OCR

前言

HOG 特征：全称 Histogram of Oriented Gradient（方向梯度直方图），由图像的局部区域梯度方向直方图构成特征； LBP 特征：全称 Local Binary Pattern（局部二值模式），通过比较中心与邻域像素灰度值构成图像局部纹理特征； Haar 特征：描述图像的灰度变化，由各模块的像素差值构成特征； 核函数（Kernels）：从低维空间到高维空间的映射，把低维空间中线性不可分的两类点变成线性可分的；

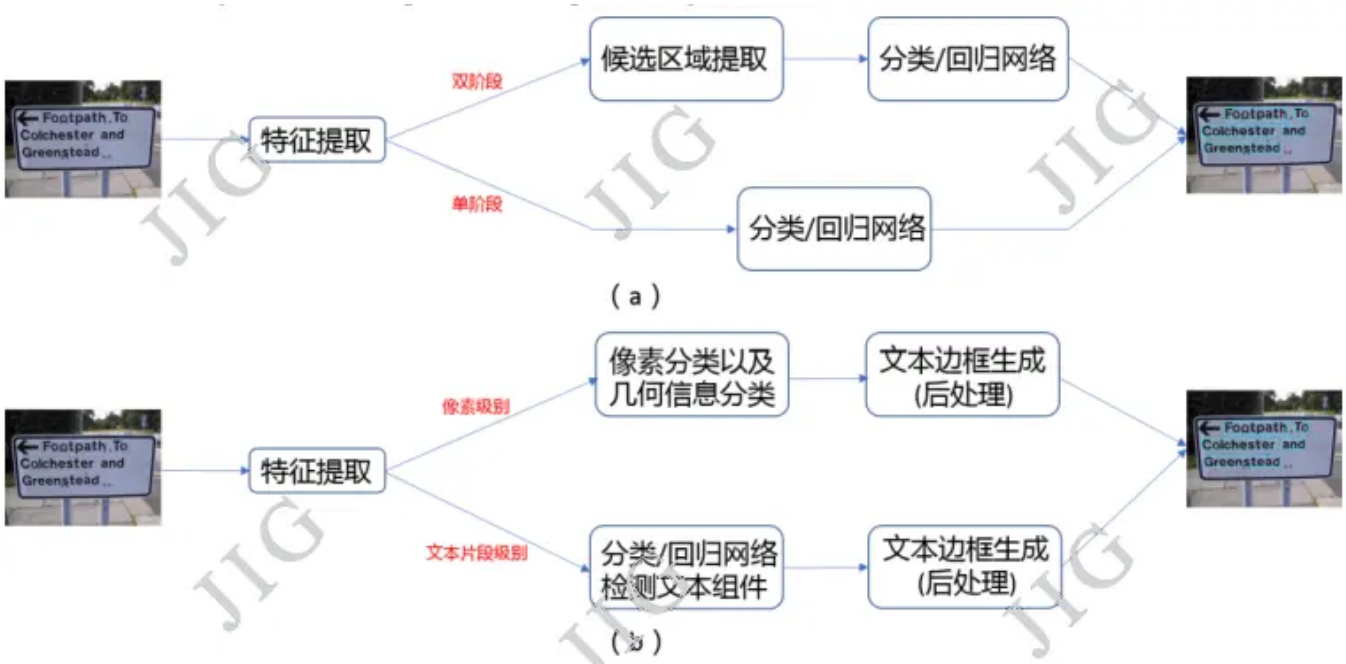
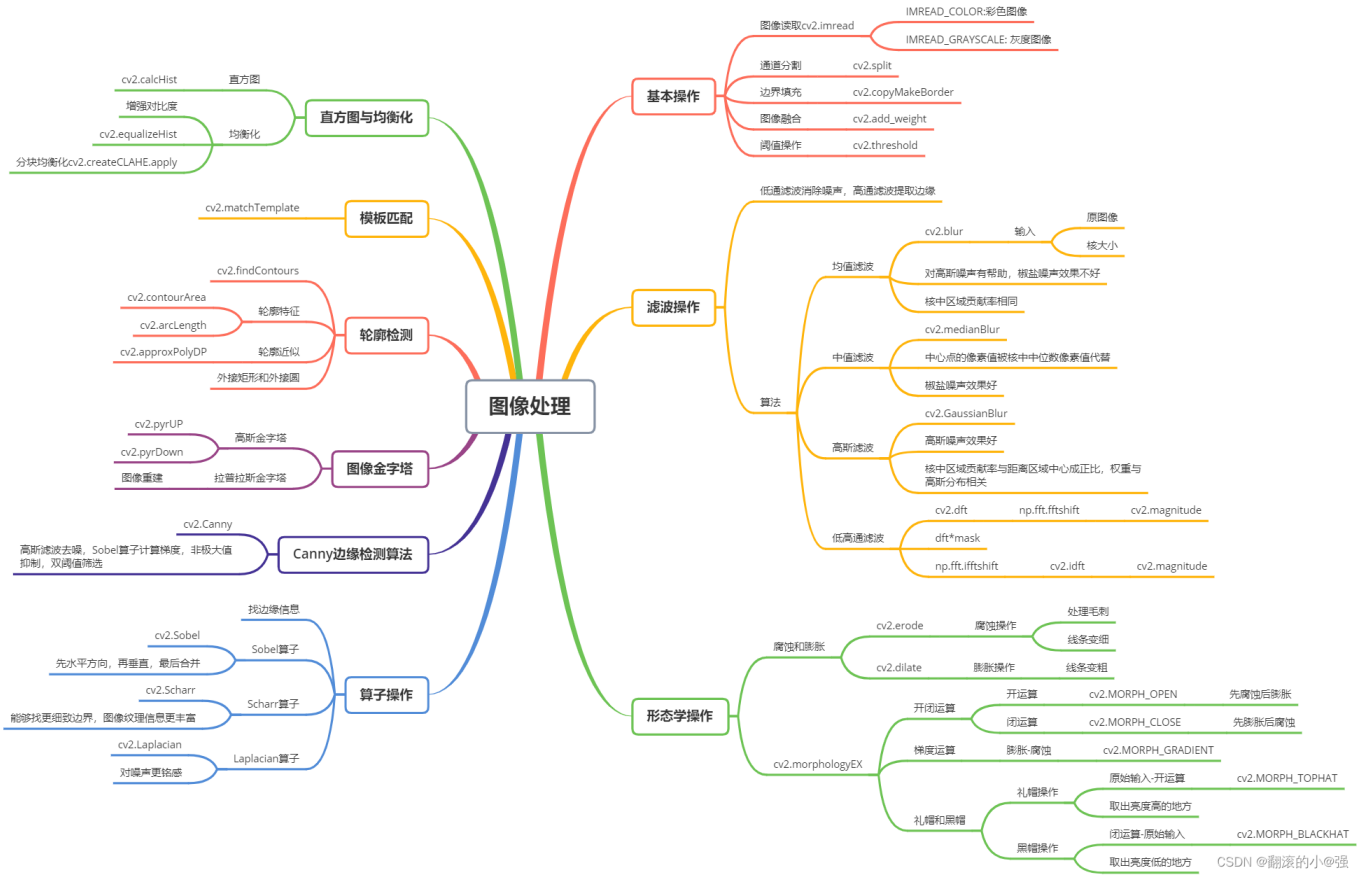


