自然场景下的文字识别

Challenge on Reading Chinese Text on Signboards

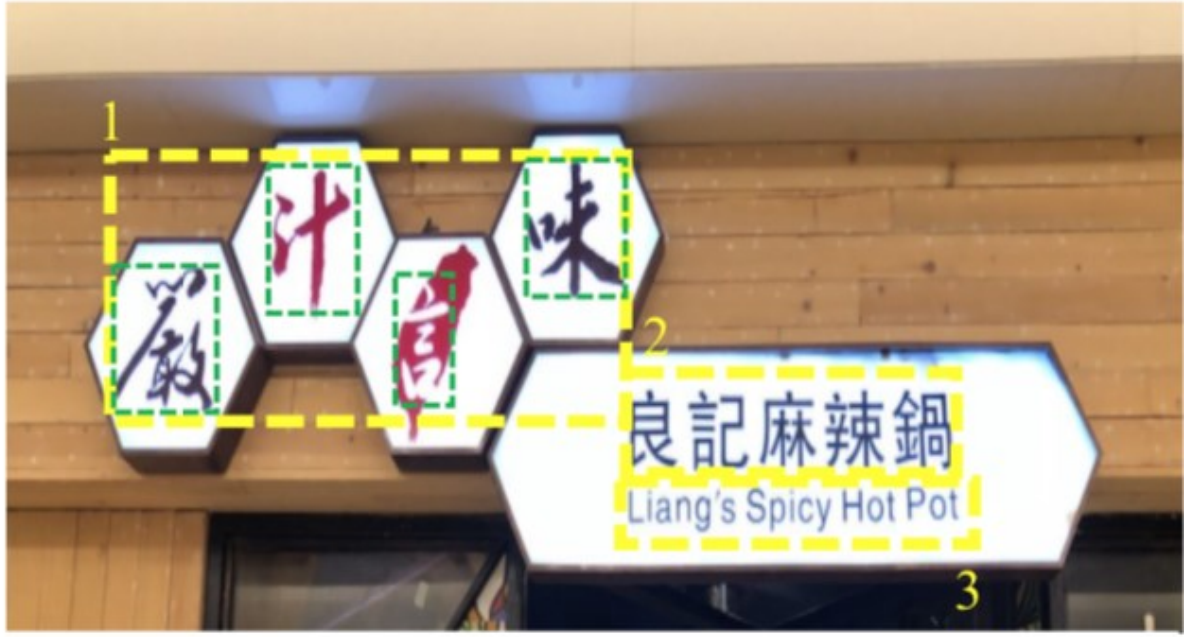
- ICDAR 2019将于今年9月20-25日在澳大利亚悉尼举办，其中有"中文门脸招牌文字识别"比赛（ICDAR 2019 Robust Reading Challenge on Reading Chinese Text on Signboards）。 http://rrc.cvc.uab.es/?ch=12
- 近年来业界围绕着文字检测和文字识别提出了许多有效的算法和技术方案。由于之前公开的数据集普遍以英文为主，因此所提出的技术方案对中文特有问题关注不足。表现在以中文为主的实际应用场景中，这些技术方案的结果与应用预期差距较大。
- 美团拥有由遍布全国的市场人员所拍摄的众多门脸招牌图片数据。每张图片都是由全国的不同个人，采用不同设备，在不同地点，不同时间和不同环境下所拍摄的不同目标，是难得的可以公正评价算法鲁棒性和识别效果的图片数据，挑战也非常大。
- 在此次 ICDAR2019上，美团挑选出很能代表中文特点的餐饮商家的门脸招牌图片来组织竞赛，这些招牌上的文字存在中文特有的设计和排版，同时也兼有自然场景文字识别中普遍存在的拍照角度、光照变化等干扰因素。希望通过竞赛引起同行们对中文识别的关注，群策群力解决中文识别的实际问题。

数据集介绍

- 数据由遍布全国的市场人员所拍摄的众多门脸招牌图片组成，共25000张。
- 每张图片是由完全独立的不同个人，采用不同设备，在不同地点，不同时间和不同环境下所拍摄的不同商家。
- 数据集以中文文字为主，也包含一定数量的英文和数字，英文和数字的占比介于 10% 和 30% 之间。
- 标注内容比较完备，每张图片均标注了单个字符的位置和文本，以及各字符串的位置和文本。是难得的用于研发和评估中文识别技术的数据集。
- 20000张图片用于训练，2000张用于验证，3000张用于测试。

