

面向精准价格牌识别的多任务循环神经网络

牟永强¹ 范宝杰^{1,2} 孙超¹ 严蕤¹ 郭怡适¹

摘要 为了促进智能新零售在线下业务场景的发展,提高作为销售关键信息价格牌的识别精度.本文对价格牌识别问题进行研究,有效地提高了价格牌的识别精度,并解决小数点定位不准确的难题.通过深度卷积神经网络提取价格牌的深度语义表达特征,将提取到的特征图送入多任务循环神经网络进行编码,然后根据解码网络设计的注意力机制解码出价格数字,最后将多个分支的结果整合并输出完整价格.本文所提出的方法能够非常有效地提高线下零售场景价格牌的识别精度,并解决了一些领域难题如小数点的定位问题,此外,为了验证本文方法的普适性,在其他场景数据集上进行了对比实验,相关结果也验证了本文方法的有效性.

关键词 卷积神经网络, 循环神经网络, 文本识别, 多任务学习, 价格牌识别

引用格式 牟永强, 范宝杰, 孙超, 严蕤, 郭怡适. 面向精准价格牌识别的多任务循环神经网络. 自动化学报, 2022, 48(2): 608-614

DOI 10.16383/j.aas.c190633

Towards Accurate Price Tag Recognition Algorithm With Multi-task RNN

MOU Yong-Qiang¹ FAN Bao-Jie^{1,2} SUN Chao¹ YAN Rui¹ GUO Yi-Shi¹

Abstract In order to promote the development of smart new retail in the offline scenario and improve the recognition accuracy of price tags, which is a key sales information, this paper studies this application scene to improve the recognition accuracy effectively of the price tag and solves the difficulty of locating decimal point. The deep semantic expression features of price tag are extracted by the deep convolutional neural network, sent to the multi-task recurrent network layer for encoding, and then the price number is decoded according to the attention mechanism of the decoding network, and finally the results of multiple branches are integrated to output the complete price. The method we proposed can effectively improve the price recognition accuracy in the smart new retail scenario and solve some challenge problems such as locating the position of decimal point. In addition, in order to verify the universality of the method, comparative experiments are carried out on datasets of other scenarios, and the related results also verify the effectiveness of the method in this paper.

Key words Convolutional neural network, recurrent neural network, text recognition, multi-task learning, price tag recognition

Citation Mou Yong-Qiang, Fan Bao-Jie, Sun Chao, Yan Rui, Guo Yi-Shi. Towards accurate price tag recognition algorithm with multi-task RNN. *Acta Automatica Sinica*, 2022, 48(2): 608-614

传统零售业抑或是近年来兴起的快消新零售,渠道核查是其中的必要环节.传统的作业方式主要分为业务代表现场考察以及第三方外包核查,但都存在人工误差大、核查周期长、核查成本高以及误差数据无法溯源等缺点.随着深度学习的迅速发展,AI(人工智能)已经成为高端科技的代名词,各行各业的AI应用层出不穷.基于深度学习的图像识别技术凭借着高精度、高泛化性,非常适合应用于

渠道核查的业务场景,是核查工作强大的助力.渠道核查主要包含两大识别内容,SKU(Stock keeping unit), (库存量单位)识别和价格牌识别,本文工作主要针对价格牌识别的需求.价格作为销售数据的基石,对识别精度非常敏感,目前基于深度学习的价格牌识别技术容易受到其外观样式、拍摄质量等因素的影响,如模糊、倾斜、光照不均匀等.因此,如何克服实际应用中可能遇到的复杂场景,准确识别价格牌中的信息是OCR(Optical character recognition)领域的一个重要研究目标.

目前,应用性较广的价格牌识别算法大多以文本识别算法为基础.基于卷积循环神经网络(Convolution recurrent neural network, CRNN)^[1]的识别方法,为序列识别任务带来了突破性的进展,也为文本识别领域打开了一扇大门.随后基于CRNN

收稿日期 2019-09-06 录用日期 2020-02-23

Manuscript received September 6, 2019; accepted February 23, 2020

本文责任编辑 金连文

Recommended by Associate Editor JIN Lian-Wen

1. 广州图匠数据科技有限公司人工智能实验室 广州 510310
2. 广州工业大学信息工程学院 广州 510006

1. AI-Labs, Guangzhou Image Data Technology Co., Ltd., Guangzhou 510310 2. College of Information Engineering, Guangdong University of Technology, Guangzhou 510006