图 2 自然场景文本检测方法流程图。（a）自顶向下；（b）自底向上

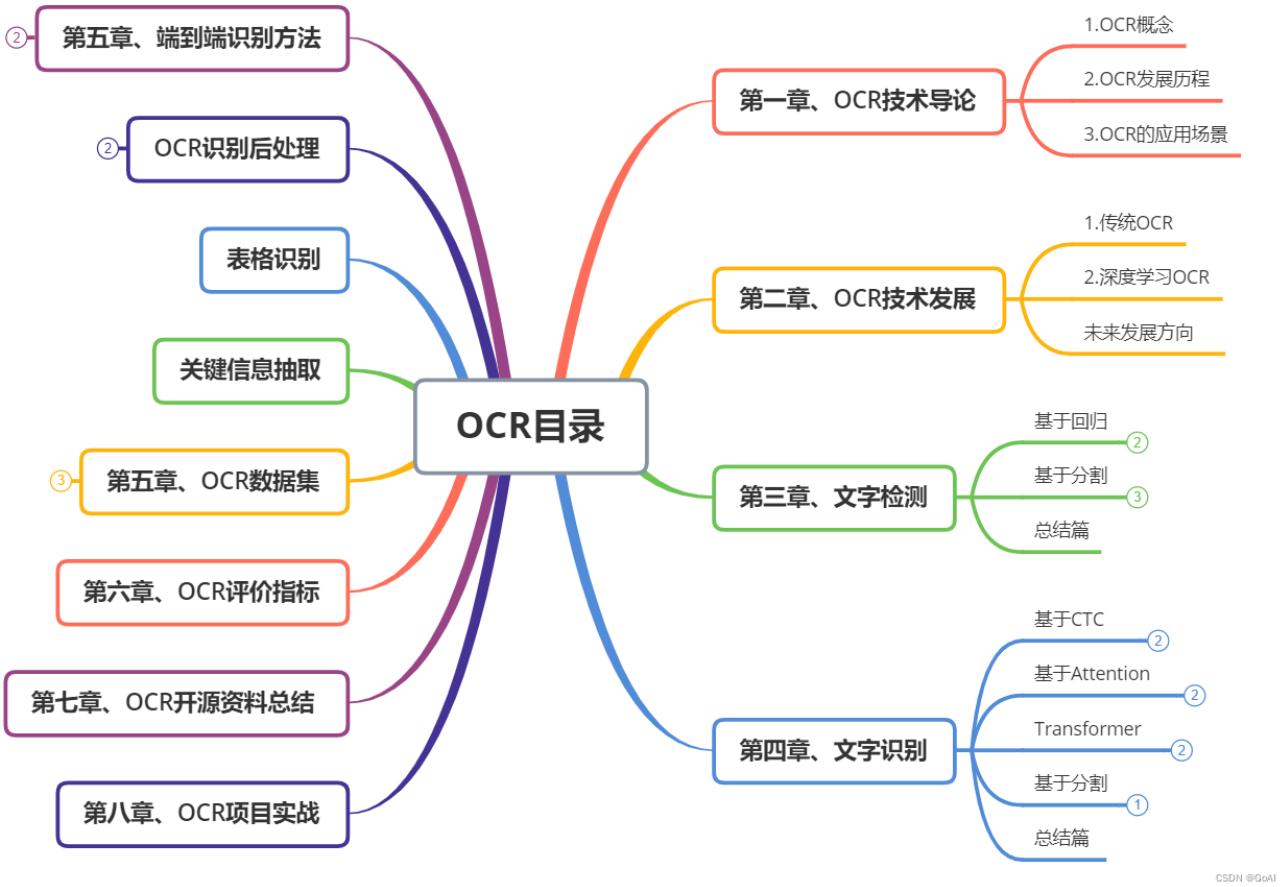
图像处理导图


图像处理
基本操作
图像读取cv2.imread
IMREAD_COLOR:彩色图像
IMREAD_GRAYSCALE: 灰度图像
通道分割
cv2.split
边界填充
cv2.copyMakeBorder
图像融合
cv2.add_weight
阈值操作
cv2.threshold
滤波操作
低通滤波消除噪声，高通滤波提取边缘
算法
均值滤波
cv2.blur
输入
原图像
核大小
对高斯噪声有帮助，椒盐噪声效果不好
核中区域贡献率相同
中值滤波
cv2.medianBlur
中心点的像素值被核中中位数像素值代替
椒盐噪声效果好
高斯滤波
cv2.GaussianBlur
高斯噪声效果好
核中区域贡献率与距离区域中心成正比，权重与高斯分布相关
低高通滤波
cv2.dft
np.fft.fftshift
cv2.magnitude
dft*mask
np.fft.ifftshift
cv2.idft
cv2.magnitude
形态学操作
腐蚀和膨胀
cv2.erode
腐蚀操作
处理毛刺
线条变细
cv2.dilate
膨胀操作
线条变粗
cv2.morphologyEX
开闭运算
开运算
cv2.MORPH_OPEN
先腐蚀后膨胀
闭运算
cv2.MORPH_CLOSE
先膨胀后腐蚀
梯度运算
膨胀-腐蚀
cv2.MORPH_GRADIENT
礼帽和黑帽
礼帽操作
原始输入-开运算
cv2.MORPH_TOPHAT
取出亮度高的地方
黑帽操作
闭运算-原始输入
cv2.MORPH_BLACKHAT
取出亮度低的地方
直方图与均衡化
直方图
cv2.calcHist
均衡化
增强对比度
cv2.equalizeHist
分块均衡化cv2.createCLAHE.apply
模板匹配
cv2.matchTemplate
轮廓检测
cv2.findContours
轮廓特征
cv2.contourArea
cv2.arcLength
轮廓近似
cv2.approxPolyDP
外接矩形和外接圆
图像金字塔
高斯金字塔
cv2.pyrUP
cv2.pyrDown
拉普拉斯金字塔
图像重建
Canny边缘检测算法
cv2.Canny
高斯滤波去噪，Sobel算子计算梯度，非极大值抑制，双阈值筛选
算子操作
找边缘信息
Sobel算子
cv2.Sobel
先水平方向，再垂直，最后合并
Scharr算子
cv2.Scharr
能够找更细致边界，图像纹理信息更丰富
Laplacian算子
cv2.Laplacian
对噪声更敏感
CSDN @翻滚的小@强

