了模型陷入结果单一化(mode collapse,也称模式坍塌)的陷阱。在上述成果之外,GAN 还被应用于图像填充<sup>[19]</sup>、表示学习<sup>[20]</sup>与 3D 对象生成<sup>[21]</sup>等任务中,均取得了极佳的效果。

### 1.3 训练数据生成

在文本定位<sup>[22]</sup>与文本识别<sup>[23]</sup>的问题中,数据生成的技术早有应用,也取得了很好的效果。除此之外,在字体识别<sup>[24]</sup>、行人检测<sup>[25]</sup>、语义分割<sup>[26]</sup>等领域,数据生成也在发挥着它的价值。但是,许多研究受模型描述能力所限,生成的图像还很难比拟真实的数据,二者在风格、清晰度以及信息的丰富性上存在很大差距。Zheng 等人<sup>[27]</sup>使用 vanilla DCGAN 生成的无标签数据做半监督学习,一定程度上提升了模型的准确性。本文将传统图像处理与 GAN 相结合(也是人工先验与学习算法的结合),生成带标签训练数据。为了证实生成数据的训练效果,本文还设计了用无标签生成数据做半监督学习的对比实验。

## 2 网络结构与训练

这一节将对本文方法做详细的描述. 如图 1 所示, 首先借助人工粗合成图像与原始的真实图像训练 GAN, 获得大量的精合成图像, 然后使用精合成图像对 C-RNN 识别模型进行预训练, 最后用原始真实图像微调模型.

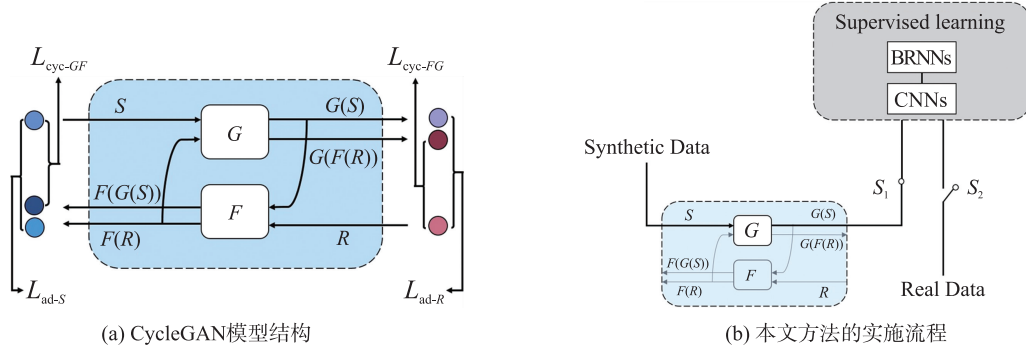


图 1 CycleGAN 模型结构与本文方法的实施流程

Fig. 1 The architecture of the CycleGAN model and the pipeline of proposed approach

### 2.1 生成对抗网络的结构与训练

生成对抗网络一般包含一个生成器和一个判别器, 其训练可视为是一个最优化过程: 生成器与判别器都希望最小化自身的损失同时最大化对方的损失. 理论上, 优化的最终结果是: 生成器生成了接近真实的数据, 判别器也捕获到了更加深层的特征. 本文最终选定的网络框架是 CycleGAN, 并使用 Wasserstein 损失监督训练.

#### 2.1.1 CycleGAN

使用 GAN 的目的是获取从粗合成图像空间  $S$  到真实图像空间  $R$  的映射,  $G$  与  $F$  代表相应的生成器:  $G: S \rightarrow R, F: R \rightarrow S, D_S$  与  $D_R$  则代表判别器. CycleGAN 的核心要求在于确保循环一致性, 即不仅使  $G(S) \approx R$ , 还要使  $F(G(S)) \approx S$ , 这一原则也预防了结果单一化的问题. 本文的生成器包含 2 个双步卷积, 6 个残差块, 以及 2 个反卷积层; 判别器则参照了 PatchGANs<sup>[17]</sup> 的设计, 根据  $70 \times 70$  的图像块判别真伪. 原始的损失函数如式(1):

$$L(G, F, D_S, D_R) = L_{\text{LSGAN}}(G, D_R, S, R) + L_{\text{LSGAN}}(F, D_S, R, S) + \lambda L_{\text{cyc}}(G, F), \quad (1)$$

式中,  $L_{\text{LSGAN}}(G, D_R, S, R)$  和  $L_{\text{LSGAN}}(F, D_S, R, S)$  项代表对抗损失, 由式(2)定义, 由于 CycleGAN 中需要学习双向的映射, 所以需计算两个方向的对抗损失.

