



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

OCR及领域优化

鄧睿 杨瑞 沈国童 窦义通 李钊斌 申志宇

2021/10/11



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

目录

CONTENTS

1

OCR简介及应用场景

2

图像预处理

3

文本检测

4

文本识别

5

领域优化

6

Demo展示



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

1

OCR简介及应用场景



OCR概念定义

光学字符识别 (Optical Character Recognition, OCR) 是指对文本资料的图像文件进行分析识别处理, 获取文字及版面信息的过程。即将图像中的文字进行识别, 并以文本的形式返回。

■ OCR基本原理

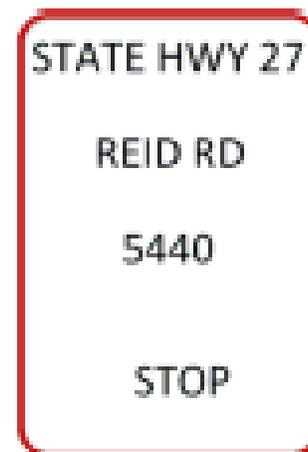
OCR文字识别的原理是计算机对图像文字进行预处理、分割和坐标定位，通过检测暗、亮，放大图像确定其形状特征，最终通过黑白点阵的图像文件与字符编码进行匹配，根据最相近的匹配度将图像中的文字转换成文本。



(a)



(b)

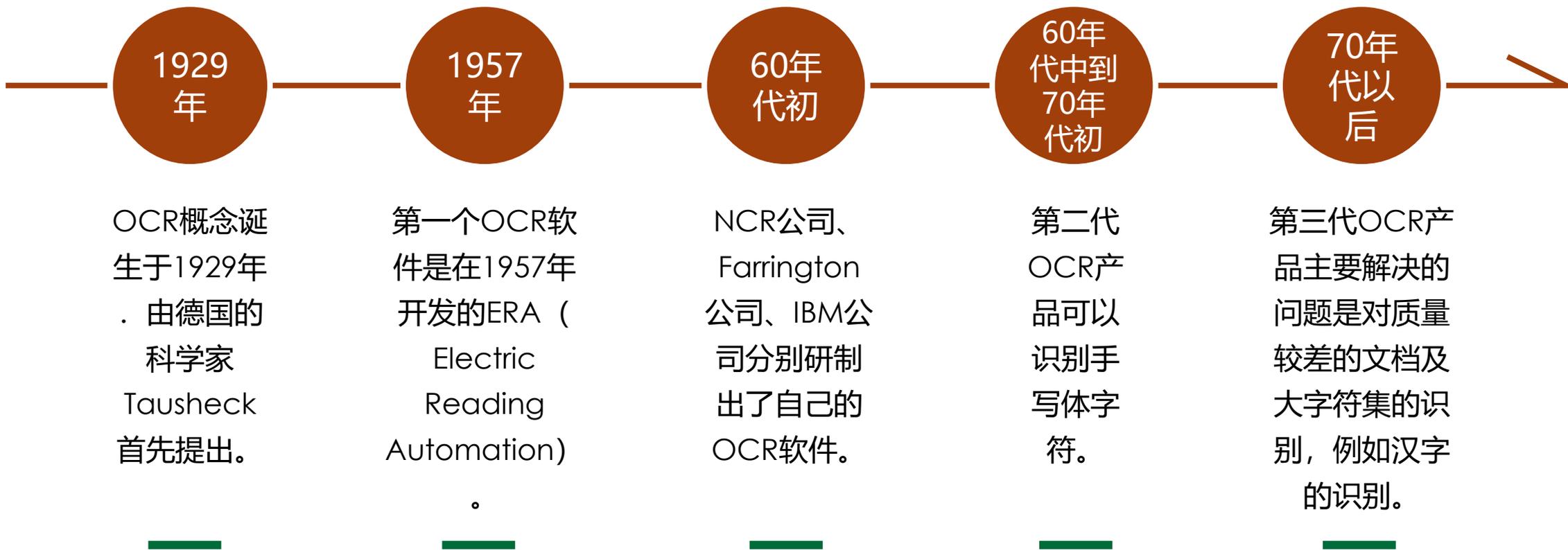


(c)



- 图像预处理通常是针对图像的成像问题进行修正。常见的预处理过程包括：几何变换（透视、扭曲、旋转等）、畸变校正、去除模糊、图像增强和光线校正等。
- 文字检测和文本识别是影响识别准确率的技术瓶颈，也是OCR技术的重点。
- 文字检测即检测文本的位置、范围和布局。通常也包括版面分析和文字行检测等。文字检测主要解决的问题是哪里有文字，文字的范围有多大。
- 文本识别是在文本检测的基础上，对文本内容进行识别，将图像中的文字转化为文本信息。文字识别主要解决的问题是每个文字是什么。

诞生 首个OCR软件 第一阶段 第二阶段 第三阶段



1.4 中文OCR技术的发展

探索阶段(1979—1985)