Opencv 思维导图


opencv思维导图
图像处理
平滑
膨胀腐蚀
开运算和闭运算
礼帽变换和黑帽变换
漫水填充
图像金字塔
阈值化和自适应阈值化
图像变换
卷积，梯度和sobel导数
Scharr滤波器
拉普拉斯变换，x，y和二阶导和
canny算子
霍夫变换
重映射
仿射变换2x3和透视变换3x3
Cart，Polar，LogPolar变换
离散傅立叶变换和离散余弦变换
积分图像，距离变换和直方图均衡化
直方图与匹配
反向投影
模板匹配
图像局部与分割
背景减除
平均背景法
背景学习模型
分水岭算法
Inpainting修补图像
Delaunay三角划分和Voronoi划分
轮廓
序列
freeman链码
查找轮廓
多边形逼近和边界框
矩，中心矩和归一化矩
等级匹配和轮廓树
凸包和凸缺陷
成对几何直方图
机器学习
Mahalanobis距离
K均值
朴素贝叶斯分类
二叉决策树
boosting
随机森林
人脸识别和Haar分类器
期望最大化EM
K近邻
神经网络
支持向量机SVM
投影三维视觉
鸟瞰图变换
POSIT
立体成像
来自运动的结构
二维三维直线拟合
跟踪与运动
寻找角点
亚像素级角点
不变特征SIFT
光流
Lucas-Kannade算法
Hom-Schunck方法
块匹配方法
mean-shift和camshift
运动模块
预估器卡尔曼滤波
condensation算法与粒子滤波
摄像机模型与标定
投影与畸变
标定，旋转和平移
矫正
罗德里格斯变换

目录



OCR 入门教程系列（一）：OCR 基础导论

OCR 发展

文档文字识别：可以将图书馆、报社、博物馆、档案馆等的纸质版图书、报纸、杂志、历史文献档案资料等进行电子化管理，实现精准地保存文献资料。

自然场景文字识别：识别自然场景图像中的文字信息如车牌、广告干词、路牌等信息。对车辆进行识别可以实现停车场收费管理、交通流量控制指标测量、车辆定位、防盗、高速公路超速自动化监管等功能。

票据文字识别：可以对增值税发票、报销单、车票等不同格式的票据进行文字识别，可以避免财务人员手动输入大量票据信息，如今已广泛应用于财务管理、银行、金融等众多领域。。

证件识别：可以快速识别身份证、银行卡、驾驶证等卡证类信息，将证件文字信息直接转换为可编辑文本，可以大大提高工作效率、减少人工成本、还可以实时进行相关人员的身份核验，以便安全管理。