$$L_{\text{LSGAN}}(G, D_R, S, R) = E_{r \sim \text{Pdata}(R)} [(D_R(r) - 1)^2] + E_{s \sim \text{Pdata}(S)} [D_R(G(s))^2]. \quad (2)$$

$L_{\text{cyc}}(G, F)$  代表循环一致性损失,

$$L_{\text{cyc}}(G, F) = E_{s \sim \text{Pdata}(S)} [\|F(G(s)) - s\|] + E_{r \sim \text{Pdata}(R)} [\|G(F(r)) - r\|], \quad (3)$$

式(1)中,  $\lambda$  为循环一致性损失的系数.

#### 2.1.2 Wasserstein 损失

不同于原始的 CycleGAN, 本文将 Wasserstein 损失与 CycleGAN 结合, 提出了 CycleWGAN. 在 GAN 的训练过程中, 损失函数有着十分重要的作用, 选择不当会使模型难以收敛并引起结果单一化. Wasserstein 损失很好地应对了这一问题, 两个分布之间的 EM 距离由式(4)定义:

$$EM(P_r, P_s) = \inf_{\lambda \in \Pi(P_r, P_s)} E_{(x, y) \sim \lambda} [\|x - y\|]. \quad (4)$$

EM 距离的统计学含义是使一个分布尽量向另一个分布靠拢, 为了将其转化为可执行的优化问题, 根据 Kantorovich-Rubinstein 定理, 得到式(5):

$$EM(P_r, P_s) = \sup_{\|f\|_L \leq 1} E_{x \sim P_r} [f(x)] - E_{x \sim P_s} [f(x)], \quad (5)$$

$$\max_{\|f\|_L \leq 1} E_{x \sim P_r} [f(x)] - E_{x \sim P_s} [f(x)], \quad (6)$$

式(5)中,  $f$  是一阶 Lipschitz 函数. 通过优化式(6), 可以确定一组函数  $\{f_\omega\}_{\omega \in \mathbb{W}}$ , 其中各项均为  $K$  阶 Lipschitz 函数, 满足  $\|f\|_L \leq K$ . 由此得到式(7):

$$\max_{\omega \in \mathbb{W}} E_{x \sim P_r} [f(x)] - E_{x \sim P_s} [f(x)]. \tag{7}$$

为了将  $K$  阶 Lipschitz 函数中的  $\omega$  限制在约定范围内,我们将每一轮权值的变化量约束在固定范围. 具体地,在 GAN 的训练中,判别器充当了式(7)中  $f$  的作用.

综上,我们对原始的 CycleGAN 进行了如下几点改进:

(1)用式(8)的 Wasserstein 损失取代原来的对抗损失( $L_{\text{LSGAN}}(F,D_S,R,S)$ 同理),

$$L_{\text{WGAN}}(G,D_R,S,R) = E_{r \sim P_{\text{data}}(R)} [D_R(R)] - E_{r \sim P_{\text{data}}(S)} [D_R(G(S))]. \tag{8}$$

(2)增加每次生成器更新之前判别器更新的轮次.

(3)在每次参数更新后,限制判别器权重的大小.

2.2 C-RNN 的结构与训练

识别网络将 CNN 与 RNN 相结合,识别过程包含 3 步:序列特征提取、序列识别与序列解码,这 3 个步骤与网络中的卷积部分、循环部分与解码部分一一对应,网络细节如表 1.

表 1 C-RNN 模型的配置

Table 1 Configuration of the C-RNN model

层类型	配置	层类型	配置
Bidirectional-LSTM	#hidden units:256	MaxPooling	s:1×2,p:1×0
Bidirectional-LSTM	#hidden units:256	ReLu	—
ReLu	—	Convolution	#filters:256,k:3×3,s:1,p:1
BatchNormalization	—	ReLu	—
Convolution	#filters:512,k:2×2,s:1,p:0	BatchNormalization	—
ReLu	—	Convolution	#filters:256,k:3×3,s:1,p:1
BatchNormalization	—	MaxPooling	p:2×2,s:2
Convolution	#filters:512,k:2×2,s:1,p:0	ReLu	—
MaxPooling	s:1×2,p:1×0	Convolution	#filters:128,k:3×3,s:1,p:1
ReLu	—	MaxPooling	p:2×2,s:2
Convolution	#filters:512,k:3×3,s:1,p:1	ReLu	—
ReLu	—	Convolution	#filters:64,k:3×3,s:1,p:1
BatchNormalization	—	Input	160×48×3 RGB images
Convolution	#filters:512,k:3×3,s:1,p:1		

2.2.1 序列特征提取

现有研究证明,CNN 在图像特征提取上有着极大的优势<sup>[28]</sup>. 本文使用 8 层的 CNN 网络,输入为 160×48×3 的 RGB 图像,输出的特征序列,即为后续 RNN 的输入.