我国在OCR技术方面的研究工作起步较晚，在70年代才开始对数字、英文字母及符号的识别进行研究，70年代末开始进行汉字识别的研究。

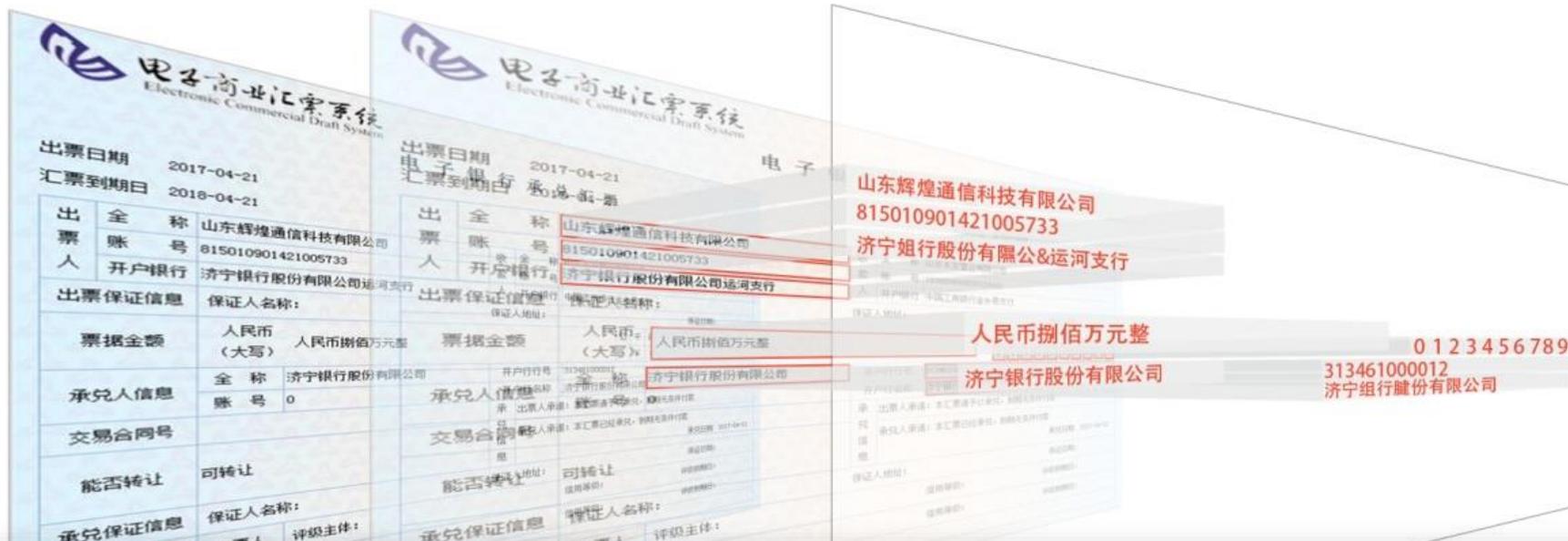
研发阶段(1986~1988)

1986年初到1988年底，汉字识别的研究进入一个实质性的阶段，总共有11个单位进行了14次印刷体汉字识别的成果鉴定。

实用阶段(1989—至今)

1989年，清华大学推出了国内第一套中文OCR软件，至此中文OCR正式从实验室走向了市场。目前，汉字识别技术的研究热点已经从单纯的文字识别转移到了表格的自动识别与录入，图文混排和多语种混排的版面分析、理解和恢复，金融票据识别等内容上。未来随着传统行业的数字化转型，OCR应用范围将进一步扩展。

1.5 OCR技术应用举例



- 1.OCR在银行中的应用
- 借助OCR识别技术，使票据处理达到“自动集中处理+电子安全保存+数据有效应用”的目标，不但保留了原始单据的图像文件，而且对图像进行OCR识别后可以自动建立票据索引，完成单据与前台数据的比对，快速找出有问题的单据，提高银行业务的自动化和智能化水平。

- 2.OCR在快递行业的应用
- OCR识别技术能快速提取快递面单上的联系人、手机号、地址等重要信息，并与系统数据进行匹配，在短时间内，完成大量包裹的自动分拣。



1.5 OCR技术应用举例

- 3.OCR在视频审查中的应用
- 借助OCR技术识别视频中的文字，对互联网视频内容进行识别审核、监控，筛除掉包含敏感、不健康词汇的视频、广告，达到高效审查的目的。





IOU

IOU 的全称为交并比 (Intersection over Union)

IOU 计算的是“检测的边框”和“真实的边框”的交集和并集的比值。

$$IoU = \frac{|A \cap B|}{|A \cup B|}$$

$$IoU = \frac{\text{Area of Overlap}}{\text{Area of Union}}$$



衡量OCR系统性能的指标：

分为检测阶段和识别阶段

■ 检测阶段：

按照检测框和标注框的IOU评估，IOU大于某个阈值判断为检测准确。

- 1.检测准确率：正确的检测框个数在全部检测框的占比，主要是判断检测的指标。
- 2.检测召回率：正确的检测框个数在全部标注框的占比，主要是判断漏检的指标。

■ 识别阶段：

字符识别准确率：正确识别的文本行占标注的文本行数量的比例，只有整行文本识别对才算正确识别。



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

2

图像预处理

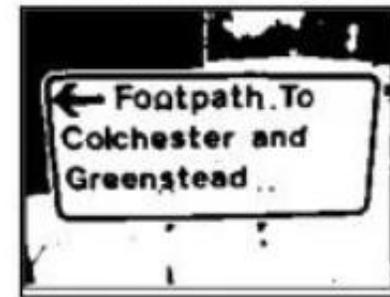
- 图像预处理的作用：要识别的图片很可能是倾斜的，或者是带噪声的，所以需要进行图片预处理，把文本位置矫正，把噪声去除，然后才能进行进一步的字符分割和文字识别。
- 图像预处理过程：通常包含二值化、平滑噪声、倾斜角检测和矫正三个步骤。