数据集介绍 – Various Layout

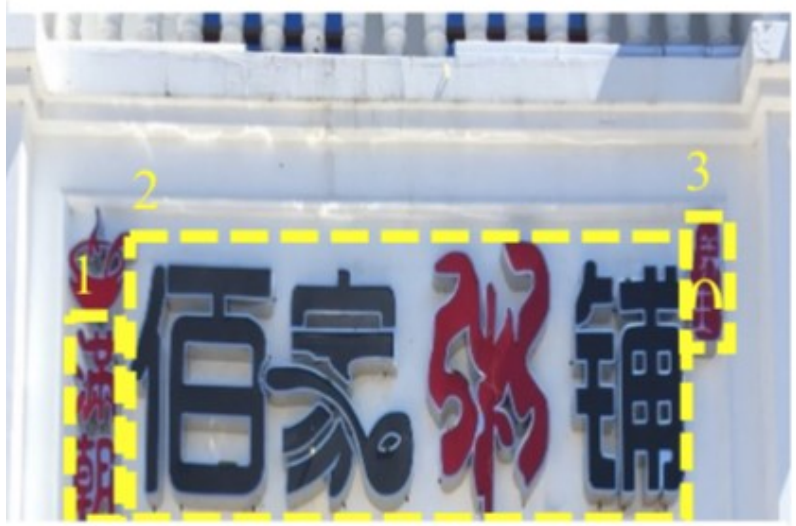
(1) Labeled text lines are “左左香，老潼关”, where the yellow boxes mean text lines, green boxes mean characters. The three characters in No. 1 box are arranged in triangles, the No.2 box is horizontal



(2) Labeled text lines are “厳汁言味，良記麻辣鍋，Liang’s Spicy Hot Pot”. The four characters in No. 1 box are arranged in broken line, the No.2 and No.3 box are both horizontal

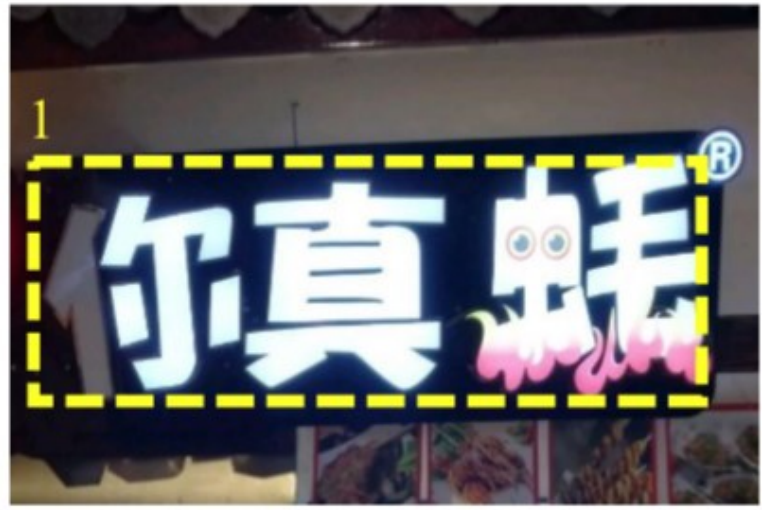
(a) Various layouts

数据集介绍 – Diverse Fonts



(3) Labeled text lines are “满朝，佰家粥铺，养生”.

(4) Labeled text lines are “靓串串，宵夜”.