2.2.2 序列识别

为了捕获车牌字符间的上下文关系,我们在 CNN 的末端串联了一个双向 RNN 网络,并采用 LSTM 来预防梯度消失的问题. 在双向 LSTM 之后,是一个输出宽度为 68 的全连接层,对应着车牌中所有可能出现的字符:31 个中文字符,26 个英文字符,10 个阿拉伯数字,以及空白位. 最终,图像的特征序列被转换为置信度序列  $X = \{x^1, x^2, \dots, x^T\}$ ,  $T$  代表输入特征序列的长度,等同于时序数据处理中的时间片数量.  $x^t$  代表第  $t$  个时间片的概率分布; $x_k^t$  则是在第  $t$  个时间片标签  $k$  的置信度. 输出的图形化表示如图 2.

2.2.3 序列解码

得到置信度分布序列之后,可以利用最佳路径算法<sup>[29]</sup>得到字符序列. 输入  $t$  对应路径  $\pi$  的置信度由式(9)得到:

$$p(\pi | x) = \prod_{t=1}^T y_{\pi_t}^t. \tag{9}$$

依照上文的描述, $y_{\pi_t}^t$  代表了在时间片  $t$  得到标签  $\pi_t$  的可能性. 如式(10),借助最佳路径算法,可以找出置信度最大的路径.

$$l = \mathcal{B}(\pi^*), \quad \text{where } \pi^* = \operatorname{argmax} p(\pi | x), \tag{10}$$

这里  $\mathcal{B}$  代表移除所有空白格与连续重复项,如  $\mathcal{B}(aa--ab--)=$

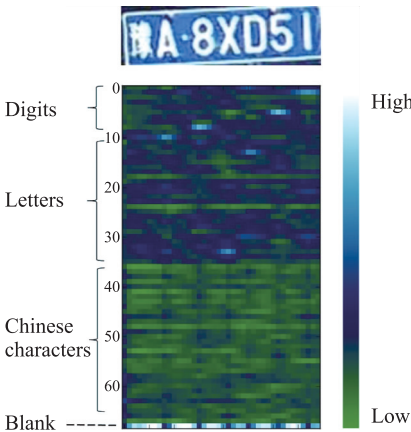


图 2 车牌识别置信度序列

Fig. 2 License plate recognition confidence map



$\mathcal{B}(-a-aa-b)=aab$ . 理论上,最优路径算法得到的结果并非一定是最优解,因为它在每一个时间片选择置信度最高的标签加入序列,但实践证实,这样的策略足以在绝大多数情况下获得正确的结果. 训练识别模型时,损失的计算根据时序分类损失(CTC)<sup>[29]</sup>,使用的优化算法为 Adadelta<sup>[30]</sup>.

### 3 实验设置

#### 3.1 数据集

我们收集了大量动态车牌照片,组成两个数据集,这些照片都来自于真实的交通监控,视角、动态模糊、光照等条件都相对复杂. 如表 2 所示,一号数据集包含 203 774 张训练图片和 9 986 张测试图片;二号数据集包含 45 139 张训练图片和 5 825 张测试图片,二者数据来源不同,在分布上存在一些差别. 图 3 与图 4 展示一些车牌照片的样例,用这些动态场景下的照片训练模型是极富挑战性的.

表 2 真实数据集概况

Table 2 Composition of the real datasets

名称	宽(像素)	高(像素)	采集时间	车牌底色	训练图像数/张	测试图像数/张
一号数据集	100~160	30~50	白天,夜晚	蓝,黄	203 774	9 986
二号数据集	50~150	15~40	白天,夜晚	蓝,黄	45 139	5 825



图 3 一号数据集部分实例

Fig. 3 Some samples of dataset-1



图 4 二号数据集部分实例

Fig. 4 Some samples of dataset-2

#### 3.2 评价指标

本文用车牌识别准确度(recognition accuracy, RA)和字符识别准确度(character recognition accuracy, CRA)衡量模型性能,前者是识别正确的序列数与全部序列数之比,后者是识别正确的字符数与全部字符数之比.

此外,我们还计算了 Top- $N$  的识别准确率,即所有的结果中置信度最高的前  $N$  个中包含正确结果的比率. 根据 LSTM 的输出,利用定向搜索算法很容易得到 Top- $N$  的结果. 在诸如刑侦和车牌验证等实际应用中,Top- $N$  的结果也是有较高参考价值的.