变体和各种注意力机制的文本识别算法^[2-3]层出不穷, 相较于前者, 增加的注意力机制主要用于关联输入信息的相关性, 这种方式显著提高了通用文本的识别精度。

目前国内外的文本识别研究, 普遍关注没有符号的文字序列。对于价格牌这类带有符号的序列识别, 一些在通用文本数据集上表现优异的算法^[4-5], 性能并不能令人满意。为此本文提出了一种多任务的卷积神经网络, 有效地提高了价格牌的识别精度。

在价格牌的识别任务中, 精准地识别所占像素比例很小或直接被省略的小数点, 是非常困难的一项任务, 也是其区别其他文本图像识别任务的重点。现有的绝大部分算法是将价格牌的整体进行无差别的识别, 但是由于价格牌的种类繁多, 以及一些客观因素的影响, 导致其在图像中的特征并不明显, 即使采用基于上下文关系的序列识别算法也很难准确定位小数点的位置。为此本文提出了一种将整数部分与小数部分分开, 协同识别整体的方法, 实现对小数点的准确定位。使用端对端的多任务训练策略进行学习, 降低训练的难度。经过实验证明, 本文提出的方法不仅在识别精度上有着优越的指标, 对于小数点的识别更是超越了以往深度学习算法的成绩。

由于已开源的数据集中暂无价格牌这一特定场景, 我们将实验中使用的价格牌数据集开源出来以供研究使用。我们的数据集采集自真实货架场景图像, 涵盖不同样式, 不同拍摄角度, 不同光照变化等, 其中包含训练集 10 000 张, 测试集 1 000 张, 困难测试集 1 000 张 (包含了手写价格、模糊价格以及其他影响因素的价格数据), 训练集及测试集的数字区域比较清晰, 辨识度较高, 而困难测试集的数字区域大都存在干扰项 (如反光、拍摄重影、双价格标签等), 辨识度较低。此外, 为了进一步验证本文所提出方法的泛化能力, 我们在类似的车牌数据集中也进行了相关实验, 实验结果表明了本文所提出方案的有效性。

1 相关工作

1.1 卷积循环神经网络

OCR (Optical character recognition, 光学字符识别, 现泛指所有图像文字检测和识别技术) 的研究, 一直是图像识别领域的重要研究方向之一。随着深度学习研究的飞跃, 关于自然场景的图像文本识别算法不胜枚举, 掀起了一轮又一轮的竞赛狂潮。

CRNN 主要用于图像的序列识别问题, 包含卷积层、循环层和转录层, 结构如图 1 所示, 是 OCR 技术的常用模型。CRNN 主要可以分为以下几个部

分: 首先输入图像预处理后通过深层卷积神经网络, 得到输出的高级特征图 (Feature map); 随后将 feature map 的每一列或每几列作为一个时间序列输入由双向 LSTM (Bi-directional Long short-term memory) 网络构成的循环层; 最后输出一个序列标签 (预测特征序列中的每一个特征向量的标签分布——真实结果的概率列表)。转录层采用 CTC (Connectionist temporal classification), (时序连接分类)^[6] 或者其他高效的序列分类方法^[7] 进行转录, 处理循环层所输出的序列标签, 将所有可能的“字符定位”结果进行整合, 转换为最终的识别结果。

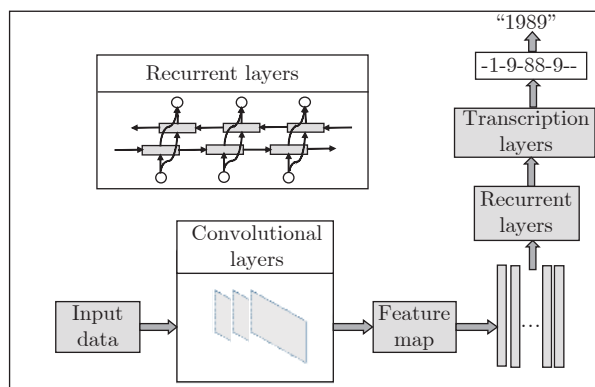


图 1 卷积循环神经网络结构

Fig. 1 The structure of convolutional recurrent neural network

虽然 CRNN 的结构理论上可以预测任意的序列对应关系, 但实际中编码和解码的准确度很大程度上依赖于语义向量。语义向量在编码压缩过程中存在信息丢失, 而语义向量的信息偏差会严重影响解码端的准确率。其次, 解码过程在每个时间步使用的内容向量是相同的, 这也会对解码准确率造成一定程度的影响。为了解决以上问题, CRNN 模型加入了注意力机制^[8]。

