知乎 首发于 SIGAI 关注专栏

自然场景文本检测识别技术综述

SIGAI
已认证的官方帐号

157 人赞同了该文章

本文及其它机器学习、深度学习算法的全面系统讲解可以阅读《机器学习与应用》，清华大学出版社，雷明著，由SIGAI公众号作者倾力打造，自2019年1月出版以来已重印3次。

- 书的购买链接
- 书的勘误，优化，源代码资源

> 番外
> 青蛇: 姐, 图像文本检测和识别领域现在的研究热点是什么?
> 白蛇: 白纸黑字的扫描文档识别技术已经很成熟，而自然场景图像文本识别的效果还不理想。倾斜字、艺术字、变形字、模糊字、形似字、残缺字、光影遮蔽、多语言混合文本等应用落地面临的技术难题还没被彻底解决。
> 青蛇: 文本检测模型CTPN中为什么选用VGG16作基础网络?
> 白蛇: CTPN是2016年被推出的, 而VGG16是那年很流行的特征提取基础网络。如果今年实施文本检测, 可以试试Resnet、FCN、Densenet等后起之秀作基础网络，或许有惊喜。

知乎 首发于 SIGAI 关注专栏

摘要

本文介绍图像文本识别（OCR）领域的最新技术进展。首先介绍应用背景，包括面临的技术挑战、典型应用场景、系统实施框架等。接着介绍搭建图文识别模型过程中经常被引用到的多种特征提取基础网络、物体检测网络框架，以及它们被应用到图文识别任务中所面临的场景适配问题。然后介绍最近三年来出现的各种文本边框检测模型、文字内容识别模型、端到端图文识别模型。最后介绍图文识别领域的大型公开数据集。

应用概述

OCR（Optical Character Recognition, 光学字符识别）传统上指对输入扫描文档图像进行分析处理，识别出图像中文字信息。场景文字识别（Scene Text Recognition，STR） 指识别自然场景图片中的文字信息。自然场景图像中的文字识别，其难度远大于扫描文档图像中的文字识别，因为它的文字展现形式极其丰富：

- 允许多种语言文本混合，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；文字区域附近有复杂的干扰纹理

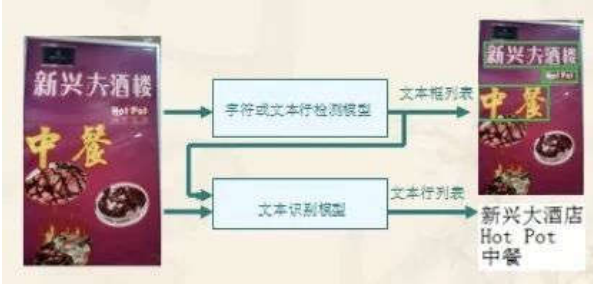
（本图摘自新浪微博《光学字符识别技术：让电脑像人一样阅读》）

也有人用OCR技术泛指所有图像文字检测和识别技术， 包括传统OCR技术与场景文字识别技术。这是因为，场景文字识别技术可以被看成是传统OCR技术的自然演进与升级换代。

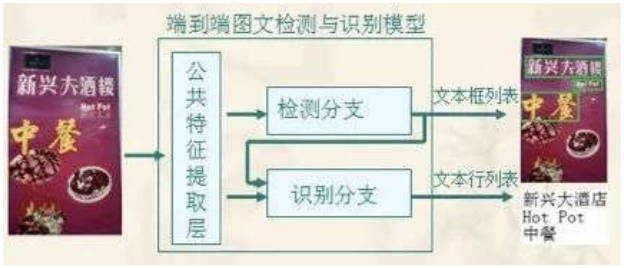
图像文字检测和识别技术有着广泛的应用场景。已经被互联网公司落地的相关应用涉及了识别名片、识别菜单、识别快递单、识别身份证、识别营业证、识别银行卡、识别车牌、识别路牌、识别商品包装袋、识别会议白板、识别广告主干词、识别试卷、识别单据等等。

已经有不少服务商在提供图像文字检测和识别服务，这些服务商既包括了腾讯、百度、阿里、微软、亚马逊、谷歌等大型云服务企业，也包括了一些活跃在物流、教育、安防、视频直播、电子政务、直接

如下图所示，传统技术解决方案中，是先分别训练文字检测和文本识别两个模型，然后在服务实施阶段将这两个模型串联到数据流水线中组成图文识别系统。



如下图所示，最近流行的技术解决方案中，是用一个多目标网络直接训练出一个端到端的模型。在训练阶段，该模型的输入是训练图像及图中文本坐标、文本内容，模型优化目标是输出端边框坐标预测误差与文本内容预测误差的加权和。在服务实施阶段，原始图片流过该模型直接输出预测文本信息。相比于传统方案，该方案中模型训练效率更高、服务运营阶段资源开销更少。



文本检测和识别技术处于一个学科交叉点，其技术演进不断受益于计算机视觉处理和自然语言处理两个领域的技术进步。它既需要使用视觉处理技术来提取图像中文字区域的图像特征向量，又需要借助自然语言处理技术来解码图像特征向量为文字结果。

模型基础

从公开论文中可以看到，起源于图像分类、检测、语义分割等视觉处理任务的各个基础网络（backbone network），纷纷被征用来提取图像中文字区域的特征向量。同时，起源于物体检测、语义分割任务的多个网络框架，也被改造后用于提升图文识别任务中的准确率和执行速度。本章将简单温习一下这些基础网络、网络框架的实现原理，并介绍图文识别任务中应用它们时所面临的各种场景适配问题。