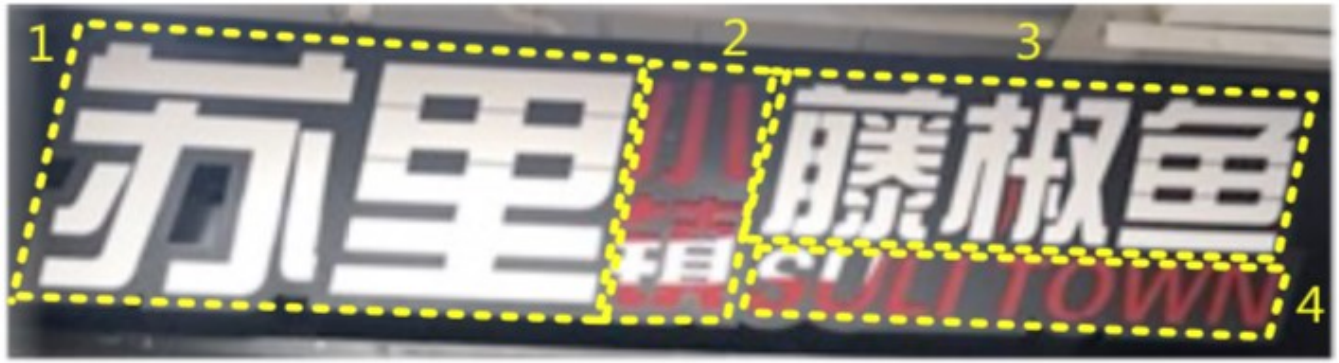


(5) Labeled text line is “你真蚝”.

(b) Diverse fonts

数据集介绍 – Various Orientation

(6) Labeled text lines are “精选猪脚，秘制卤水，正宗原味汤粉王，火红，潮汕隆江猪脚饭，订餐电话：13450994961，看似肥腻，吃而不腻”. The No.1,2,7,8 boxes are vertical, the No.3,4,5,6 boxes are



(7) Labeled text lines are “苏里，小镇，藤椒鱼，SULI TOWN”. The No.1,3,4 boxes are horizontal, the No.2 boxes is vertical

(c) Various orientations

比赛内容

- 本次门脸招牌识别，共定义了 4 个任务，分别如下：
  - TASK 1：招牌端到端文字识别
  - TASK 2：招牌文字行定位
  - TASK 3：招牌区域内单字识别
  - TASK 4：招牌区域内字符串识别

重要日期

- 2019年3月1日：报名通道开放
- 2019年3月18日：训练数据集开放
- 2019年4月15日：测试数据集分批开放
- 2019年4月16日：提交通道开放
- 2019年4月30日：提交截止日期
- 2019年5月10日：比赛最终报告提交
- 2019年9月20日：ICDAR 2019 大会召开

自然场景文字识别的难点

- 存在多种语言文本混合，字符可以有不同的大小、字体、颜色、亮度、对比度等。
- 文本行可能有横向、竖向、弯曲、旋转、扭曲等式样。
- 图像中的文字区域还可能会产生变形(透视、仿射变换)、残缺、模糊等现象。
- 自然场景图像的背景极其多样。如文字可以出现在平面、曲面或折皱面上或者非文字区域有近似文字的纹理，比如沙地、草丛、栅栏、砖墙等。

文本检测模型

视觉领域常规物体检测方法不理想的主要原因

- 相比于常规物体，文字行长度、长宽比例变化范围很大。
- 文本行是有方向性的。常规物体边框BBox的四元组描述方式信息量不充足。
- 自然场景中某些物体局部图像与字母形状相似，如果不参考图像全局信息将有误报。
- 有些艺术字体使用了弯曲的文本行，而手写字体变化模式也很多。
- 由于丰富的背景图像干扰，手工设计特征在自然场景文本识别任务中不够鲁棒。

解决方案

从各种角度对常规物体检测方法进行改造，极大提升了自然场景图像中文本检测的准确率：

- CTPN方案中，用BLSTM模块提取字符所在图像上下文特征，以提高文本块识别精度
- RRPN等方案中，文本框标注采用BBOX +方向角度值的形式，模型中产生出可旋转的文字区域候选框，并在边框回归计算过程中找到待测文本行的倾斜角度
- DMPNet等方案中，使用四边形（非矩形）标注文本框，来更紧凑的包围文本区域
- SegLink 将单词切割为更易检测的小文字块，再预测邻近连接将小文字块连成词
- TextBoxes等方案中，调整了文字区域参考框的长宽比例，并将特征层卷积核调整为长方形，从而更适合检测出细长型的文本行
- FTSN方案中，作者使用Mask-NMS代替传统BBOX的NMS算法来过滤候选框
- WordSup方案中，采用半监督学习策略，用单词级标注数据来训练字符级文本检测模型

CTPN模型

- CTPN是目前流传最广、影响最大的开源文本检测模型，可以检测水平或微斜的文本行。
  - Github（caffe版本）:https://github.com/tianzhi0549/CTPN
  - （tensorflow版本）:https://github.com/eragonruan/text-detection-ctpn
- 优点:

  对于检测的边框在上下左右4个点上都比较准确，这点比EAST要好。
- 缺点:

（1）CTPN只可以检测水平方向的文本，竖直方向的话就会出现一个字一个字断开的情况。倾斜角度的话需要修改后处理anchor的连接方式，但是应该会引入新的问题。

（2）CTPN由于涉及到anchor合并的问题，何时合并，何时断开，这是一个问题。所以对于双栏，三栏的这种文本，CTPN会都当做一个框处理，有时也会分开处理。

RRPN模型

- RRPN非常创新的提出了使用带角度的锚点处理场景文字检测中最常见的倾斜问题
- 与CTPN相比，RRPN采用的是旋转anchor，用于检测任意方向的文本行；而CTPN采用的是垂直anchor，用于检测水平方向的文本
- https://github.com/mjq11302010044/RRPN

FTSN模型

- FTSN（Fused Text Segmentation Networks）模型使用分割网络支持倾斜文本检测

DMPNet模型

- 使用四边形（非矩形）来更紧凑地标注文本区域边界，其训练出的模型对倾斜文本块检测效果更好。

EAST模型

- 根据开源工程中预训练模型的测试，该模型检测英文单词效果较好、检测中文长文本行效果欠佳
- 省略了其他模型中常见的区域建议、单词分割、子块合并等步骤，因此该模型的执行速度很快。
- https://github.com/argman/EAST

SegLink模型

- 优点：

（1）相比于CTPN等文本检测模型，SegLink的图片处理速度快很多

（2）与CTPN方法相比，SegLink引入了带方向的bbox，它可以检测任意方向的文本行

- 缺点：

（1）不能检测形变或者曲线文本，这是因为算法在合并的时候采用的是直线拟合．这里可以通过修改合并算法，来检测变形或曲线文本

（2）不能检测很大的文本，这是因为link主要是用于连接相邻的segments，而不能用于检测相距较远的文本行

- https://github.com/bgshih/seglink

PixelLink模型

- 与SegLink一样，不能检测很大的文本，这是因为link主要是用于连接相邻的segments，而不能用于检测相距较远的文本行
- https://github.com/ZJULearning/pixel_link

Textboxes/Textboxes++模型

- 调整了文字区域参考框的长宽比例，并将特征层卷积核调整为长方形，从而更适合检测出细长型的文本行。
- 与CTPN一样，TextBoxes在水平方向检测效果的好，因为其default box是水平框，回归的是水平矩形框
- 不足之处：对于曝光过度的地方并不能识别出文字，对于字符之间间距过大的单词识别效率也不高
- https://github.com/MhLiao/TextBoxes

WordSup模型

- WordSup提出了一种弱监督的训练框架，　可以文本行、单词级标注数据集上训练出字符级检测模型。

Shape Robust Text Detection with Progressive Scale Expansion Network

- arxiv: https://arxiv.org/abs/1806.02559
- github: https://github.com/whai362/PSENet

文本识别模型

CRNN模型

- CRNN(Convolutional Recurrent Neural Network）是目前较为流行的图文识别模型，可识别较长的文本序列
- https://github.com/bgshih/crnn
- 优势：

(1)可以端到端训练

(2)不需要进行字符分割和水平缩放操作，只需要垂直方向缩放到固定长度既可，同时可以识别任意长度的序列

(3)可以训练基于词典的模型和不基于词典的任意模型

(4)模型速度快，并且很小

RARE模型

- RARE（Robust text recognizer with Automatic Rectification）模型在识别变形的图像文本时效果很好
- RARE中支持一种称为TPS（thin-plate splines）的空间变换，从而能够比较准确地识别透视变换过的文本、以及弯曲的文本.

端到端模型

FOTS Rotation-Sensitive Regression

- 检测+识别一体化的框架，具有模型小，速度快，精度高，支持多角度的特点。
- 支持倾斜文本的识别
- Github：https://github.com/jiangxiluning/FOTS.PyTorch

STN-OCR模型

- Github地址： https://github.com/Bartzi/stn-ocr

MaskTextSpotter

- 该方法可以用于识别任意形状的文本，包括水平、多方向和曲线文本
- 相比于之前的方法，本文提出了通过语义分割来实现精确的文本检测和文本识别
- https://github.com/lvpengyuan/masktextspotter.caffe2

An end-to-end TextSpotter with Explicit Alignment and Attention

- https://github.com/tonghe90/textspotter

目前流行的解决方案

- 水平文字检测:
- 倾斜文字检测:
- 文字识别:

训练数据集

Chinese Text in the Wild(CTW)

- 该数据集包含32285张图像，1018402个中文字符(来自于腾讯街景), 包含平面文本，凸起文本，城市文本，农村文本，低亮度文本，远处文本，部分遮挡文本。图像大小2048*2048，数据集大小为31GB。以(8:1:1)的比例将数据集分为训练集(25887张图像，812872个汉字)，测试集(3269张图像，103519个汉字)，验证集(3129张图像，103519个汉字)。
- 文献链接：https://arxiv.org/pdf/1803.00085.pdf
- 数据集下载地址：https://ctwdataset.github.io/

Reading Chinese Text in the Wild(RCTW-17)

- 该数据集包含12263张图像，训练集8034张，测试集4229张，共11.4GB。大部分图像由手机相机拍摄，含有少量的屏幕截图，图像中包含中文文本与少量英文文本。图像分辨率大小不等。
- 下载地址 http://mclab.eic.hust.edu.cn/icdar2017chinese/dataset.html
- 文献：http://arxiv.org/pdf/1708.09585v2

ICPR MWI 2018 挑战赛

- 大赛提供20000张图像作为数据集，其中50%作为训练集，50%作为测试集。主要由合成图像，产品描述，网络广告构成。该数据集数据量充分，中英文混合，涵盖数十种字体，字体大小不一，多种版式，背景复杂。文件大小为2GB。
- 下载地址：

https://tianchi.aliyun.com/competition/information.htm?raceId=231651&_is_login_redirect=true&accounttraceid=595a06c3-7530-4b8a-ad3d-40165e22dbfe

Total-Text

- 该数据集共1555张图像，11459文本行，包含水平文本，倾斜文本，弯曲文本。文件大小441MB。大部分为英文文本，少量中文文本。训练集：1255张 测试集：300
- 下载地址：

http://www.cs-chan.com/source/ICDAR2017/totaltext.zip

- 文献： http:// arxiv.org/pdf/1710.10400v

Google FSNS(谷歌街景文本数据集)

- 该数据集是从谷歌法国街景图片上获得的一百多万张街道名字标志，每一张包含同一街道标志牌的不同视角，图像大小为600*150，训练集1044868张，验证集16150张，测试集20404张。
- 下载地址：http://rrc.cvc.uab.es/?ch=6&com=downloads
- 文献：http:// arxiv.org/pdf/1702.03970v1

COCO-TEXT

- 该数据集，包括63686幅图像，173589个文本实例，包括手写版和打印版，清晰版和非清晰版。文件大小12.58GB，训练集：43686张，测试集：10000张，验证集：10000张
- 文献: http://arxiv.org/pdf/1601.07140v2
- 下载地址： https://vision.cornell.edu/se3/coco-text-2/

Synthetic Data for Text Localisation

- 在复杂背景下人工合成的自然场景文本数据。包含858750张图像，共7266866个单词实例，28971487个字符，文件大小为41GB。该合成算法，不需要人工标注就可知道文字的label信息和位置信息，可得到大量自然场景文本标注数据。
- 下载地址：http://www.robots.ox.ac.uk/~vgg/data/scenetext/
- 文献：http://www.robots.ox.ac.uk/~ankush/textloc.pdf
- Code:

https://github.com/ankush-me/SynthText (英文版)

https://github.com/wang-tf/Chinese_OCR_synthetic_data(中文版)

Synthetic Word Dataset

- 合成文本识别数据集，包含9百万张图像，涵盖了9万个英语单词。文件大小为10GB
- 下载地址：http://www.robots.ox.ac.uk/~vgg/data/text/

Caffe-OCR中文合成数据

- 数据利用中文语料库，通过字体、大小、灰度、模糊、透视、拉伸等变化随机生成，共360万张图片，图像分辨率为280x32，涵盖了汉字、标点、英文、数字共5990个字符。文件大小约为8.6GB
- 下载地址：https://pan.baidu.com/s/1dFda6R3

相关补充资料

- https://github.com/handong1587/handong1587.github.io/blob/master/_posts/deep_learning/2015-10-09-ocr.md
- https://zhuanlan.zhihu.com/p/52335619
- https://github.com/Jyouhou/SceneTextPapers
- ALPR:
- 链接： https://pan.baidu.com/s/1Dk1SuZlgbS79PjBOMuE_lQ 密码: zlve