不同的注意力机制对序列的处理方法不同, 应用较广泛的注意力机制^[8]是由编码器将输入数据编码成一个向量的序列后, 在解码阶段的每一个时间步, 注意力模型都会选择性地从向量序列中挑选出一个子集进行输出预测 (这种选择基于解码阶段隐层状态与输入序列的相关性)。这种机制可以保证在产生每一个输出的时候, 都能找到当前输入序列应该重点关注的信息, 也表明每一个输出所参考的语义向量都是不同的。

1.2 多任务学习

深度学习中单任务学习模型关注点通常是对某一个特定度量进行优化, 比如分类精度、识别精度

或者回归指标等. 在训练的基准模型上, 我们不断地微调模型, 直到模型的结果不能继续优化. 虽然这种方法可以得到高于基准模型的结果, 但我们选择性地忽略了可能提升特征度量指标的其他信息.

区别于单任务模型将注意力聚焦于某个度量, 多任务学习可以共享相关任务之间的表征, 使模型可以更好地学习原始任务. 某种程度上, 多任务学习可以认为是人类学习的思维延伸, 通过人类学习的先验知识, 关联多任务之间的表征信息. 从信息学的角度, 可以将多任务学习视为信息归纳转移的一种方式.

分析价格牌数据, 识别过程最大的困难便是小数点的定位. 如图 2 所示, 小数点的位置总是模糊不清或被省略, 单任务的端到端网络包括针对复杂文本的^[9]也很难做到定位小数点. 因此, 提出拆分价格牌的整数部分和小数部分, 通过多任务学习的策略联合学习小数点的特征信息, 价格牌拆分示意图如图 3 所示. 这种策略需要价格牌数据结构的先验知识, 将分支结果与小数点后期拼接, 得到完整的价格牌数据.



图 2 价格牌图像
Fig. 2 Images of some price tag samples

在计算机视觉领域, 最常见的多任务学习方法便是共享卷积层^[10]参数, 同时独立学习特定任务的其他层参数.

2 价格牌识别算法

2.1 多任务模型分析

CRNN 及其变体的结构在 Coco^[11]、ICDAR2015^[12]等通用文本数据集上取得了优异成绩, 证明了其方法的有效性. 文献 [13] 对近年来具有代表性的文本识别算法结构进行了总结, 通过实验分析, 确定了在自然文本数据集上表现最优的 CRNN 结构.

沿着 CRNN 的方法, 我们使用卷积网络提取文本的特征, 沿宽度方向切片作为输入特征送入

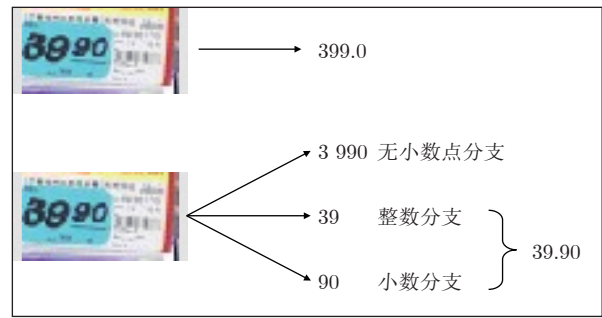


图 3 基准识别与多分支识别结果的生成方式
Fig. 3 Baseline method compared with multi-branch method

循环层, 得到特征序列的标签分布, 之后用基于 LSTM 的编码器和解码器将特征序列转换为最终的识别结果, 网络结构如图 4 所示.