#### 3.3 实施细节

##### 3.3.1 识别框架

C-RNN 搭建与训练用 TensorFlow<sup>[31]</sup>完成. 图片被放缩为  $160 \times 48$  的大小输入到网络中,网络最后的全连接层包含 68 个输出节点. 得到输出后,使用上述基于贪心原则的搜索算法得到 Top-1 和 Top- $N$  的结果. Graves 等人<sup>[29]</sup>使用了一种前向搜索的算法,能略微提高识别准确度,但考虑到计算复杂度,本文仍沿用之前的方法.

##### 3.3.2 粗合成图片

仿照真实车牌的字体、颜色和字符分布,用传统图像处理方法得到总计 200 k 张粗合成图片. 并且适当地引入了高斯噪声、运动模糊以及仿射变换来模拟真实状况. 此外,还有一些具体限制需要注意,如第二个字符必须是英文字符,又如后面的 5 个字符中至多有 2 个英文字符(除了“ $I$ ”与“ $O$ ”),图 5(a)展示了部分粗合成样例.

##### 3.3.3 GAN 训练和测试

首先用 4 500 张粗合成蓝底车牌和 4 500 张真实蓝底车牌(来源于一号数据集)训练 CycleGAN. 式(1)中参数  $\lambda$  取 10. 所有图片放缩到  $143 \times 143$  的大小,并随机截取  $128 \times 128$  的图像块. 优化器选用 Adam,  $\beta_1=0.5, \beta_2=0.999$ ,学习率为 0.000 2. 训练持续 30 个 epoch,每个 epoch 后保存中间模型. 测试时用 20 个

模型共生成 100 k 蓝底图像. 对黄底车牌也进行同样的操作. 最终我们获得了 200 k 张合成图片. Wasserstein 的  $d_{\text{iter}}$  设为 5, 判别器的参数被截取到  $[-0.01, 0.01]$ . 后续训练时, 200 k 张图片又被随机截取, 得到总计 800 k 例训练图片.

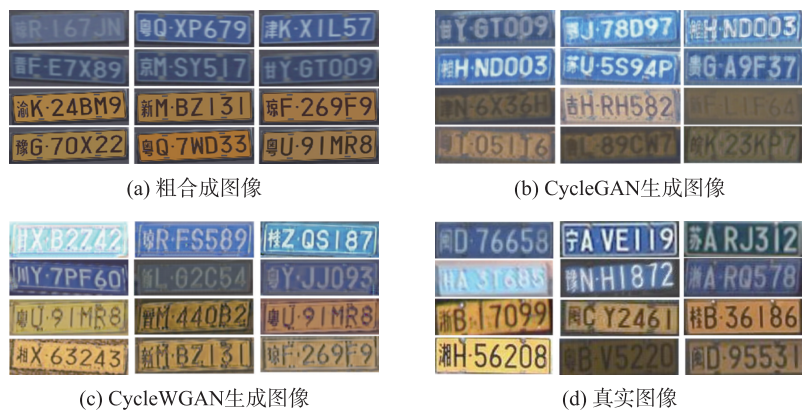


图 5 各类合成图像与真实图像对比

Fig. 5 Comparison of various synthetic images, generated images and real images

4 实验结果

4.1 模型评估

我们在两个真实数据集上直接训练 C-RNN, 得到基线, 结果如表 3 和表 4. 为了验证 C-RNN 框架的有效性, 本文引入 EasyPR<sup>[32]</sup> 进行对比, EasyPR 是基于支持向量机的开源 LPR 框架, 属于有分割方法. 如表 5 所示, 本文使用的 C-RNN 模型在效果上大大超越了 EasyPR.

为了衡量数据生成方法对模型的影响, 我们直接用粗合成数据训练了模型, 并将其与用 CycleGAN 和 CycleWGAN 辅助训练的模型相对比, 结果如表 6. 可以看到, 只使用粗合成数据训练的模型效果很差, 对比图 3 与图 5(a), 不难发现目前的粗合成数据风格过于单一, 不利于泛化.