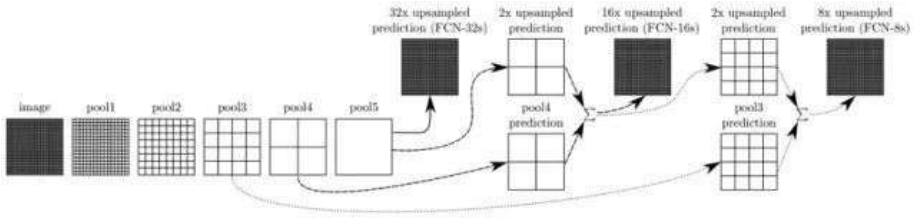
基础网络

图文识别任务中充当特征提取模块的基础网络，可以来源于通用场景的图像分类模型。例如，VGGNet，ResNet、InceptionNet、DenseNet、Inside-Outside Net、Se-Net等。

图文识别任务中的基础网络，也可以来源于特定场景的专用网络模型。例如，擅长提取图像细节特征的FCN网络，擅长做图形矫正的STN网络。

由于大家对通用网络模型已经很熟悉，所以本节只简单介绍上述专用网络模型。

采样（upsampling）操作，将特征矩阵恢复到接近原图尺寸，然后对每一个位置上的像素做类别预测，从而能识别出更清晰的物体边界。基于FCN的检测网络，不再经过候选区域回归出物体边框，而是根据高分辨率的特征图直接预测物体边框。因为不需要像Faster-RCNN那样在训练前定义好候选框长宽比例，FCN在预测不规则物体边界时更加鲁棒。由于FCN网络最后一层特征图的像素分辨率较高，而图文识别任务中需要依赖清晰的文字笔画来区分不同字符（特别是汉字），所以FCN网络很适合用来提取文本特征。当FCN被用于图文识别任务时，最后一层特征图中每个像素将被分成文字行（前景）和非文字行（背景）两个类别。



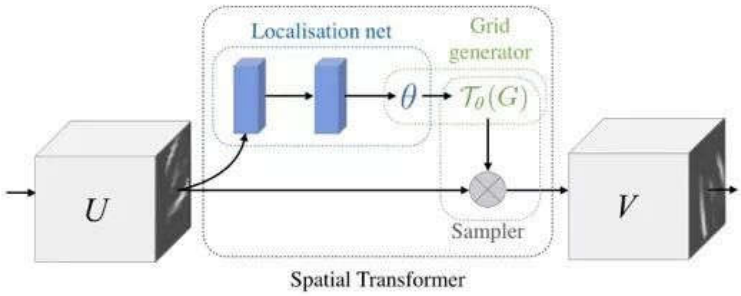
（选自arXiv:1411.4038，‘ Fully Convolutional Networks for Semantic Segmentation’）

STN网络

空间变换网络（STN，Spatial Transformer Networks）的作用是对输入特征图进行空间位置矫正得到输出特征图，这个矫正过程是可以进行梯度传导的，从而能够支持端到端的模型训练。

如下图所示，STN网络由定位网络（Localization Network），网格生成器（Grid generator），采样器（Sampler）共3个部分组成。定位网络根据原始特征图U计算出一套控制参数，网格生成器这套控制参数产生采样网格（sampling grid），采样器根据采样网格核函数将原始图U中像素对应采样到目标图V中。

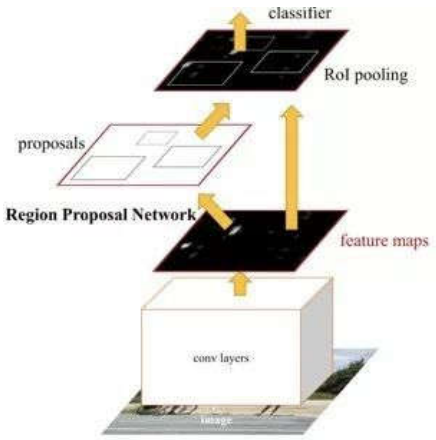
空间变换的控制参数是根据原始特征图U动态生成的，生成空间变换控制参数的元参数则是在模型训练阶段学习到的、并且存放于定位网络的权重（weights）矩阵中。



（选自arXiv: 1506.02025，‘ Spatial Transformer Networks’）

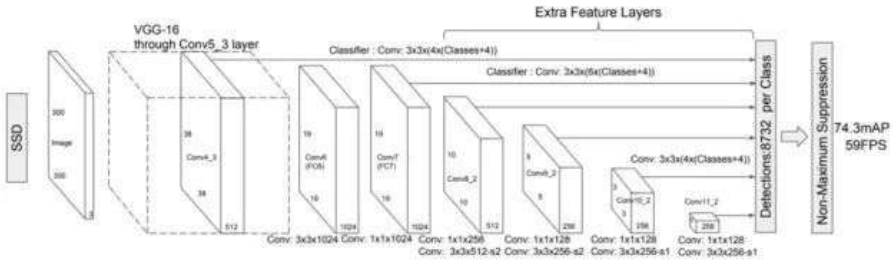
检测网络框架

Faster RCNN作为一个检测网络框架，其目标是寻找紧凑包围被检测对象的边框（BBOX，Bounding Box）。如下图所示，它在Fast RCNN检测框架基础上引入区域建议网络（RPN，Region Proposal Network），来快速产生与目标物体长宽比例接近的多个候选区域参考框（anchor）；它通过ROI（Region of Interest） Pooling层为多种尺寸参考框产生出归一化固定尺寸的区域特征；它利用共享的CNN卷积网络同时向上述RPN网络和ROI Pooling层输入特征映射（Feature Maps），从而减少卷积层参数量和计算量。训练过程中使用到了多目标损失函数，包括RPN网络、ROI Pooling层的边框分类loss和坐标回归loss，通过这些loss的梯度反向传