OCR 生态

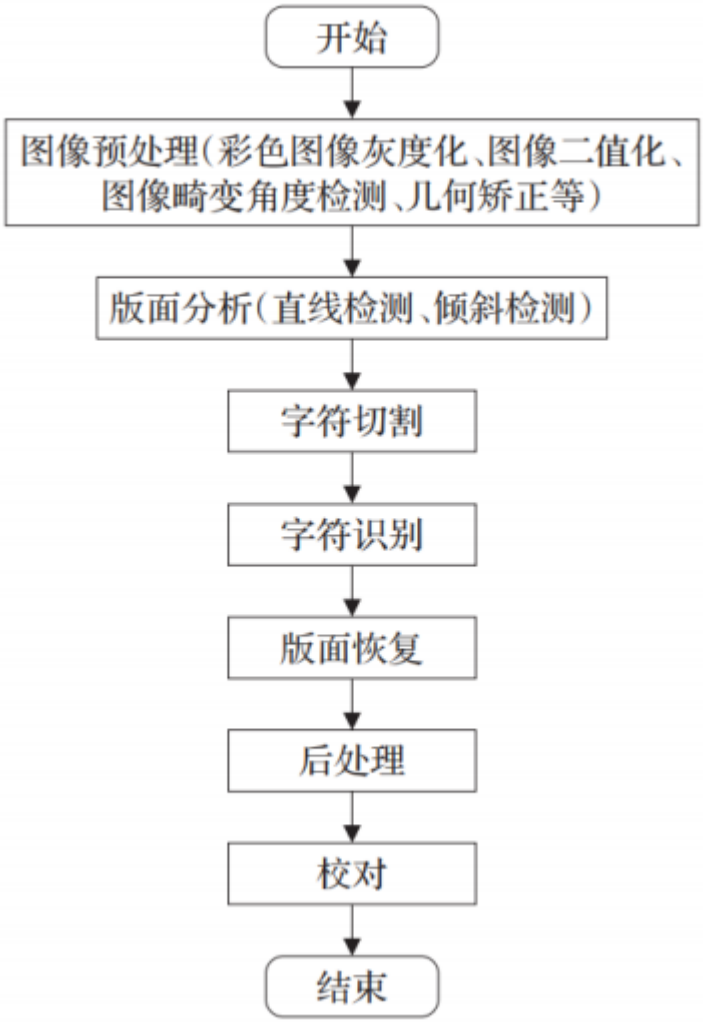
OCR 的技术路线

其中 OCR 识别的关键路径在于文字检测和文本识别部分，这也是深度学习技术可以充分发挥功效的地方。

传统 OCR 技术流程

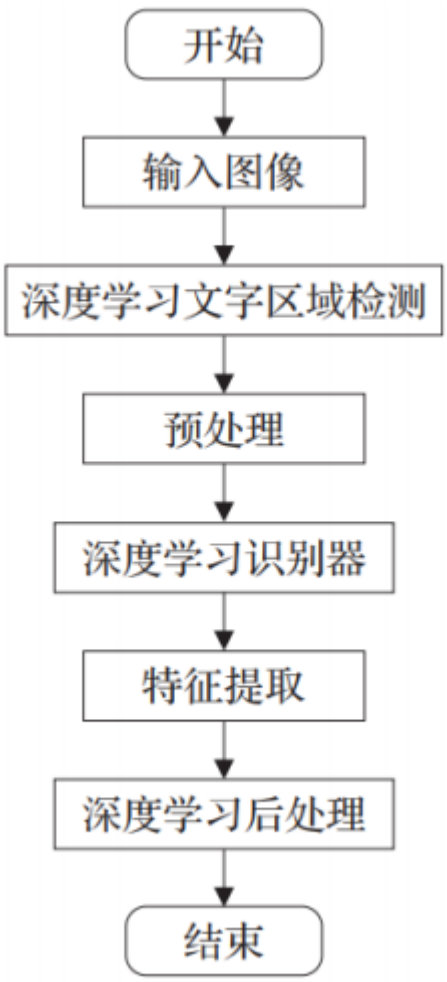
1、水平投影垂直投影 2、模板匹配 3、查找轮廓 findcontours

传统的光学字符识别过程为：图像预处理（彩色图像灰度化、二值化处理、图像变化角度检测、矫正处理等）、版面划分（直线检测、倾斜检测）、字符定位切分、字符识别、版面恢复、后处理、校对等。