表 3 不同生成数据量训练结果比较

Table 3 Comparison of models using different volumes of the real data and synthetic data					%
训练数据	模型	准确率 (Top-1)	字符识别准确率	准确率 (Top-3)	准确率 (Top-5)
9 k	Baseline	86.1	94.9	90.1	92.3
	Script	90.2	97.0	94.6	95.6
	CycleWGAN	93.6	98.4	96.8	97.4
50 k	Baseline	93.1	97.9	96.4	97.2
	Script	95.2	98.8	97.7	98.1
	CycleWGAN	96.3	99.2	98.3	98.8
200 k (All)	Baseline	96.1	98.9	98.0	98.5
	Script	96.7	99.1	98.6	98.8
	CycleWGAN	97.6	99.5	98.9	99.2

表 4 本文方法与基线在二号数据集的效果对比

Table 4 Comparison of our method and baseline on dataset-2					%
方法	准确率 (Top-1)	字符识别准确率	准确率 (Top-3)	准确率 (Top-5)	
Baseline	94.5	98.4	97.6	98.1	
CycleWGAN	96.2	99.2	98.7	99.1	

表 5 不同框架在一号数据集上的效果比较

Table 5 Comparison of different frameworks on dataset-1			%
框架	准确率 (Top-1)	字符识别准确率	
SVM+ANN <sup>[19]</sup>	68.2	82.5	
CNN+LSTMs	96.1	98.9	

表 6 不同生成模型在一号数据集上的性能对比

Table 6 Comparison of models that use different generated methods on dataset-1					%
方法	准确率(Top-1)	字符识别准确率	中文字符识别准确率	数字和英文字符识别准确率	
Script	4.4	30.0	20.0	31.7	
CycleGAN	34.6	82.8	41.3	89.8	
CycleWGAN	61.3	90.6	66.2	94.8	

CycleGAN 的引入将准确度由 4.4%提升到了 34.6%,说明此时生成图片已经更接近真实的数据. 图 5(b) 展示了部分样例. CycleGAN 还有较大的提升空间,因为此时所有的图片倾向于同一种风格,证实了目前还是存在一定的结果单一化现象<sup>[14,33]</sup>.

如图 5(c)所示,CycleWGAN 生成图像的风格更加多样化,也说明 Wasserstein 损失的有效性,许多图片甚至已经达到以假乱真的程度. 使用 CycleWGAN 之后,Top-1 准确度提升至 61.3%. 非中文字符的准确度更是达到了 94.8%. 可见,生成图像越接近真实情况,得到的模型效果越好.

4.2 结果分析

当前并没有研究能从理论上说明 GAN 生成数据对模型训练的影响. 本文从实验出发给出解释. 深度模型可以视作是从给定输入到输出的条件分布<sup>[34-35]</sup>. 输入 5 000 张随机噪声图像,再求取输出的平均,得到图 6 的输出. 图 6(b)展示了单独使用生成图像训练的模型的结果,以第二个字符对应的列为例,反应主要集中在中间区域,而这一区域对应的正是英文字母,说明模型已经一定程度上捕获了字符的位置分布信息.

此外,本文还筛选出了基线模型判别错误,而后续模型判别正确的测试用例,发现其中约有 35%是在中文字符上误判,25%则是因为一些易混淆的字符,如“0”与“D”、“8”与“B”以及“5”与“S”. 综上,无论是