- 二值化，就是通过将像素点的灰度值设为0或255，将灰度值图像信号转化成只有黑(1)和白(0)的二值图像信号。在传统方法甚至是现在的流行方法中，高质量的二值化图像仍然可以显著提升OCR效果，一方面减少了数据维度，另一方面排除噪声凸显有效区域。目前，二值化方法主要分为四种：1.全局阈值方法；2.局部阈值方法；3.基于深度学习的方法；4.基于形态学和阈值的文档图像二值化方法。

input image

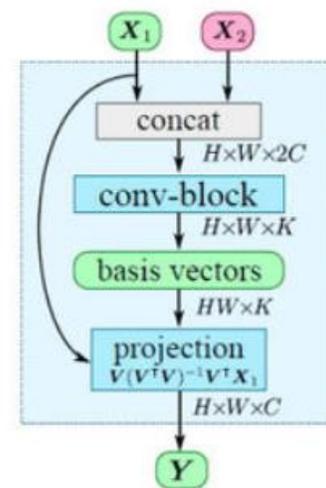
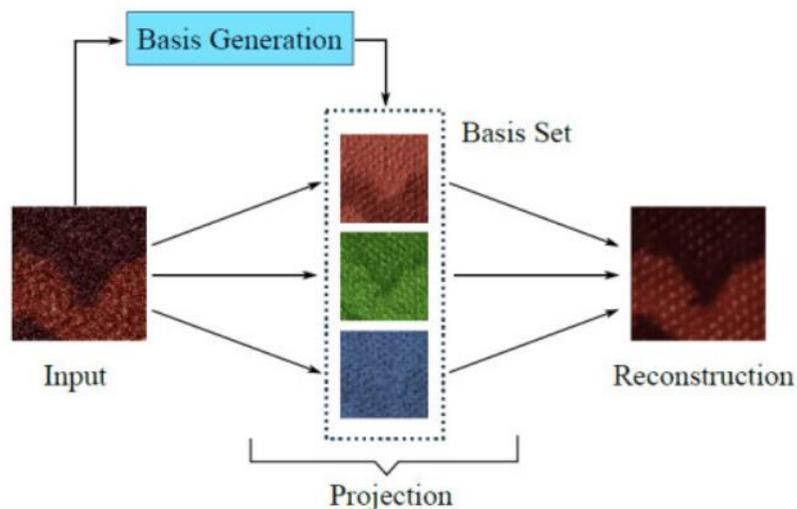