(摘自arXiv:1506.01497，‘ Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks’)

SSD（Single Shot MultiBox Detector），是2016年提出的一种全卷积目标检测算法，截止到目前仍是主要的目标检测框架之一，相比Faster RCNN有着明显的速度优势。如下图所示，SSD是一种one stage算法，直接预测被检测对象的边框和得分。检测过程中，SSD算法利用多尺度思想进行检测，在不同尺度的特征图(feature maps)上产生与目标物体长宽比例接近的多个默认框(Default boxes)，进行回归与分类。最后利用非极大值抑制(Non-maximum suppression)得到最终的检测结果。训练过程中，SSD采用Hard negative mining策略进行训练，使正负样本比例保持为1：3，同时使用多种数据增广(Data augmentation)方式进行训练，提高模型性能。



(摘自arxiv: 1512.02325， “SSD: Single Shot MultiBox Detector”)

文本检测模型

文本检测模型的目标是从图片中尽可能准确地找出文字所在区域。

但是，视觉领域常规物体检测方法(SSD, YOLO, Faster-RCNN等)直接套用于文字检测任务效果并不理想， 主要原因如下：

- 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。

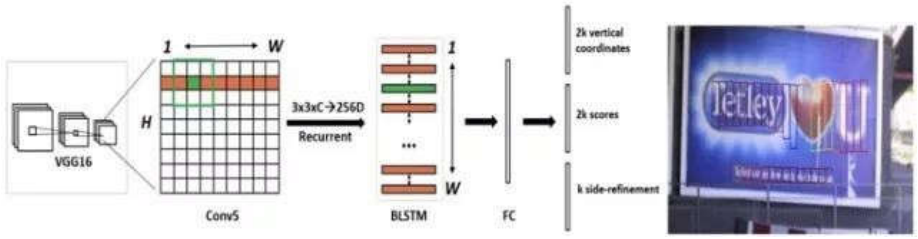
针对上述问题根因，近年来出现了各种基于深度学习的技术解决方案。它们从特征提取、区域建议网络(RPN)、多目标协同训练、Loss改进、非极大值抑制（NMS）、半监督学习等角度对常规物

·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
·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。

下面用近年来出现的多个模型案例，介绍如何应用上述各方法提升图像文本检测的效果。

CTPN模型

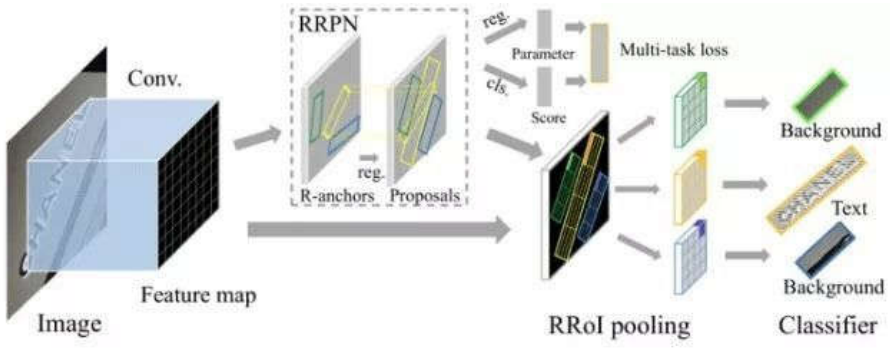
CTPN是目前流传最广、影响最大的开源文本检测模型，可以检测水平或微斜的文本行。文本行可以被看成一个字符sequence，而不是一般物体检测中单个独立的目标。同一文本行上各个字符图像间可以互为上下文，在训练阶段让检测模型学习图像中蕴含的这种上下文统计规律，可以使得预测阶段有效提升文本块预测准确率。CTPN模型的图像预测流程中，前端使用当时流行的VGG16做基础网络来提取各字符的局部图像特征，中间使用BLSTM层提取字符序列上下文特征，然后通过FC全连接层，末端经过预测分支输出各个文字块的坐标值和分类结果概率值。在数据后处理阶段，将合并相邻的小文字块为文本行。



（选自arXiv: 1609.03605，‘Detecting Text in Natural Image with Connectionist Text Proposal Network’）

RRPN模型

基于旋转区域候选网络（RRPN, Rotation Region Proposal Networks）的方案，将旋转因素并入经典区域候选网络（如Faster RCNN）。这种方案中，一个文本区域的ground truth被表示为具有5元组(x,y,h,w,θ)的旋转边框, 坐标(x,y)表示边框的几何中心, 高度h设定为边框的短边，宽度w为长边，方向是长边的方向。训练时，首先生成含有文本方向角的倾斜候选框，然后在边框回归过程中学习文本方向角。



（选自arXiv: 1703.01086，‘Arbitrary-Oriented Scene Text Detection via Rotation Proposals’）

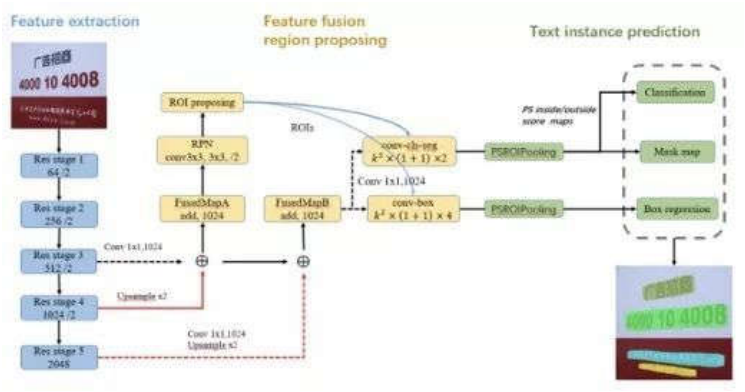
RRPN中方案中提出了旋转感兴趣区域（RRoI，Rotation Region-of-Interest）池化层，将任意方向的区域建议先划分成子区域，然后对这些子区域分别做max pooling，并将结果投影到具有固

(a) divide arbitrary-oriented proposal into subregions; (b) max pooling from inclined proposals to RROIs.