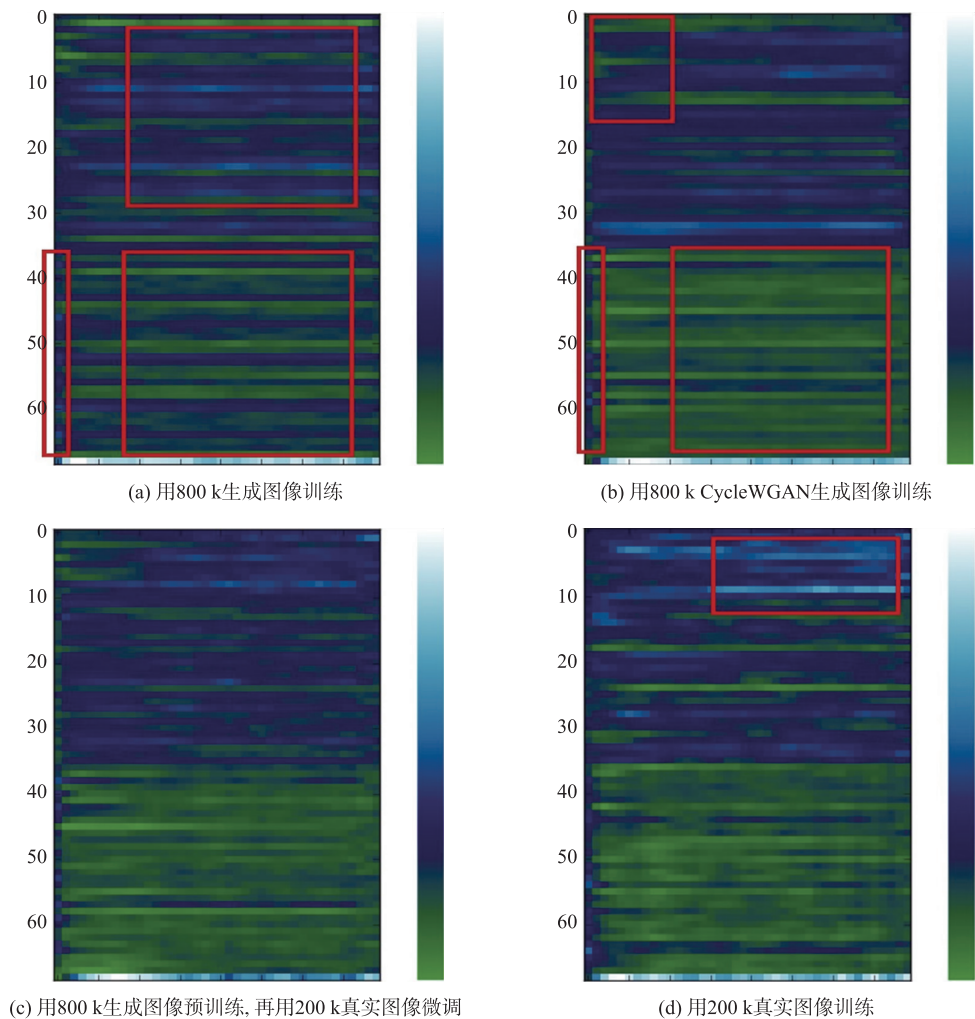


图 6 在不同数据集上训练的模型的比较

Fig. 6 Comparison of models trained on different datasets



粗合成时的人工先验信息,还是后续 GAN 捕获到的真实图像特征,都对模型的训练有所裨益.

4.3 模型压缩

为了进一步满足移动端和嵌入式端的需求,本文用深度可分离卷积<sup>[36]</sup>替代标准卷积,并使用了[37]中的超参数宽度乘数. 一个深度可分离卷积模块包含一个分离卷积层和 1×1 卷积层,分离卷积的核心思想是一个卷积核只用于一个通道. 最终得到了更轻量级的模型——LightCRNN,其整体结构如表 7.

我们在不使用 GPU,且只使用 1 个 CPU 核(2.40 GHz Intel (R) Xeon (R) E5-2620 CPU and 64 GB RAM)的条件下进行测试,结果如表 8. 模型大小由 71.4 MB 降低到了 40.3 MB,运行速度由 7.2 FPS 上升到 13.9 FPS,同时准确度也有所上升,出现这一状况可能是因为参数量的减少避免了模型陷入过拟合,表中 1.2 代表宽度乘数.

表 7 LightCRNN 模型配置  
Table 7 Configuration of LightCRNN model

层类型	配置	层类型	配置
Bidirectional-LSTM	#hidden units:256	MaxPooling	s:1×2,p:1×0
Bidirectional-LSTM	#hidden units:256	Conv	#filters:256,k:1×1,s:1,p:0
Conv	#filters:512,k:1×1,s:1,p:0	Depthwise Conv	k:3×3,s:2,p:1
Depthwise Conv	k:2×2,s:1,p:0	Conv	#filters:128,k:1×1,s:1,p:0
Conv	#filters:512,k:1×1,s:1,p:0	Conv	#filters:128,k:1×1,s:1,p:0
Depthwise Conv	k:2×2,s:1,p:0	Depthwise Conv	k:3×3,s:2,p:1
MaxPooling	s:1×2,p:1×0	Conv	#filters:64,k:3×3,s:1,p:1
Conv	#filters:512,k:1×1,s:1,p:0	Input	
Depthwise Conv	k:3×3,s:1,p:1		

表 8 LightCRNN 模型效果比较  
Table 8 Comparison of using different models

模型	准确率(Top-1)/%	模型大小/MB	速度/FPS
LightCRNN	96.5	40.3	13.9
1.2 LightCRNN	97.0	44.2	11.5
1.2 LightCRNN+CycleWGAN images	98.6	44.2	11.5

4.4 动态 LPR

为了检验在更复杂数据上的作用,我们在一辆车上安装了摄像机,让其在行进时拍摄周围车辆的车牌. 这些照片无论视角还是对象都是运动的,比静态或是只有一者运动的情况都更复杂,此类问题称为动态 LPR,该技术在辅助驾驶、道路监控和移动机器人领域都有广泛的应用.