output image

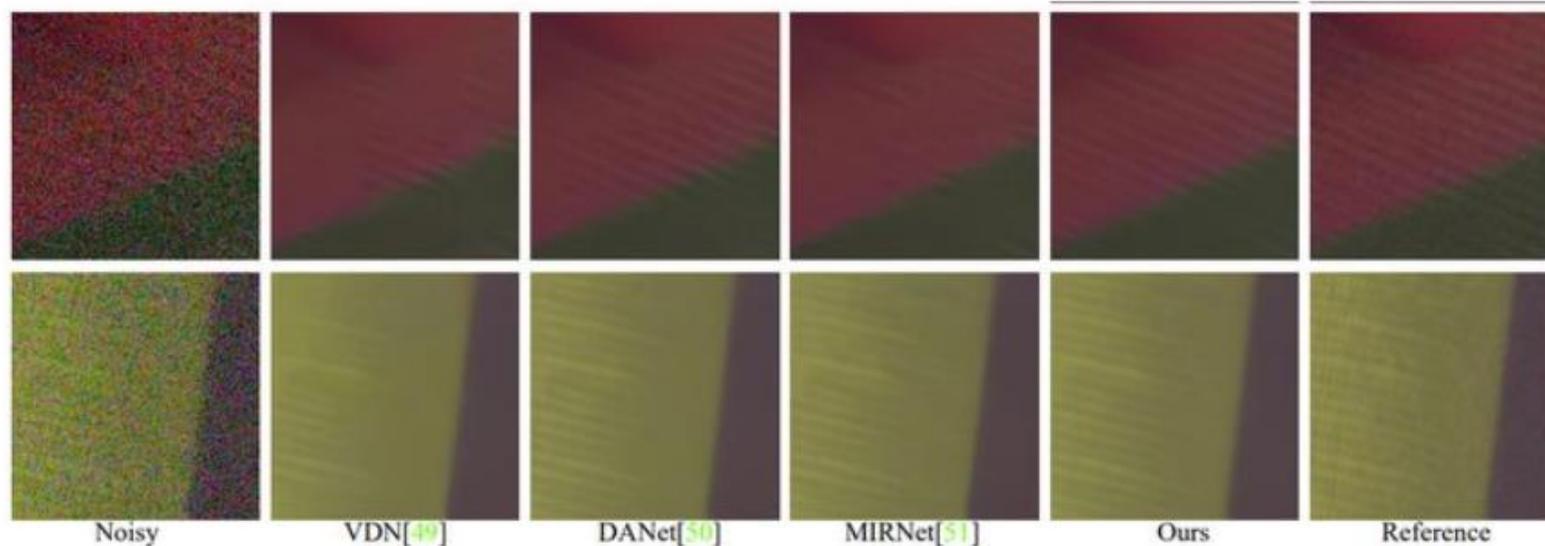
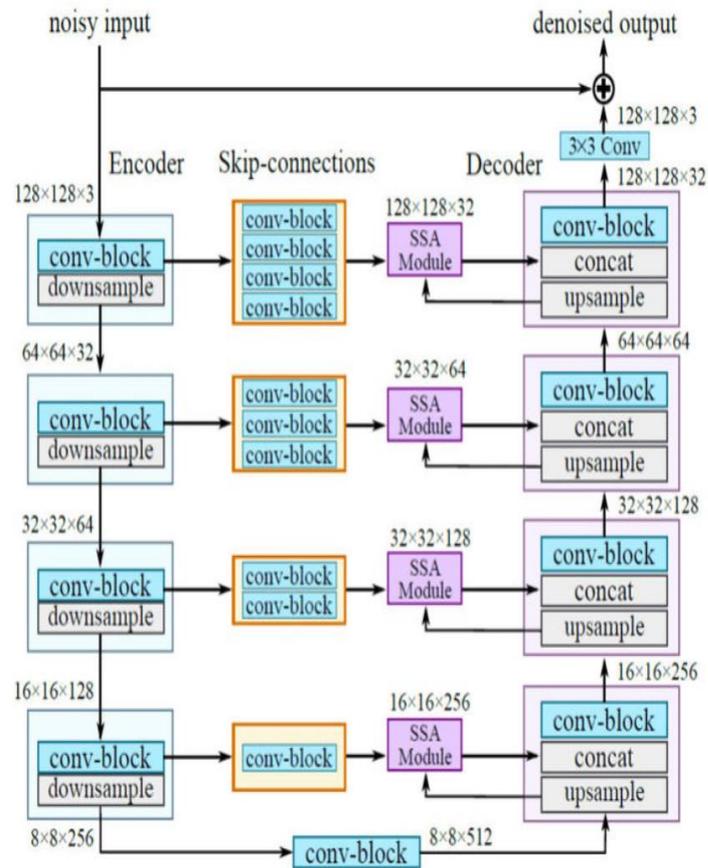


- 图像噪声是指存在于图像数据中的不必要的或多余的干扰信息，产生于图像的采集、量化或传输过程，对图像的后续处理、分析均会产生极大的影响。
- 传统方法是利用图像先验和噪声模型估计图像或是噪声，例如NLM和BM3D利用图像局部相似性和噪声独立性，小波去噪方法利用图像在变换域上的稀疏性。
- 近年来，基于深度学习的去噪方法隐式地利用图像先验或从大规模的成对数据集中学习噪声分布。尽管基于深度学习的方法取得了很大成功，但在弱纹理或高频细节这样的困难场景下，降噪的同时也同样会丢失许多图像细节，恢复高质量图像仍然是一项挑战。

- 为解决上述问题提出了一种新的图像去噪网络NBNet--通过图像自适应投影进行降噪。投影可以较好的保留图像的结构信息，投影也是一种捕获全局相关性的方式。投影可以训练一种信息与噪声分离的网络。具体做法就是从输入图像生成一系列的图像基底向量，然后从这些基底向量构成的子空间重建出去噪图像。
- NBNet整体是UNet形式的网络，其中关键的是子空间注意力模块SSA。NBNet的创新在于子空间投影，包括两个主要的步骤：基底向量生成和子空间投影。



(c) Subspace Attention (SSA) Module



实验结果显示NBNet在线条、点状等弱纹理图案处优于其他方法

- 图像在扫描过程中，很容易出现文档旋转和位移的情况，常见的倾斜角检测校正方法有霍夫变换、Randon变换以及基于PCA的方法。
- 其中最常用的方法是霍夫变换，霍夫变换的使用一般分为3个步骤：1.用霍夫变换探测出图像中的所有直线；2.计算出每条直线的倾斜角，求它们的平均值；3.根据倾斜角旋转矫正图片。
- 其原理是将图片进行膨胀处理，将断续的文字连成一条直线，便于直线检测。计算出直线的角度后就可以利用旋转算法，将倾斜图片矫正到水平位置。

2.1.3 倾斜角检测和矫正





北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

3 文本检测

3.1 文本检测是什么

- 1.文本检测就是要定位图像中的文字区域，然后通常以边界框的形式将单词或文本行标记出来。
- 2.作用：将输入图片中的文字区域位置检测出来，用于后续的文字识别，因为只有找到了文本所在区域，才能对其内容进行识别。