（选自arXiv: 1703.01086，‘Arbitrary-Oriented Scene Text Detection via Rotation Proposals’）

FTSN模型

FTSN（Fused Text Segmentation Networks）模型使用分割网络支持倾斜文本检测。它使用Resnet-101做基础网络，使用了多尺度融合的特征图。标注数据包括文本实例的像素掩码和边框，使用像素预测与边框检测多目标联合训练。



（选自arXiv: 1709.03272，‘Fused Text Segmentation Networks for Multi-oriented Scene Text Detection’）

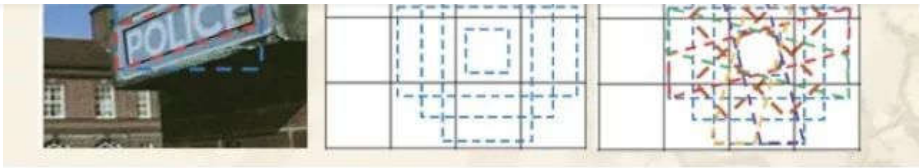
基于文本实例间像素级重合度的Mask-NMS， 替代了传统基于水平边框间重合度的NMS算法。下图左边子图是传统NMS算法执行结果，中间白色边框被错误地抑制掉了。下图右边子图是Mask-NMS算法执行结果， 三个边框都被成功保留下来。

（选自arXiv: 1709.03272，‘Fused Text Segmentation Networks for Multi-oriented Scene Text Detection’）

DMPNet模型

DMPNet（Deep Matching Prior Network）中，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。

如下图所示，它使用滑动窗口在特征图上获取文本区域候选框，候选框既有正方形的、也有倾斜四边形的。接着，使用基于像素点采样的Monte-Carlo方法，来快速计算四边形候选框与标注框间的面积重合度。然后，计算四个顶点坐标到四边形中心点的距离，将它们与标注值相比计算出目标



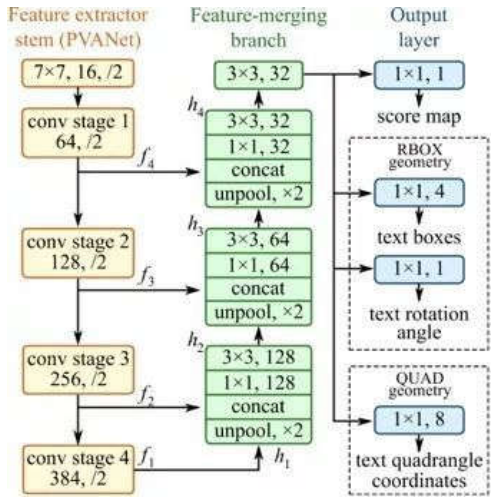
（选自arXiv:1703.01425，’ Deep Matching Prior Network: Toward Tighter Multi-oriented Text Detection’ ）

EAST模型

EAST（Efficient and Accuracy Scene Text detection pipeline）模型中，首先使用全卷积网络（FCN）生成多尺度融合的特征图，然后在此基础上直接进行像素级的文本块预测。该模型中，支持旋转矩形框、任意四边形两种文本区域标注形式。对应于四边形标注，模型执行时会对特征图中每个像素预测其到四个顶点的坐标差值。对应于旋转矩形框标注，模型执行时会对特征图中每个像素预测其到矩形框四边的距离、以及矩形框的方向角。

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。或许，根据中文数据特点进行针对性训练后，检测效果还有提升空间。

上述过程中，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。



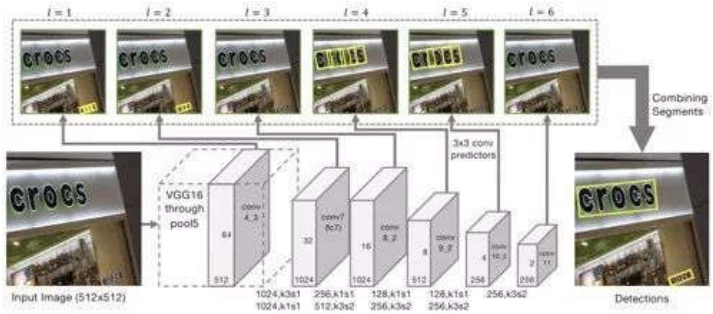
（选自arXiv: 1704.03155，’ EAST: An Efficient and Accurate Scene Text Detector’ ）

SegLink模型

SegLink模型的标注数据中，先将每个单词切割为更易检测的有方向的小文字块（segment），然后用邻近连接（link ）将各个小文字块连接成单词。这种方案方便于识别长度变化范围很大的、带方向的单词和文本行，它不会象Faster-RCNN等方案因为候选框长宽比例原因检测不出长文本行。相比于CTPN等文本检测模型，SegLink的图片处理速度快很多。

（选自arXiv: 1703.06520，‘ Detecting Oriented Text in Natural Images by Linking Segments’ ）

如下图所示，该模型能够同时从6种尺度的特征图中检测小文字块。同一层特征图、或者相邻层特征图上的小文字块都有可能被连接入同一个单词中。换句话说，位置邻近、并且尺寸接近的文字块都有可能被预测到同一单词中。