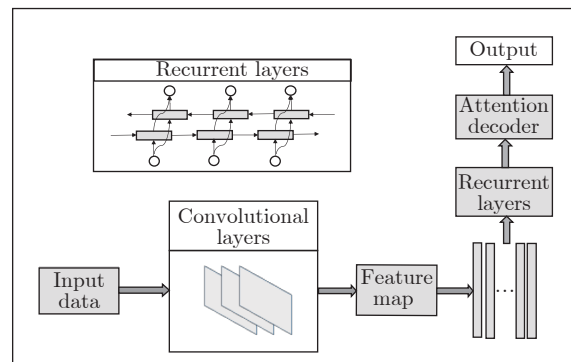


图 4 基础单任务识别网络结构
Fig. 4 The structure of our basic single recognition network

本文设计的多任务学习模型不同于一般联合学习^[14], 而是基于价格牌可拆分的数据结构知识. 整体结构如图 5 所示, 其中 IB (Integer branch) 表示整数分支, DB (Decimal branch) 表示小数分支, NDPB (No decimal point branch) 表示去小数点的数字分支, 如图 3 所示. 模型分支结构完全相同, 在特征提取阶段后, 学习序列不同感受野的信息. 无小数点字符串分支作为辅助损失抑制整数分支与小分支的过拟合, 共同优化共享的卷积块参数. 三分支网络结构与损失函数完全相同, 通过对应不同的标签优化网络参数, 极大地简化训练流程. 这里我们之所以选择三分支的模型, 也是由于应用场景的特殊性, 在实验阶段我们也会输出不同分支组合结果进行分析.

相比于单一任务的方法, 我们所提出的多任务模型机制也拥有更好的可分析性: 对于价格牌识别问题, 我们选取多任务的结构可以数据化模型对整

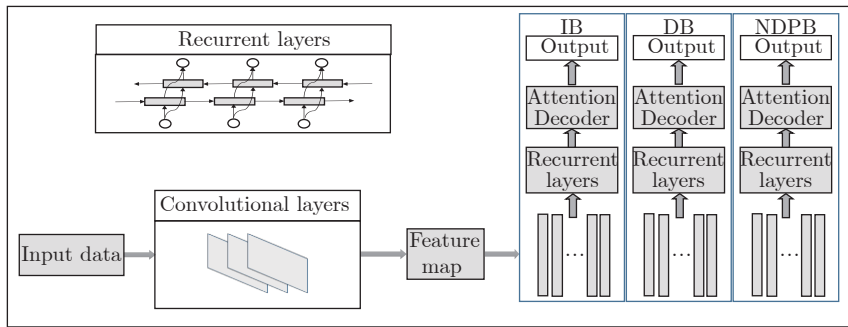


图 5 多任务循环卷积网络结构

Fig.5 The structure of multi-task RNN

数部分以及小数部分的识别准确度, 从而分析误判问题. 对于不同分支的识别难度有初步估计, 从而制定相应的训练策略, 如去小数点分支融入的可训练超参数权重, 这种策略对模型精度有可观的改善.

2.2 解码流程分析

解码阶段以单向 LSTM 作为解码网络, 增加了循环层注意力机制, 结构如图 6 所示. 所提出的多任务模型采用相同方式解码, 分支损失函数为式 (1) 所示的交叉熵函数, 其中 M 为每批次数列数, N 为解码端单向 LSTM 时间步长. 网络损失函数设置为整数损失与小数损失之和, 去小数点分支损失乘以超参数 η 作为损失函数正则化项, 整体损失函数如式 (2) 所示. 该设计的出发点是考虑到实际场景应用中小数部分会存在很大一部分全为零的情况, 网络存在过拟合风险. 训练相对复杂一点的去小数点分支可以起到正则化的作用, 且加入的超参数可训练, 根据验证集的反馈自适应学习, 实验阶段中我们建议的超参数值为 0.5.

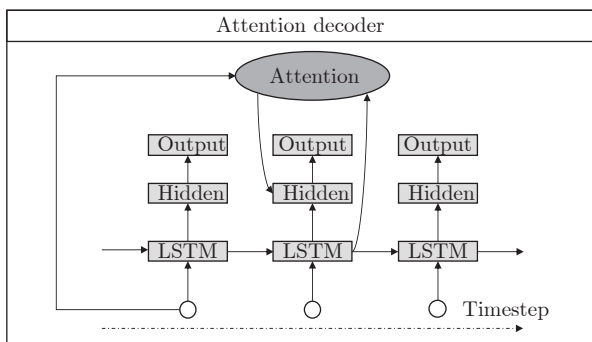


图 6 注意力机制网络解码流程图

Fig.6 Flowchart of decoder network based on attention

损失函数的改进^[15] 以及其他改进策略也可以一定程度上提高模型精度, 后续会考虑融入到我们的工作当中.

$$L = \frac{-1}{MN} \sum_{i=1}^M \sum_{j=1}^N y_{i,j} \ln(s_{i,j}) \quad (1)$$

$$L = L_{integer} + L_{decimal} + \eta L_{NDPB} \quad (2)$$

3 实验与分析

3.1 训练网络

为提高模型性能, 在训练网络之前, 需要对训练数据进行数据预处理操作. 本文采集的数据集来源于真实的货架图像, 数据丰富多样, 涵盖不同设计样式以及角度、光照的变化. 将数据归一化处理为相同的规格 (本文规格为 96×200), 并处理数据标签. 例如, 价格数据原标签为 79.99, 处理得到整数标签 79、小数标签 99 以及去小数点标签 7999.