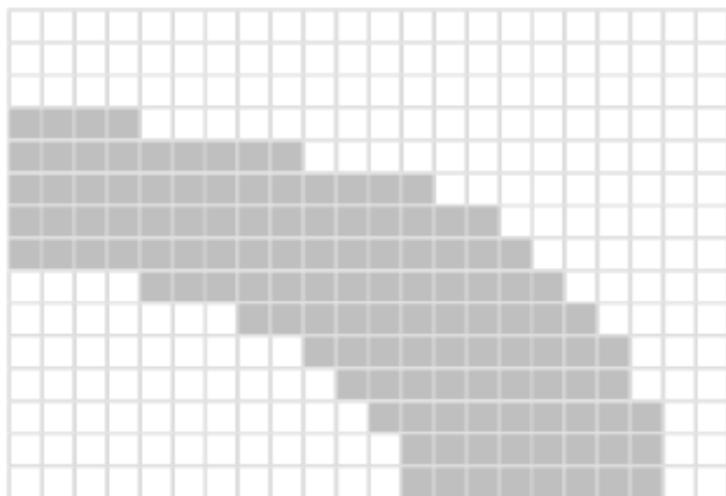


传统的文本检测方法一般采用手工特征提取的方式进行检测文本，比如 SWT、MSER等方法，然后采用模板匹配或模型训练的方法对检测到的文本进行识别。而现在的深度学习方法使用卷积神经网络代替手工提取特征方法进行文本检测。下面以SWT为例来简要介绍一下传统的文本检测方法。

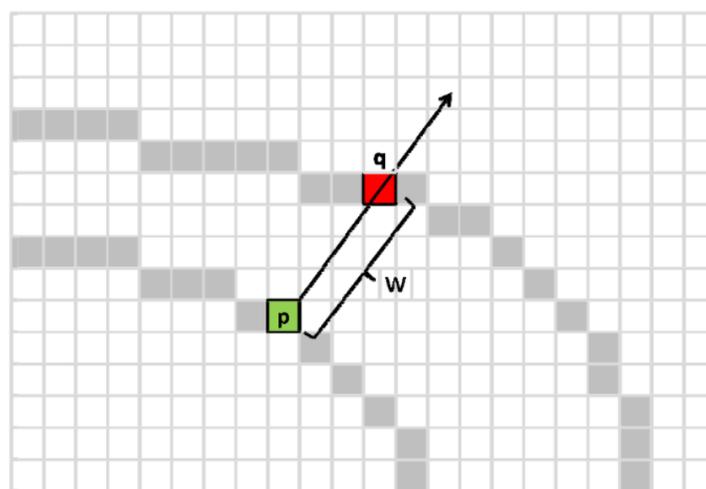
3.3 SWT (Stroke Width Transform)

笔画宽度变换算法步骤：

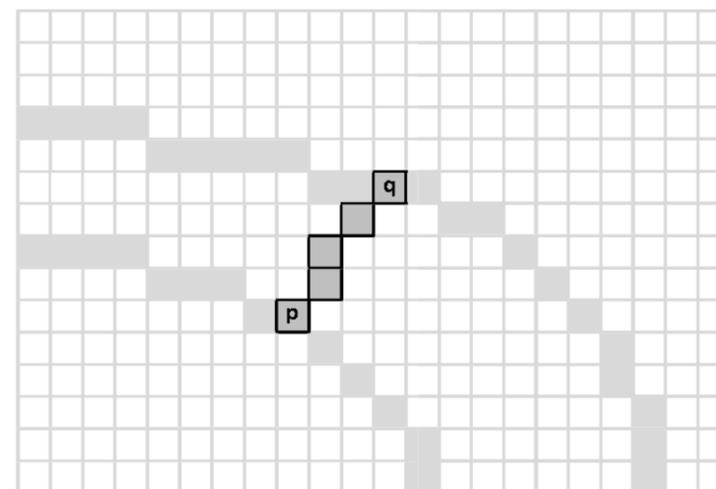
1. 笔画宽度变换，输出SWT图像
2. 通过SWT图像得到多个连通域
3. 通过自定义的规则过滤一些连通域，得到候选连通域
4. 将连通域合并得到文本行



(a)



(b)



(c)

后来传统的文本检测方法被日益发展的基于深度学习的方法代替了。

目前基于深度学习的方法主要有两大类：

(1) 从目标检测发展来的基于候选框的文本检测算法(Proposal-based)，其模型有CTPN，TextBoxes，TextBoxes++等等。

(2) 基于分割方法的文本检测(Segmentation-based)主要有EAST，PSENet，DBNet等等。

接下来我们分别来看一下

几个目标检测中的基本概念：

■ (1) 边界框 (bounding box, bbox) :

两种格式：

1. (x_1, y_1, x_2, y_2) , 其中 (x_1, y_1) 和 (x_2, y_2) 分别表示左上角和右下角的坐标。
2. (x, y, w, h) , 其中 (x, y) 是矩形框中心点的坐标, w 和 h 是矩形框的宽度和高度。

■ (2) 锚框 (anchor box) :

由人们假想出来的一种框。在目标检测模型中, 模型对这些候选区域是否包含物体进行预测, 如果包含目标物体, 则还需要进一步预测出物体所属的类别。

几个目标检测中的基本概念：

■ (3) 交并比：

使用Jaccard系数可以衡量两个集合的相似度， $J(A,B) = |A \cap B| / |A \cup B|$ 。其中，A和B集合分别为锚框与真实边界框的像素。此时计算出的 $J(A,B)$ 即为交并比，其为0表示两个框无重合，1表示完成重合。