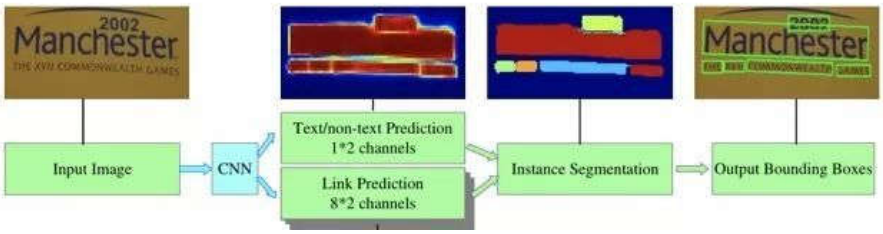
（选自arXiv: 1703.06520，‘ Detecting Oriented Text in Natural Images by Linking Segments’ ）

PixelLink模型

自然场景图像中一组文字块经常紧挨在一起，通过语义分割方法很难将它们识别开来，所以PixelLink模型尝试用实例分割方法解决这个问题。

该模型的特征提取部分，为VGG16基础上构建的FCN网络。模型执行流程如下图所示。首先，借助于CNN 模块执行两个像素级预测：一个文本二分类预测，一个链接二分类预测。接着，用正链接去连接邻居正文本像素，得到文字块实例分割结果。然后，由分割结果直接就获得文字块边框，而且允许生成倾斜边框。

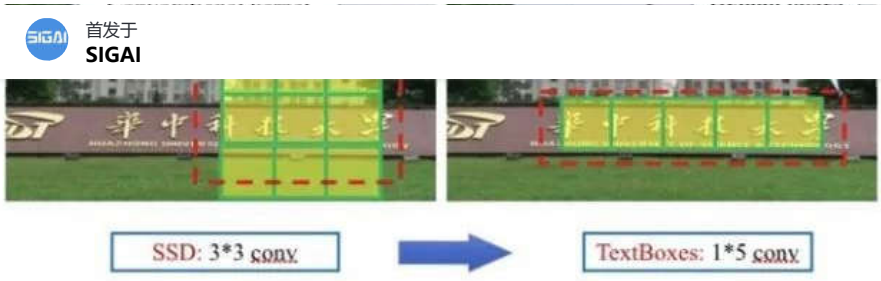
上述过程中，省掉了其他模型中常见的边框回归步骤，因此训练收敛速度更快些。训练阶段，使用了平衡策略，使得每个文字块在总LOSS中的权值相同。训练过程中，通过预处理增加了各种方向角度的文字块实例。



（选自arXiv: 1801.01315，‘ Detecting Scene Text via Instance Segmentation’ ）

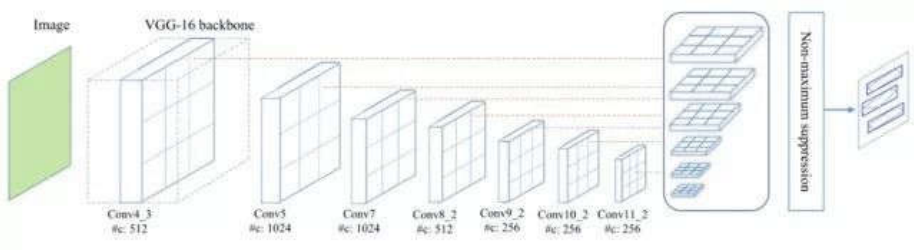
Textboxes/Textboxes++模型

Textboxes是基于SSD框架的图文检测模型，训练方式是端到端的，运行速度也较快。如下图所示，为了适应文字行细长型的特点，候选框的长宽比增加了1,2,3,5,7,10这样初始值。为了适应文本行细长型特点，特征层也用长条形卷积核代替了其他模型中常见的正方形卷积核。为了防止漏检文本行，还在垂直方向增加了候选框数量。为了检测大小不同的字符块，在多个尺度的特征图上并行预测文本框， 然后对预测结果做NMS过滤。



（选自arXiv: 1611.06779，‘TextBoxes: A Fast Text Detector with a Single Deep Neural Network’）

Textboxes++是Textboxes的升级版本，目的是增加对倾斜文本的支持。为此，将标注数据改为了旋转矩形框和不规则四边形的格式；对候选框的长宽比例、特征图层卷积核的形状都作了相应调整。

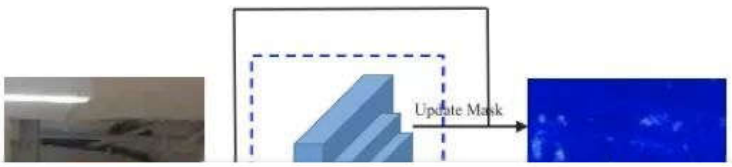


（选自arXiv: 1801.02765，‘TextBoxes++: A Single-Shot Oriented Scene Text Detector’）

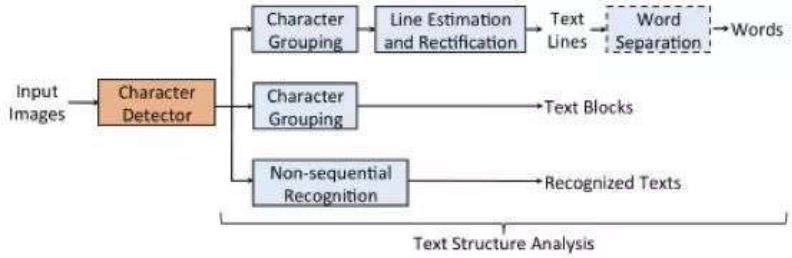
WordSup模型

如下图所示，在数学公式图文识别、不规则形变文本行识别等应用中，字符级检测模型是一个关键基础模块。由于字符级自然场景图文标注成本很高、相关公开数据集稀少，导致现在多数图文检测模型只能在文本行、单词级标注数据上做训练。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。

如下图所示，WordSup弱监督训练框架中，两个训练步骤被交替执行：给定当前字符检测模型，并结合单词级标注数据，计算出字符中心点掩码图； 给定字符中心点掩码图，有监督地训练字符级检测模型.