预处理后的图像送入卷积块, 得到规格为 $12 \times 25 \times 512$ 的高层特征. 沿宽度方向切片 reshape 成 $25 \times 6 \times 144$ 的序列格式输入循环层. 循环层如前文所述, 由双向 LSTM 堆叠组成. 解码求得每个时间步的输出, 通过与标签计算交叉熵, 反馈训练网络. 对于我们的双分支网络, 网络的输出结果取决于两个分支的结果合并. 以去小数点分支与整数分支为例, 将去小数点分支结果沿着整数分支结果截断即得到小数部分, 通过小数点拼接输出完整价格.

关于模型的训练, 我们提供了一些训练策略来提升精度. 考虑到实际场景的条件影响, 增加饱和度和随机调整和随机旋转的数据增强策略, 可以很好地增强模型的泛化能力. 由于整体网络较深, 需要较大的学习率初始值加速网络收敛. 通过实验测试, 学习率初始值为 0.3 时, 伴随随机梯度下降策略效果最优.

3.2 实验结果分析

3.2.1 多任务结构分析

本文实验目的在于介绍多任务机制对于特殊结

构文本的贡献, 因此对于 Baseline 的选取, 我们只对前沿场景文本识别算法^[13]的主干结构进行实验分析, 而暂不考虑相关训练策略. 实验结果如表 1 所示, ResNet 作为卷积块, BiLSTM 作为循环层, 通过注意力机制解码的结构能够达到最高的精度.

表 1 模块的研究 (%)
Table 1 Study of modules (%)

Model	General-data	Hard-data
VGG-BiLSTM-CTC	50.20	20.20
VGG-BiLSTM-Attn	61.20	38.60
ResNet-BiLSTM-CTC	55.60	28.80
ResNet-BiLSTM-Attn	68.10	41.40

我们采用文献 [13] 中表现最优的模型作为 Baseline, 实验测试了基准方法并与我们的多任务分支进行比较分析. 根据价格牌结构的切分方式, 价格牌识别任务可以划分为: 去小数点的数据分支识别 (NDPB)、整数分支识别 (IB) 以及小数分支识别 (DB), 实验测试了多种分支组合方案, 精度结果如表 2 所示. 相较于在文本识别上的突出成绩, 基准模型很难在价格牌数据集中取得满意的成绩, 而本文提出的多任务模型则非常适用于价格牌这一特定场景, 为了体现出多分支结果的优点, 我们将基准方案与每个分支的输出进行了可视化分析, 图 7 给出了本方法是如何通过三支识别的方式规避了困难的小数点识别并通过各分支的结果推断出最终识别结果的机制. 实验结果表明, 不同双分支组合的结构相较于基准模型均取得较优的成绩, 这便验证了我们最初的信息拆分识别思路, 通过多任务的

方式独立地识别各分支是行之有效的且对最终的结果有促进作用. 整数分支与小数分支, 以及去小数点分支与小数点分支的多任务模型分别在普通测试集与困难测试集上取得了最优成绩, 这也是由于数据结构的最佳切分与相应多任务模型的组合. 进而在整数分支与小数分支的基础上以正则化的方式融入去小数点分支, 也让我们的多任务模型更进一步有所提升, 在普通测试集取得了 93.20 % 的最好成绩, 困难测试集上取得了 75.20 % 的最好成绩.

表 2 多任务模型结果 (%)
Table 2 Results of multitask model (%)

Model	General-data	Hard-data
Baseline ^[13]	68.10	41.40
NDPB&IB	90.10	72.90
NDPB&DB	91.70	74.30
IB&DB	92.20	73.20
NDPB&IB&DB	93.20	75.20

实验结果表明, 多任务机制可以充分有效地解决价格牌的识别问题. 在没有其他策略的优化下, 仅以多任务机制便可取得优异的成绩. 现阶段端对端的模型已成为深度学习主流, 而一些特殊的任务如价格牌中的小数点, 却很难用单任务的端对端模型解决. 因此, 我们提议从数据结构上分析, 以多任务结构联合进行分离式的识别是一个可行的解决方案.