■ (4) NMS（非极大值抑制）：

步骤：1.将预测的边界框按照置信度排序；2.选择最高的添加到最终输出，并将其从边界框列表中删除；3.计算置信度最高的边界框与其他候选框的交并比；4.删除交并比大于阈值的边界框；5.重复上述步骤直到边界框列表为空。

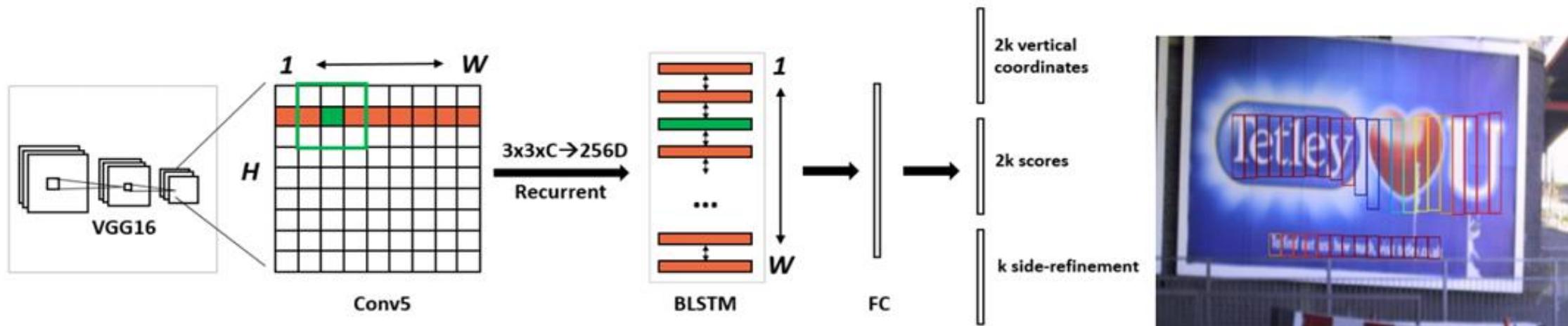
本质：搜索局部最大值，抑制非极大值

基于候选框的文本检测大致的思想就是根据设置的锚产生一系列的候选文本框，再进行一系列调整、筛选，最终通过NMS得到最终的文本边界框。

CTPN (Detecting Text in Natural Image with Connectionist Text Proposal Network, 基于连接预选框网络的文本检测) 是在 Faster RCNN 的基础上进行改进, 是目前应用最广的文本检测模型之一。其关键点如下:

1. 采用垂直 anchor 回归机制, 检测小尺度的文本候选框
2. 提出了一个 vertical anchor 的方法来解决文本长度变化不定的问题。只预测文本的竖直方向上的位置, 水平方向的位置不预测, 水平位置的确定只需要我们检测一个一个小的固定宽度的文本段, 将他们对应的高度预测准确, 最后再将他们连接在一起, 就得到了文本行。
3. 采用 RNN 将检测的小尺度文本进行连接, 得到文本行。
4. 采用 CNN + RNN 端到端的训练方式, 支持多尺度和多语言, 避免后处理。

CTPN网络总体结构：



1. 对输入的N个图像，使用VGG16卷积得到conv5的特征作为feature map，大小为 $N \times W \times H \times C$ 。
2. 对得到的feature map用3*3的滑动窗口提取特征，最后输出 $N \times W \times H \times 9C$ 的feature map
3. 将feature map reshape成为 $(NH) \times W \times 9C$ ，然后双向LSTM，学习每一行的序列特征。最后双向LSTM输出 $(NH) \times W \times 256$ ，然后重新reshape回 $N \times 256 \times H \times W$ 。该特征既包含空间特征，也包含了LSTM学习到的序列特征。
4. 经过一个卷积层，变成 $N \times H \times W \times 512$ 的特征。
5. $N \times H \times W \times 512$ 最后经过类似Faster R-CNN的RPN网络获得text proposals，然后通过NMS来过滤掉多余的文本框。
6. 文本位置水平，会将上述得到的一个文本小框使用文本线构造方法合成一个完整文本行，如果还有些倾斜，先进行矫正。