如下图，训练好字符检测器后，可以在数据流水线中加入合适的文本结构分析模块，以输出符合应用场景格式要求的文本内容。该文作者例举了多种文本结构分析模块的实现方法。



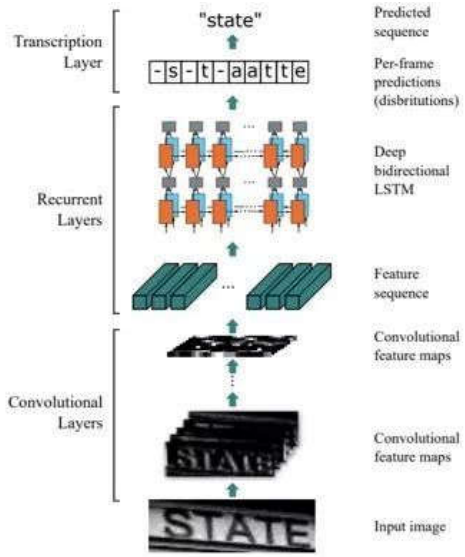
（选自arXiv: 1708.06720，‘WordSup：Exploiting Word Annotations for Character based Text Detection’）

文本识别模型

文本识别模型的目标是从已分割出的文字区域中识别出文本内容。

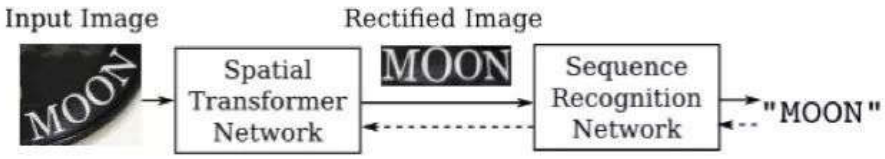
CRNN模型

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。它包含CNN特征提取层和BLSTM序列特征提取层，能够进行端到端的联合训练。它利用BLSTM和CTC部件学习字符图像中的上下文关系， 从而有效提升文本识别准确率，使得模型更加鲁棒。预测过程中，前端使用标准的CNN网络提取文本图像的特征，利用BLSTM将特征向量进行融合以提取字符序列的上下文特征，然后得到每列特征的概率分布，最后通过转录层(CTC rule)进行预测得到文本序列。

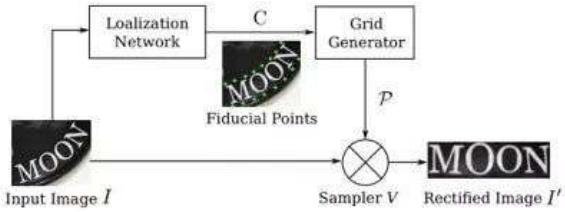


（选自arXiv: 1507.05717，‘An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition’）

RARE模型



如下图所示，空间变换网络内部包含定位网络、网格生成器、采样器三个部件。经过训练后，它可以根据输入图像的特征图动态地产生空间变换网格，然后采样器根据变换网格核函数从原始图像中采样获得一个矩形的文本图像。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.



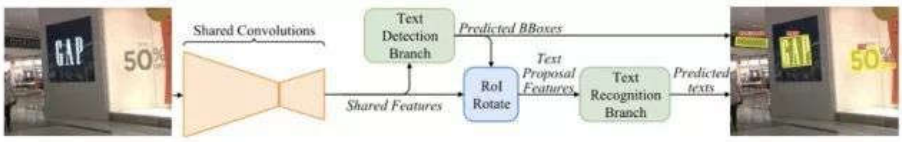
（选自arXiv: 1603.03915，‘ Robust Scene Text Recognition with Automatic Rectification’ ）

端到端模型

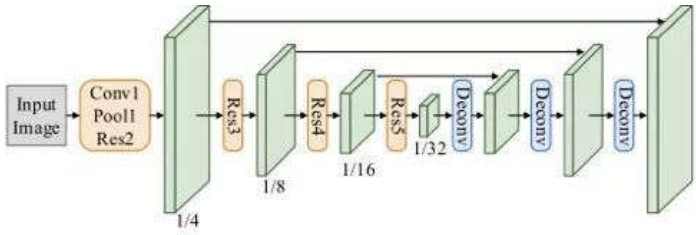
端到端模型的目标是一站式直接从图片中定位和识别出所有文本内容来。

FOTS Rotation-Sensitive Regression

FOTS（Fast Oriented Text Spotting）是图像文本检测与识别同步训练、端到端可学习的网络模型。检测和识别任务共享卷积特征层，既节省了计算时间，也比两阶段训练方式学习到更多图像特征。引入了旋转感兴趣区域（RoIRotate），可以从卷积特征图中产生出定向的文本区域，从而支持倾斜文本的识别.



Overall architecture. The network predicts both text regions and text labels in a single forward pass.