3.2.2 模型分析

本文提出的价格牌识别网络有效地提高了价格牌的识别精度, 该方法也可以应用到其他 OCR 场



输入图像				
标签值	270.00	268.00	278.00	7.90
直接识别	2 700.0	26.800	218.00	7.90
无小数分支识别	27 000	26 800	27 800	790
整数识别	270	268	278	7
小数识别	00	00	00	90
组合结果	270.00	268.00	278.00	7.90

图 7 与直接识别方法的比较

Fig. 7 Compared with the single-branch method

景中, 为验证方法的迁移能力, 本文通过选取类似可拆分数据结构的车牌场景^[16], 验证所提方法的泛化性. 我们将论文所提出的方法在目前最大的车牌数据集 CCPD 中与效果优异的 TE2E^[17] 以及 CCPD^[16] 网络进行对比研究. 在测试中, 我们将车牌拆分成省、市和车牌号三个部分, 使用三分支结构进行识别. CCPD 测试集中包括各种复杂场景, 例如光线不均匀、角度倾斜以及雨雪天气等, 实验结果如表 3 所示. 本文所提出的方法均高于所对比的方法, 尤其在复杂场景的测试集中, 识别精度明显提升.

表 3 车牌数据集实验结果 (%)

Table 3 Experimental results on license plate dataset (%)

	DB	FN	Rotate	Tilt	Weather	Challenge
TE2E ^[17]	96.90	94.30	90.80	92.50	87.90	85.10
CCPD ^[16]	96.90	94.30	90.80	92.50	87.90	85.10
Our method	98.24	98.81	98.12	98.79	98.19	91.92

在车牌识别的应用中, 可以将汉字为切分点, 多分支结构分为汉字分支、数字字母分支以及完整车牌分支. 同样地, 完整车牌分支以正则化项的方式融入网络中, 防止其他分支训练过拟合. 表 3 中的实验结果表明, 对于车牌识别这一特定场景的任务, 本文的多任务模型性能明显高于原论文, 在所有测试集上, 精度均有所提升. 相比于在商业应用中的 TE2E 以及学术研究中的 CCPD, 个别测试集甚至取得了高达 10% 的提升, 这为多任务机制联合学习感受野的策略提供了强力依据. 本文提出的方法主要针对那些信息可拆分的且拆分部分具有独立性的图像文本, 比如价格牌的整数与小数部分, 车牌的汉字与字母数字部分. 实验结果表明本文提出的根据数据结构进行设计的多任务学习方法具有良好的性能, 且对于复杂场景的泛化性更强, 在价格牌数据集中的困难测试集以及车牌数据集中的各种复杂场景车牌集上都取得了很好的效果.

4 结论

本文针对新零售领域价格牌识别应用提出了基于多任务的价格牌识别网络, 针对特定场景图像文本的数据结构, 将整体数据分开处理, 通过先识别整数分支与小数分支替代识别完整价格, 最后添加小数点来解决小数点难以识别问题. 我们的网络采用卷积循环网络的结构, 以循环层注意力机制解码序列, 结合多任务学习机制, 用特定的领域知识联合学习难以定位的特征信息. 本文所提出的方法在我们开源的价格牌数据集上, 相比目前主流的文本识别算法有着明显的精度提升, 并且在类似数据结构的车牌数据集中也有非常好的效果. 我们的工作

目前只针对具有特定文本结构的图像, 对于通用文本的泛化性较差, 接下来的工作将会考虑多任务机制在通用文本上的可行性研究.