动态 LPR 存在两个瓶颈. 首先数据成本更高,其次往往只能部署在移动平台,而本文方法中的数据生成机制和 LightCRNN 恰好可以应对这两个问题.

图 7 展示了部分实例. 这部分数据在苏州采集,涵盖了丰富的场景,包括夜晚、高速公路、十字路口等. 最终得到了 20 026 张训练图片和 2 000 张测试图片,这部分数据称之为三号数据集.

先用三号数据集训练 LightCRNN 基线,得到了 89.4%的准确度. 应用本文方法,最终使准确率上升了 2.7%,如表 9 所示. 图 7(b)展示了部分生成的样例.



图 7 三号数据集中部分样例及用于动态 LPR 的 CycleWGAN 生成图像样例  
Fig. 7 Some samples of dataset-3 and samples of CycleWGAN images for moving LPR



表 9 三号数据集上的测试结果

Table 9 Results on dataset-3

%

方法	准确率 (Top-1)	字符识别准确率
Baseline	89.4	97.6
CycleWGAN+	92.1	98.0

## 5 结语

本文方法借助 GAN 生成训练图片,辅助车牌识别模型进行训练. 使用 CycleWGAN 合成高质量的车牌图片,极大地拓展了训练数据的规模. 在进行车牌识别时,借助 C-RNN 模型,先用合成图片预训练,再使用真实数据微调,极大地提升了识别的准确率. 实验结果表明,本文方法提高了模型在困难数据上的表现,特别是原始数据有限的情况下. 此外,本文还用实证的方式探讨了 GAN 生成数据对模型的监督作用. 最后,为了适应轻量级的平台,本文还对模型进行了压缩,提出了 LightCRNN,使识别速度提升近 1 倍.

### [ 参考文献 ]

- [1] LI H, SHEN C. Reading car license plates using deep convolutional neural networks and LSTMs[EB/OL]. [2019-06-01]. <https://arxiv.org/abs/1601.05610v1>.
- [2] SHI B, BAI X, YAO C. An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition[J]. IEEE transactions on pattern analysis and machine intelligence, 2017, 39(11): 2298-2304.
- [3] ZHU J Y, PARK T, ISOLA P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice: IEEE Computer Society, 2017: 2223-2232.
- [4] GOU C, WANG K, YAO Y, 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.
- [5] SHIN H C, TENENHOLTZ N A, ROGERS J K, et al. Medical image synthesis for data augmentation and anonymization using generative adversarial networks [C]//International Workshop on Simulation and Synthesis in Medical Imaging. Berlin: Springer, 2018.
- [6] NOMURA S, YAMANAKA K, KATAI O, et al. A novel adaptive morphological approach for degraded character image segmentation[J]. Pattern recognition, 2005, 38(11): 1961-1975.
- [7] SARFRAZ M, AHMED M J. Exploring critical approaches of evolutionary computation: an approach to license plate recognition system using neural network[M]. Hershey: IGI Global, 2019: 20-36.
- [8] JIAO J, YE Q, HUANG Q. A configurable method for multi-style license plate recognition[J]. Pattern recognition, 2009, 42(3): 358-369.
- [9] NAIR A S, RAJU S, HARIKRISHNAN K J, et al. A survey of techniques for license plate detection and recognition[J]. Imanager's journal on image processing, 2018, 5(1): 25.
- [10] WELANDER P, KARLSSON S, EKLUND A. Generative adversarial networks for image-to-image translation on multi-contrast MR images: a comparison of CycleGAN and UNIT[EB/OL]. [2019-06-01]. <https://arxiv.org/abs/1806.07777>.
- [11] LLORENS D, MARZAL A, PALAZÓN V, et al. Car license plates extraction and recognition based on connected components analysis and HMM decoding[C]//Iberian Conference on Pattern Recognition and Image Analysis. Berlin: Springer, 2005: 571-578.
- [12] WEN Y, LU Y, YAN J, et al. An algorithm for license plate recognition applied to intelligent transportation system[J]. IEEE transactions on intelligent transportation systems, 2011, 12(3): 830-845.
- [13] TAO D, LIN X, JIN L, et al. Principal component 2-D long short-term memory for font recognition on single Chinese characters[J]. IEEE transactions on cybernetics, 2016, 46(3): 756-765.
- [14] GOODFELLOW I, POUGET A J, MIRZA M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27: 2672-2680.
- [15] RADFORD A, METZ L, CHINTALA S. Unsupervised representation learning with deep convolutional generative adversarial networks[EB/OL]. [2019-06-01]. <https://arxiv.org/abs/1511.06434>.
- [16] REED S, AKATA Z, YAN X, et al. Generative adversarial text to image synthesis[EB/OL]. [2019-06-01]. <https://arxiv.org/abs/1609.09909>.