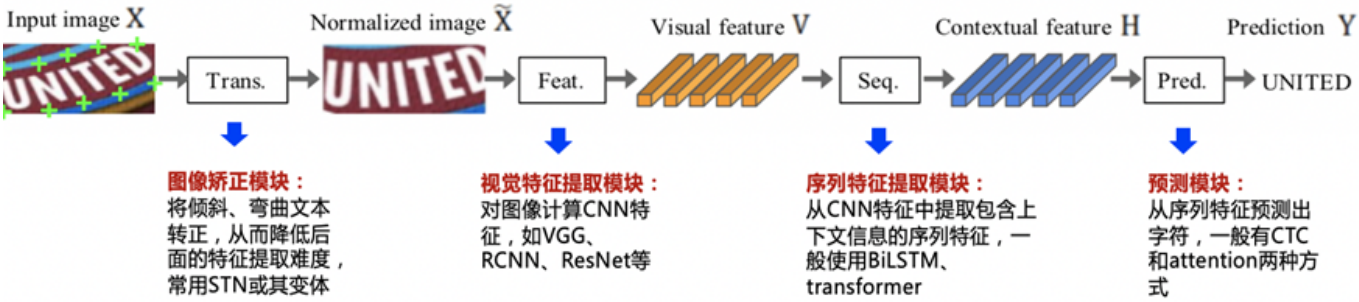


深度学习 OCR 技术流程

深度学习图像文字识别流程包括：输入图像、深度学习文字区域检测、预处理、特征提取、深度学习识别器、



深度学习后处理等。 现有多数深度学习识别算法具体流程包括图像校正、特征提取、序列预测等模块，流程如图所示：



常用文字检测框架介绍

DBNet（早期 14-16）

> 首先，DB 是一种基于分割的文本检测算法。在各种文本检测算法中，基于分割的检测算法可以更好地处理弯曲等不规则形状文本，因此往往能取得更好的检测效果。但分割法后处理步骤中将分割结果转化为检测框的流程复杂，耗时严重。因此作者提出一个可微的二值化模块（Differentiable Binarization，简称 DB），将二值化阈值加入训练中学习，可以获得更准确的检测边界，从而简化后处理流程。DB 算法最终在 5 个数据集上达到了 state-of-art 的效果和性能。

- 主要思想：先获取图像中的文本区域，再利用 **opencv、polygon 等后处理得到文本区域的最小包围曲线；**

- DB 提出可微分阈值，通过一个近似于阶跃函数的二值化函数使得分割网络在训练时学习文本分割的动态阈值，使模型提升精度，简化后处理；
- DB 的 backbone 时典型的 FCN 结构，由多层上采样和下采样的特征图 concat 完成。

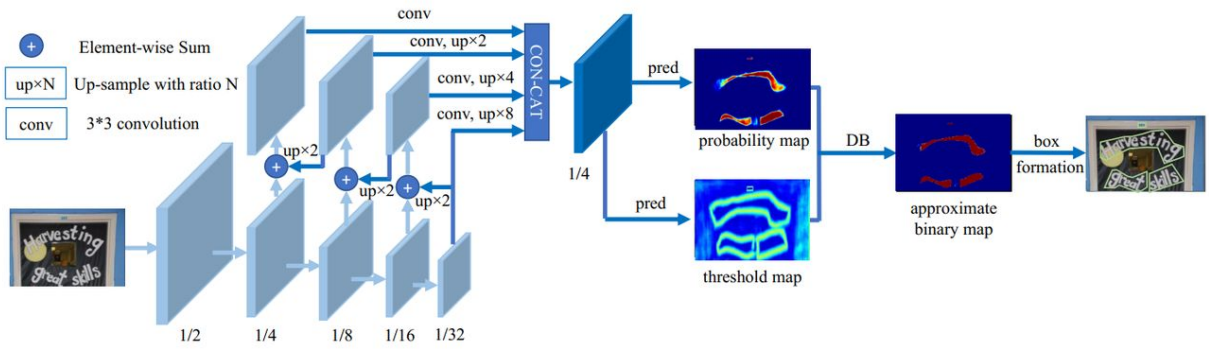
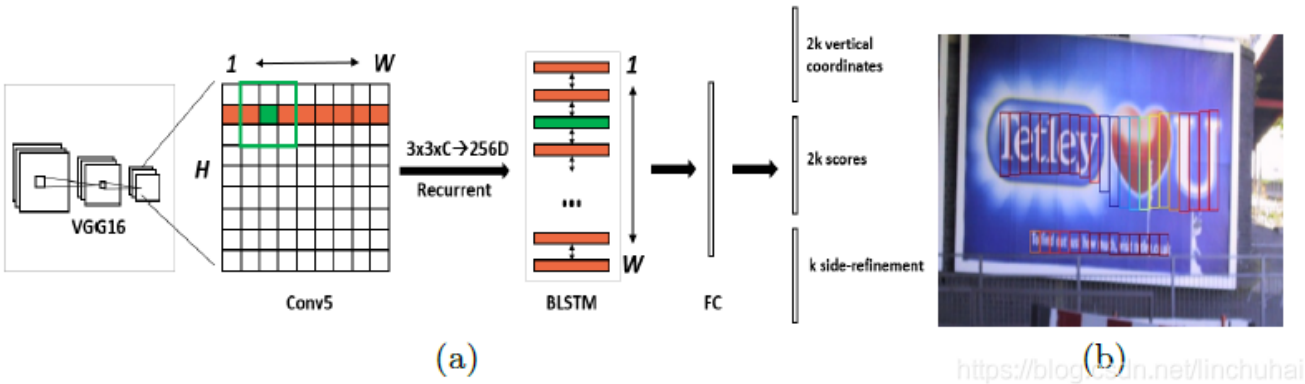


Figure 3: Architecture of our proposed method, where "pred" consists of a 3×3 convolutional operator and two de-convolutional operators with stride 2. The "1/2", "1/4", ... and "1/32" indicate the scale ratio compared to the input image.