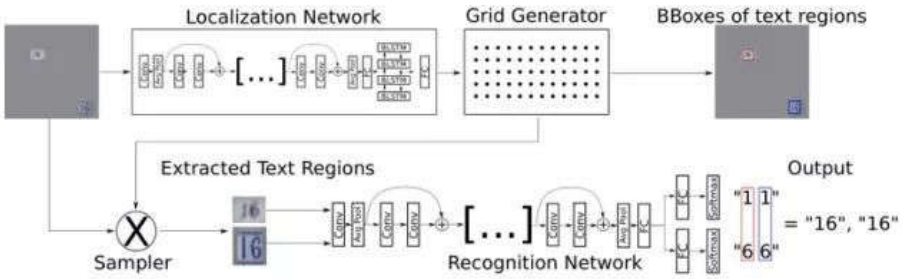


Architecture of shared convolutions.

（选自arXiv: 1801.01671，‘ FOTS: Fast Oriented Text Spotting with a Unified Network’ ）

STN-OCR模型

的识别精度。在训练上STN-OCR属于半监督学习方法，只需要提供文本内容标注，而不要求文本定位信息。作者也提到，如果从头开始训练则网络收敛速度较慢，因此建议渐进地增加训练难度。STN-OCR已经开放了工程源代码和预训练模型。



(选自arXiv: 1707.08831，‘ STN-OCR: A single Neural Network for Text Detection and Text Recognition’)

训练数据集

本章将列举可用于文本检测和识别领域模型训练的一些大型公开数据集， 不涉及仅用于模型fine-tune任务的小型数据集。

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/

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_redirec

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

Curve multi-oriented horizontal

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

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

Synthetic Word Dataset

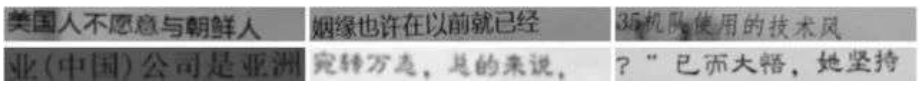
合成文本识别数据集，包含9百万张图像，涵盖了9万个英语单词。文件大小为10GB

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



参考文献

1. “光学字符识别技术：让电脑像人一样阅读”，新浪微博, 霍强

arxiv.org/pdf/1411.4038

3. “Spatial Transformer Networks”，arXiv:1506.02025，Max Jaderberg, Karen Simonyan, Andrew Zisserman, Koray Kavukcuoglu

arxiv.org/pdf/1506.0202...

4. “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks”，arXiv:1506.01497，Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun

arxiv.org/pdf/1506.0149...

5. “SSD: Single Shot MultiBox Detector”，arxiv:1512.02325，Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg

arxiv.org/pdf/1512.0232...

6. “Detecting Text in Natural Image with Connectionist Text Proposal Network”，arXiv:1609.03605，Zhi Tian, Weilin Huang, Tong He, Pan He, Yu Qiao

arxiv.org/pdf/1609.0360...

7. “Arbitrary-Oriented Scene Text Detection via Rotation Proposals”，arXiv:1703.01086，Jianqi Ma, Weiyuan Shao, Hao Ye, Li Wang, Hong Wang, Yingbin Zheng, Xiangyang Xue

arxiv.org/pdf/1703.0108...

8. “Fused Text Segmentation Networks for Multi-oriented Scene Text Detection”，arXiv:1709.03272，Yuchen Dai, Zheng Huang, Yuting Gao, Youxuan Xu, Kai Chen, Jie Guo, Weidong Qiu

arxiv.org/pdf/1709.0327...

9. “Deep Matching Prior Network: Toward Tighter Multi-oriented Text Detection”，arXiv:1703.01425，Yuliang Liu, Lianwen Jin

arxiv.org/pdf/1703.0142...

10. “EAST: An Efficient and Accurate Scene Text Detector”，arXiv:1704.03155，Xinyu Zhou, Cong Yao, He Wen, Yuzhi Wang, Shuchang Zhou, Weiran He, Jiajun Liang

arxiv.org/pdf/1704.0315...

11. “Detecting Oriented Text in Natural Images by Linking Segments”，arXiv:1703.06520，Baoguang Shi, Xiang Bai, Serge Belongie

arxiv.org/pdf/1703.0652...

arxiv.org/pdf/1801.0131...

13. “TextBoxes: A Fast Text Detector with a Single Deep Neural Network”，arXiv:1611.06779，Minghui Liao, Baoguang Shi, Xiang Bai, Xinggang Wang, Wenyu Liu

arxiv.org/pdf/1611.0677...

14. “TextBoxes++: A Single-Shot Oriented Scene Text Detector”，arXiv:1801.02765，Minghui Liao, Baoguang Shi, Xiang Bai

arxiv.org/pdf/1801.0276...

15. “WordSup： Exploiting Word Annotations for Character based Text Detection”，arXiv:1708.06720，Han Hu, Chengquan Zhang, Yuxuan Luo, Yuzhuo Wang, Junyu Han, Errui Ding

arxiv.org/pdf/1708.0672...

16. “An End-to-End Trainable Neural Network for Image-based Sequence Recognition and Its Application to Scene Text Recognition”，arXiv:1507.05717，Baoguang Shi, Xiang Bai, Cong Yao

arxiv.org/pdf/1507.0571...

17. “Robust Scene Text Recognition with Automatic Rectification”，arXiv:1603.03915，Baoguang Shi, Xinggang Wang, Pengyuan Lyu, Cong Yao, Xiang Bai

arxiv.org/pdf/1603.0391...

18. “FOTS: Fast Oriented Text Spotting with a Unified Network”，arXiv:1801.01671，Xuebo Liu, Ding Liang, Shi Yan, Dagui Chen, Yu Qiao, Junjie Yan

arxiv.org/pdf/1801.0167...

19. “STN-OCR: A single Neural Network for Text Detection and Text Recognition”，arXiv:1707.08831，Christian Bartz, Haojin Yang, Christoph Meinel

arxiv.org/pdf/1707.0883...