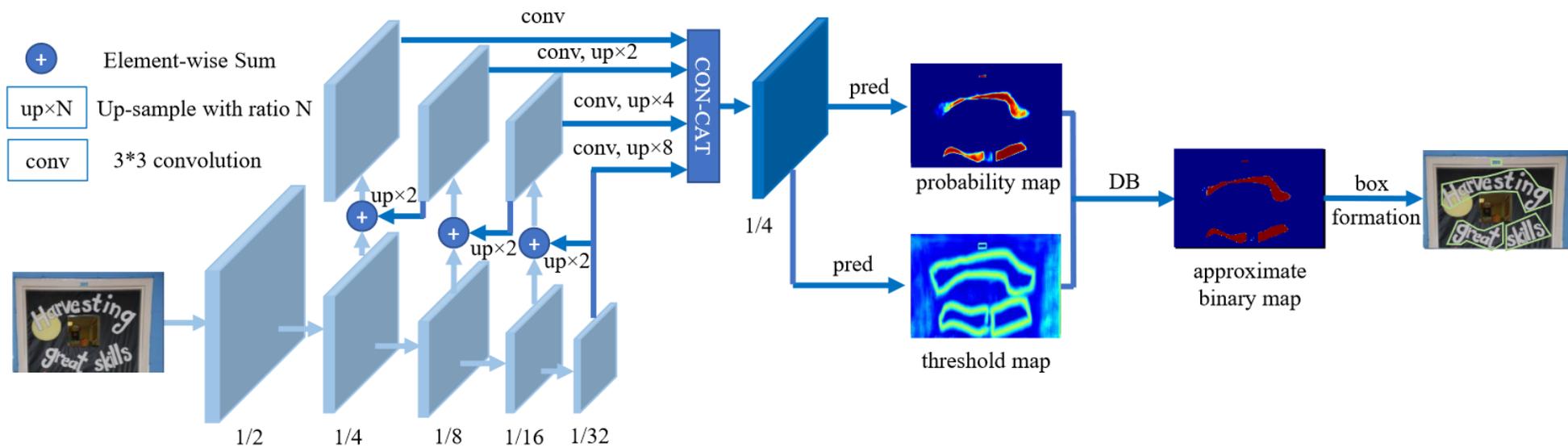
3.7 基于分割方法的文本检测

- 一般场景下，可以通过对通用目标检测算法进行配置修改，来实现对文本行的检测定位。然而在弯曲文字场景，通用目标检测算法无法实现对文字边框的精准表述。
- 基于分割方法的文本检测一般先对网络结构进行分割以达到像素级别的语义分割，然后根据分割的结果构建文本行，其对各种形状（弯曲、竖直、多方向）的场景文本检测更加精确。例如PSENet、DBNet和FTSN，它们会生成分割映射，然后再经过后续其他处理，最终得到文本边界框。
- 下面我们以DBNet为例来简要介绍基于分割方法的文本检测。



3.8 DBNet(Differentiable Binarization)结构

整体过程：



将图片输入网络后，经过特征提取和上采样融合（使用Feature Pyramid Network即FPN进行卷积，）并concat操作后得到图中蓝色的1/4大小的特征图称为F，然后使用F预测出概率图（probability map）称为P和使用F预测出阈值图（threshold map）称为T，最后通过P和T计算出近似二值图。推断过程文本框可以通过近似二值图或者概率图来获取。

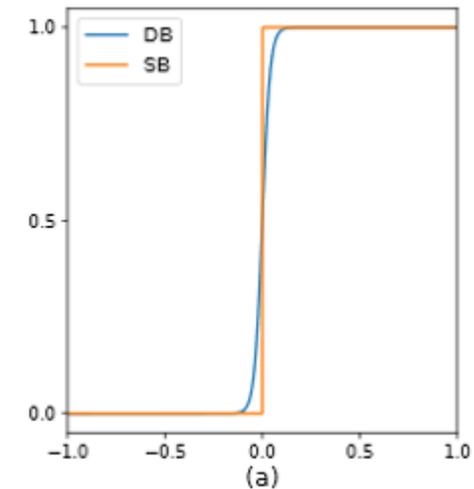
3.9 Differentiable Binarization

普通二值化

$$B_{i,j} = \begin{cases} 1 & \text{if } P_{i,j} \geq t, \\ 0 & \text{otherwise.} \end{cases} \quad (1)$$