References

- 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. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2016, **39**(11): 2298–2304
- Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, et al. Attention is all you need. In: Proceedings of the Neural Information Processing Systems. San Diego, USA: MIT, 2017. 5998–6008
- Luong M T, Pham H, Manning C D. Effective approaches to attention-based neural machine translation [Online], available: <https://arxiv.org/abs/1508.04025>, Sep 20, 2015
- Li H, Wang P, Shen C. Towards end-to-end text spotting with convolutional recurrent neural networks. In: Proceedings of the IEEE International Conference on Computer Vision. Venice, Italy: IEEE, 2017. 5238–5246
- Yuan X, He P, Li X A. Adaptive adversarial attack on scene text recognition [Online], available: <http://export.arxiv.org/abs/1807.03326>, Jul 9, 2018
- Graves A, Fernández S, Gomez F, Schmidhuber J. Connectionist temporal classification: Labelling unsegmented sequence data with recurrent neural networks. In: Proceedings of the 23rd International Conference on Machine Learning. Pittsburgh Pennsylvania, USA: ACM, 2006. 369–376
- Sutskever I, Vinyals O, Le Q V. Sequence to sequence learning with neural networks. In: Proceedings of the Neural Information Processing Systems. Montréal, Canada: MIT, 2014. 3104–3112
- Lei Z, Zhao S, Song H, Shen J. Scene text recognition using residual convolutional recurrent neural network. *Machine Vision and Applications*, 29(5), 861–871
- Shi B, Yang M, Wang X, Lyu P, Yao C, Bai X. Aster: An attentional scene text recognizer with flexible rectification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, **41**(9), 2035–2048
- Long M, Wang J. Learning multiple tasks with deep relationship networks [Online], available: <https://arxiv.org/abs/1506.02117v1>, Jul 6, 2015
- Veit A, Matera T, Neumann L, Matas J, Belongie S. Coco-text: Dataset and benchmark for text detection and recognition in natural images [Online], available: <https://arxiv.org/abs/1601.07140v1>, Jan 26, 2016
- Karatzas D, Gomez-Bigorda L, Nicolaou A, Ghosh S, Bagdanov A, Iwamura M, Shafait F. ICDAR 2015 competition on robust reading. In: Proceedings of the 13th International Conference on Document Analysis and Recognition (ICDAR). Tunis, Tunisia: IEEE, 2015. 1156–1160
- Baek J, Kim G, Lee J, Park S, Han D. What is wrong with scene text recognition model comparisons? dataset and model analysis [Online], available: <https://arxiv.org/abs/1904.01906>, Dec 18, 2019
- Bingel J, Søgaard A. Identifying beneficial task relations for multi-task learning in deep neural networks [Online], available: <https://arxiv.org/abs/1702.08303>, Feb 27, 2017
- Xie Z, Huang Y, Zhu Y, Jin L, Liu Y, Xie L. Aggregation cross-entropy for sequence recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. Long Beach CA, USA: IEEE, 2019. 6538–6547
- Li H, Wang P, Shen C. Toward end-to-end car license plate detection and recognition with deep neural networks. *IEEE Transactions on Intelligent Transportation Systems*, 2018, **20**(3): 1126–1136

- 17 Xu Z, Yang W, Meng A, Lu N, Huang H, Ying C, Huang L. Towards end-to-end license plate detection and recognition: A large dataset and baseline. In: Proceedings of the European Conference on Computer Vision (ECCV). Munich, Germany: Springer, 2018. 255-271



牟永强 广州图匠数据科技有限公司首席 AI 架构师. 2012 年获得西安理工大学信号与信息处理专业硕士学位. 主要研究方向为机器视觉, 模式识别以及深度学习. 本文通信作者.

E-mail: yongqiang.mou@gmail.com

(**MOU Yong-Qiang** Chief AI architect at Guangzhou Image Data Technology Co., Ltd. He received his master degree in signal and information processing from Xi'an University of Technology in 2012. His research interest covers computer vision, pattern recognition, and deep learning. Corresponding author of this paper.)



范宝杰 广东工业大学硕士研究生. 主要研究方向为深度学习和计算机视觉.

E-mail: 735678367@qq.com

(**FAN Bao-Jie** Master student at Guangdong University of Technology. His research interest covers deep learning and computer vision.)



孙超 华南农业大学硕士研究生. 主要研究方向为深度学习和计算机视觉. E-mail: ice_moyan@163.com

(**SUN Chao** Master student at South China Agricultural University. His research interest covers deep learning and computer vision.)



严蕊 广州图匠数据科技有限公司高级研究员. 主要研究方向为深度学习和计算机视觉.

E-mail: reeyree@163.com

(**YAN Rui** Advanced Researcher at Guangzhou Image Data Technology Co., Ltd. His research interest

covers deep learning and computer vision.)



郭怡适 广州图匠数据科技有限公司首席执行官. 主要研究方向为深度学习和计算机视觉.

E-mail: yi.shi@imagedt.com

(**GUO Yi-Shi** Chief executive officer at Guangzhou Image Data Technology Co., Ltd. His research

interest covers deep learning and computer vision.)