20. “Chinese Text in the Wild”，arXiv:1803.00085，Tai-Ling Yuan, Zhe Zhu, Kun Xu, Cheng-Jun Li, Shi-Min Hu

arxiv.org/pdf/1803.0008...

21. “ICDAR2017 Competition on Reading Chinese Text in the Wild (RCTW-17)”，arXiv:1708.09585，Baoguang Shi, Cong Yao, Minghui Liao, Mingkun Yang, Pei Xu, Linyan Cui, Serge Belongie, Shijian Lu, Xiang Bai

arxiv.org/pdf/1710.1040...

23. “End-to-End Interpretation of the French Street Name Signs Dataset”，arXiv:1702.03970，Raymond Smith, Chunhui Gu, Dar-Shyang Lee, Huiyi Hu, Ranjith Unnikrishnan, Julian Ibarz, Sacha Arnoud, Sophia Lin

arxiv.org/pdf/1702.0397...

24. “COCO-Text: Dataset and Benchmark for Text Detection and Recognition in Natural Images”，arXiv:1601.07140，Andreas Veit, Tomas Matera, Lukas Neumann, Jiri Matas, Serge Belongie

arxiv.org/pdf/1601.0714...

25. “Synthetic Data for Text Localisation in Natural Images”，arXiv:1604.06646，Ankush Gupta， Andrea Vedaldi， Andrew Zisserman

arxiv.org/pdf/1604.0664...

推荐文章

[1] 机器学习-波澜壮阔40年 SIGAI 2018.4.13.

[2] 学好机器学习需要哪些数学知识? SIGAI 2018.4.17.

[3] 人脸识别算法演化史 SIGAI 2018.4.20.

[4] 基于深度学习的目标检测算法综述 SIGAI 2018.4.24.

[5] 卷积神经网络为什么能够称霸计算机视觉领域? SIGAI 2018.4.26.

[6] 用一张图理解SVM的脉络 SIGAI 2018.4.28.

[7] 人脸检测算法综述 SIGAI 2018.5.3.

[8] 理解神经网络的激活函数 SIGAI 2018.5.5.

[9] 深度卷积神经网络演化历史及结构改进脉络-40页长文全面解读 SIGAI 2018.5.8.

[10] 理解梯度下降法 SIGAI 2018.5.11.

[11] 循环神经网络综述—语音识别与自然语言处理的利器 SIGAI 2018.5.15

[12] 理解凸优化 SIGAI 2018.5.18

[13]【实验】理解SVM的核函数和参数 SIGAI 2018.5.22

[14]【SIGAI综述】行人检测算法 SIGAI 2018.5.25

[17]【群话题精华】5月集锦—机器学习和深度学习中一些值得思考的问题 SIGAI 2018.6.1

[18] 大话Adaboost算法 SIGAI 2018.6.2

[19] FlowNet到FlowNet2.0：基于卷积神经网络的光流预测算法 SIGAI 2018.6.4

[20] 理解主成分分析(PCA) SIGAI 2018.6.6

[21] 人体骨骼关键点检测综述 SIGAI 2018.6.8

[22] 理解决策树 SIGAI 2018.6.11

[23] 用一句话总结常用的机器学习算法 SIGAI 2018.6.13

[24] 目标检测算法之YOLO SIGAI 2018.6.15

[25] 理解过拟合 SIGAI 2018.6.18

[26] 理解计算：从√2到AlphaGo ——第1季 从√2谈起 SIGAI 2018.6.20

[27] 场景文本检测——CTPN算法介绍 SIGAI 2018.6.22

[28] 卷积神经网络的压缩和加速 SIGAI 2018.6.25

[29]k近邻算法SIGAI 2018.6.27

编辑于 2019-03-29

科技

文章被以下专栏收录

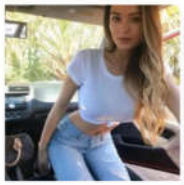
SIGAI
专注于AI技术研究与机器学习框架研发，让AI所见即所得

进入专栏

推荐阅读

图像质量评价之结构相似性SSIM（中）

在上一篇文章中，我们介绍了对图像质量进行评价的必要性、主观评价和客观评价的两种标准，以及设计符合人类直觉的评价标准的困难性和重要性。 本来这篇文章想把我



一个22万张NSFW图片的鉴黄数据集？我有个大胆的想法……

视觉多目标跟踪算法综述（上）-附开源代码下载链接…

白翔：趣谈“捕文景文字检测 | VA

深度学习大讲堂致力能，深度学习方面的品以及活动。请关注栏！ 编者按 字测的图像和 有性和更概括的表达力

程程

知乎 首发于 SIGAI 关注专栏

8 条评论 切换为时间排序

写下你的评论...

做自己 1 年前
赞
赞

龙塔路口 1 年前
很全面啊
3

一只萌新 1 年前
提到了许多最新的技术，看论文很有帮助
赞

韩熙枭雄 1 年前
有OCR数据链接，很赞了
赞

爱媳妇的好男人 1 年前
文章对自然场景下的文本检测进行了详细的分析，之前做过印刷体的数字识别系统，用BP神经网络实现，接触到深度学习后一直想多了解一下这方面的知识，文中对多种网络模型进行了详细的介绍，还分享了好多公开数据集，是一篇良心好文！👍
赞

杨子文 1 年前
很全面很清晰的综述，可以当做自然场景字符识别领域的一个目录，如果能够收录一些对应的开源代码就更好了
赞

piginzoo 6 个月前
良心之作，必须顶下！

赞

知乎用户 3 个月前
好文！看来19年句子识别最强还是15年的crnn了，没有更好的或者变种吗？
1

▲ 赞同 157 ▼ 8 条评论 分享 收藏 ···