- org/abs/1605.05396.
- [17] ISOLA P, ZHU J Y, ZHOU T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Hawaii:IEEE Computer Society, 2017:1125–1134.
- [18] ARJOVSKY M, CHINTALA S, BOTTOU L. Wasserstein generative adversarial networks[J]. Proceedings of machine learning research, 2017, 70:214–223.
- [19] PATHAK D, KRAHENBUHL P, DONAHUE J, et al. Context encoders:feature learning by inpainting[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas:IEEE Computer Society, 2016:2536–2544.
- [20] SALIMANS T, GOODFELLOW I, ZAREMBA W, et al. Improved techniques for training gans[J]. Advances in neural information processing systems, 2016, 29:2234–2242.
- [21] WU J, ZHANG C, XUE T, et al. Learning a probabilistic latent space of object shapes via 3d generative-adversarial modeling[J]. Advances in neural information processing systems, 2016 29:82–90.
- [22] GUPTA A, VEDALDI A, ZISSERMAN A. Synthetic data for text localisation in natural images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas:IEEE Computer Society, 2016:2315–2324.
- [23] JADERBERG M, SIMONYAN K, VEDALDI A, et al. Synthetic data and artificial neural networks for natural scene text recognition[EB/OL]. [2019-06-01]. <https://arxiv.org/abs/1406.2227>.
- [24] WANG Z, YANG J, JIN H, et al. Deepfont:identify your font from an image[C]//Proceedings of the 23rd ACM International Conference on Multimedia. Brisbane:ACM, 2015:451–459.
- [25] YU J, FARIN D, KRÜGER C, et al. Improving person detection using synthetic training data[C]//2010 IEEE International Conference on Image Processing. Hong Kong:IEEE Computer Society, 2010:3477–3480.
- [26] ROS G, SELLART L, MATERZYNSKA J, et al. The synthia dataset: a large collection of synthetic images for semantic segmentation of urban scenes[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Las Vegas: IEEE Computer Society, 2016:3234–3243.
- [27] ZHENG Z, ZHENG L, YANG Y. Unlabeled samples generated by gan improve the person re-identification baseline in vitro[C]//Proceedings of the IEEE International Conference on Computer Vision. Venice:IEEE Computer Society, 2017:3754–3762.
- [28] WEI Y, ZHAO Y, LU C, et al. Cross-modal retrieval with CNN visual features; a new baseline[J]. IEEE transactions on cybernetics, 2017, 47(2):449–460.
- [29] GRAVES A, FERNÁNDEZ S, GOMEZ F, et al. Connectionist temporal classification: labelling unsegmented sequence data with recurrent neural networks[C]//Proceedings of the 23rd International Conference on Machine Learning. Pittsburgh: ACM, 2006:369–376.
- [30] ZEILER M D. ADADELTA:an adaptive learning rate method[EB/OL]. [2019-06-01]. <https://arxiv.org/abs/1212.5701>.
- [31] ABADI M, AGARWAL A, BARHAM P, et al. Tensorflow:large-scale machine learning on heterogeneous distributed systems [EB/OL]. [2019-06-01]. <https://www.arxiv-vanity.com/papers/1603.04467/>.
- [32] LIU R, LI M. EasyPR[EB/OL]. [2019-06-01]. <https://github.com/liurouze/EasyPR>.
- [33] IM D J, KIM C D, JIANG H, et al. Generating images with recurrent adversarial networks[EB/OL]. [2019-06-01]. <https://arxiv.org/abs/1602.05110>.
- [34] HINTON G, VINYALS O, DEAN J. Distilling the knowledge in a neural network[EB/OL]. [2019-06-01]. <https://arxiv.org/abs/1503.02531>.
- [35] PEREYRA G, TUCKER G, CHOROWSKI J, et al. Regularizing neural networks by penalizing confident output distributions [EB/OL]. [2019-06-01]. <https://arxiv.org/abs/1701.06548>.
- [36] SIFRE L, MALLAT S. Rigid-motion scattering for image classification[D]. Palaiseau:Ecole Polytechnique CMAP, 2014.
- [37] HOWARD A G, ZHU M, CHEN B, et al. Mobilenets:efficient convolutional neural networks for mobile vision applications [EB/OL]. [2019-06-01]. <https://arxiv.org/abs/1704.04861>.

[ 责任编辑:丁 蓉 ]