CTPN（中期 17-18）

CTPN 模型主要包括三个部分，分别是卷积层、Bi-LSTM 层、全连接层，其结构如下图所示。



> 在卷积层部分，CTPN 选取 **VGG16 模型前 5 个 conv5= stage 得到 feature maps 作为图像最后的特征**，假设此时 feature maps 的尺寸为 W *H * C； 由于文本之间存在序列关系，因此，作者引入了递归神经网络，采用的是一层 Bi-LSTM 层，作者发现**引入了递归神经网络对文本检测的效果有一个很大的提升**，如下图所示，第一行是不采用递归神经网络的效果，第二行是采用了 Bi-LSTM 后的效果。具体的做法是采用一个的滑动窗口，提取 feature maps 上每个点周围的区域作为该点的特征向量表示，此时，图像的尺度变为，然后将每一行作为序列的长度，高度作为 batch_size，传入一个 128 维的 Bi-LSTM，得到 Bi-LSTM 层的输出为； 将 Bi-LSTM 的输出接入全连接层，在这一部分，作者引入了 anchor 的机制，即对每一个点用 k 个 anchor 进行预测，每个 anchor 就是一个盒子，其高度由 [273，390，...，11] 逐渐递减，每次除以 0.7，总共有 10 个。作者采用的是三个全连接层分支。

OCR 评价指标

OCR 评价指标包括字段粒度和字符粒度的识别效果评价指标。

以字段为单位的统计和分析，适用于卡证类、 票据类等结构化程度较高的 OCR 应用评测。 以字符 （文字和标点符号） 为单位的统计和分析，适用于通用印刷体、手写体类非结构化数据的 OCR 应用评测。具体指标包括以下几个：

a **字段召回率**，指被完全正确识别字段（测试输出结果与字段的所有字符完全匹配）数量与总字段数比值。

b **字段准确率**，指被完全正确识别字段（测试输出结果与字段的所有字符完全匹配）数量与测试返回识别结果的字段数量比值。

c **字符召回率**，指被完全正确识别字符数量与真实字符总数的比值，可以反应识别错和漏识别的情况。

d **字符准确率**，指被完全正确识别字符数量与测试返回的字符数的比值，可以反应识别错和多识别的情况。

e F_β - Score，可以综合反映字符识别召回效果和字符识别准确效果，计算公式如下:

$$F_\beta\text{-}Score = (1+\beta^2) * \frac{字符召回率 * 字符准确率}{\beta^2 * (字符召回率 + 字符准确率)}$$

f **最小编辑距离**，表示测试结果要与标注结果一致需要修改的字符数，忽略引擎返回行的顺序与原图标注的顺序，适用于电商广告、手机截图等样本版式不规范的 OCR 应用评测。

g **全图编辑距离**，表示测试返回结果要与标注结果一致需要修改的字符数，要求引擎返回的每一行文字顺序要和标注顺序一致，适用于文档、表格等样本版式较为规范的 OCR 应用评测。

编辑距离：

编辑距离是针对二个字符串（例如英文字）的差异程度的量化量测，量测方式是看至少需要多少次的处理才能将一个字符串变成另一个字符串。在莱文斯坦距离中，可以删除、加入、替换字符串中的任何一个字元，也是较常用的编辑距离定义，常常提到编辑距离时，指的就是莱文斯坦距离。

测试指标说明：

平均识别率：[1 - (编辑距离 /max (1, groundtruth 字符数，predict 字符数))] * 100.0% 的平均值； 平均编辑距离：编辑距离，用来评估整体的检测和识别模型； 平均替换错误：编辑距离计算时的替换操作，用于评估识别模型对相似字符的区分能力； 平均多字错误：编辑距离计算时的删除操作，用来评估检测模型的误检和识别模

型的多字错误； 平均漏字错误：编辑距离计算时的插入操作，用来评估检测模型的漏检和识别模型的少字错误；