近似的可微分二值化

$$\hat{B}_{i,j} = \frac{1}{1 + e^{-k(P_{i,j} - T_{i,j})}} \quad (2)$$



公式 (1) 中 t 为自定义的阈值。

公式 (2) 中，论文里将参数 k 取值为50， $P_{i,j}$ 是预测出的概率图像素点的值， $T_{i,j}$ 是预测出的阈值图对应像素点的值。

常规的的二值化方法不可微，所以没法放入网络学习中优化。新提出的DB二值化方法可以继续进行监督网络学习来优化。。



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

4 文本识别

4.1 文本识别定义

- 文本识别是在文本检测的基础上，对定位好的文字区域进行识别，主要解决的问题是每个文字是什么，将图像中的文字区域进转化为字符信息。

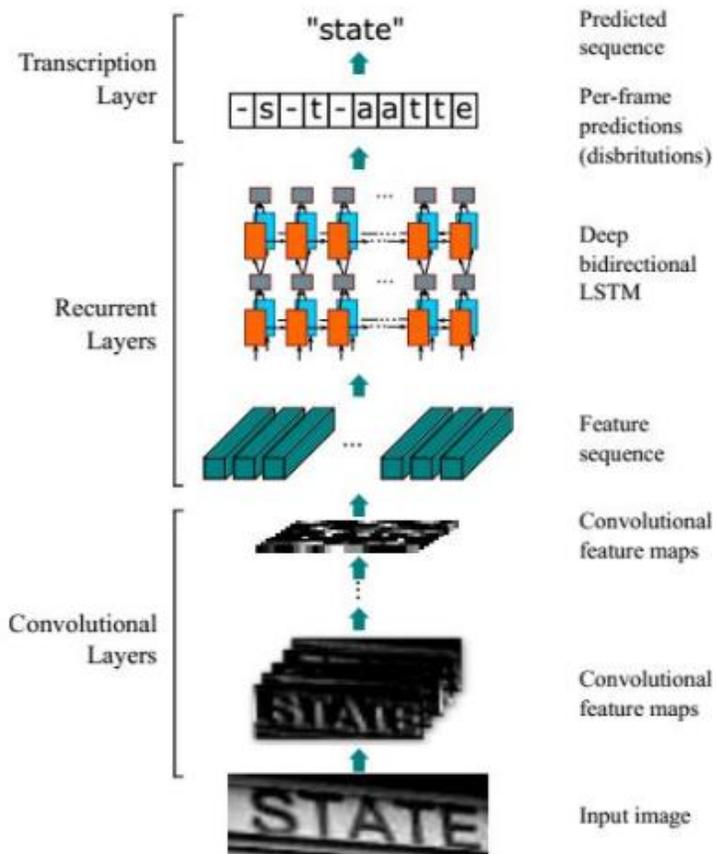


文字检测算法已经定位图中的“subway”区域（红框），接下来就是文字识别

- 文本识别在传统技术中采用模板匹配的方式进行分类。
- 但是对于文字行，只能通过识别出每一个字符来确定最终文字行从内容。因此可以对文字行进行字符切分，以得到单个文字。
- 过分割-动态规划是最常见的切分方法，对候选字符进行过分割，使其足够破碎，之后通过动态规划合并分割碎片。
- 另一种方法是通过滑动窗口对每一个可能的字符进行匹配，

- 以上的传统方法通过识别每个单字符以实现全文的识别，这一过程导致了上下文信息的丢失，对于单个字符有较高的识别正确率，其条目识别正确率也难以保证。
- 因此引入上下文的信息，成为了提升条目准确率的关键。从深度学习的角度出发，要引入上下文这样的序列信息，RNN和LSTM等依赖于时序关系的神经网络是最理想的选择。
- 常见的一种做法是利用CRNN模型，另一种方法是引入注意力机制。

- 与语音识别问题类似，OCR可建模为时序依赖的词汇或者识别问题。一些学者尝试把CTC损失函数借鉴到OCR识别中，CRNN 就是其中代表性算法。
- 以CNN特征作为输入，双向LSTM进行序列处理使得文字识别的效率大幅提升，也提升了模型的泛化能力。先由分类方法得到特征图，之后通过CTC对结果进行翻译得到输出结果。使用目前最为广泛的一种文本识别框架。需要自己构建字词库（包含常用字、各类字符等）



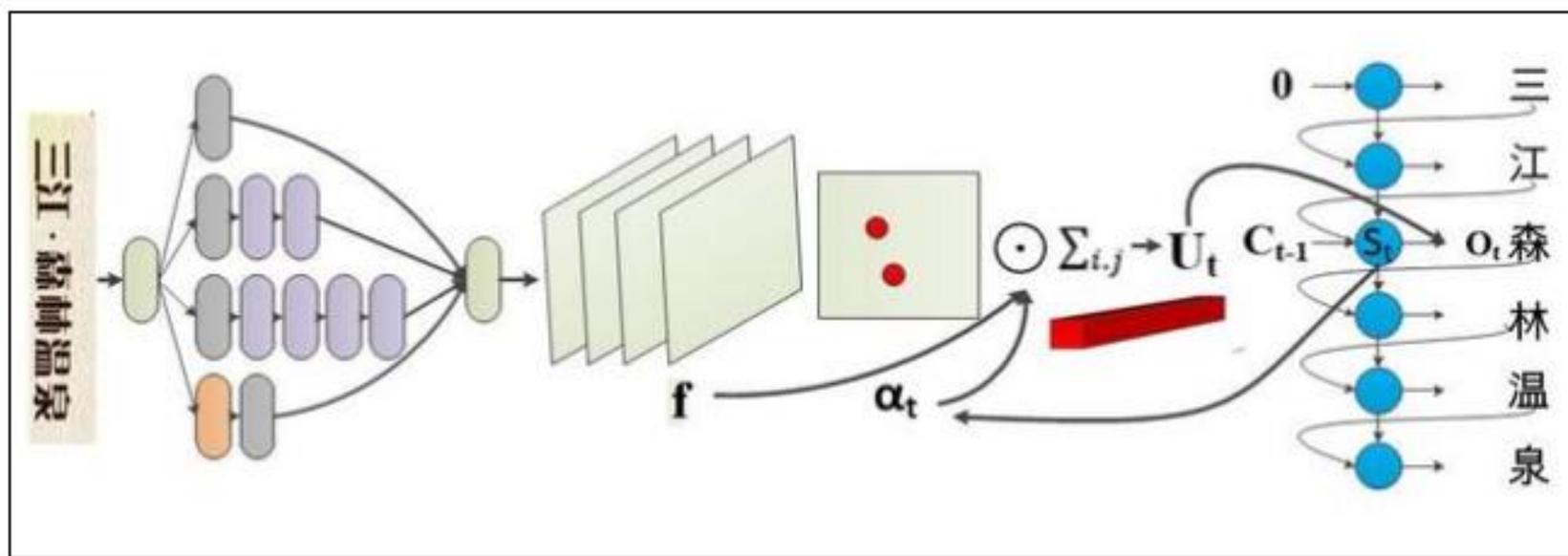
CNN（卷积层），使用深度CNN，对输入图像提取特征，得到特征图；

RNN（循环层），使用双向RNN（BLSTM）对特征序列进行预测，对序列中的每个特征向量进行学习，并输出预测标签（真实值）分布；

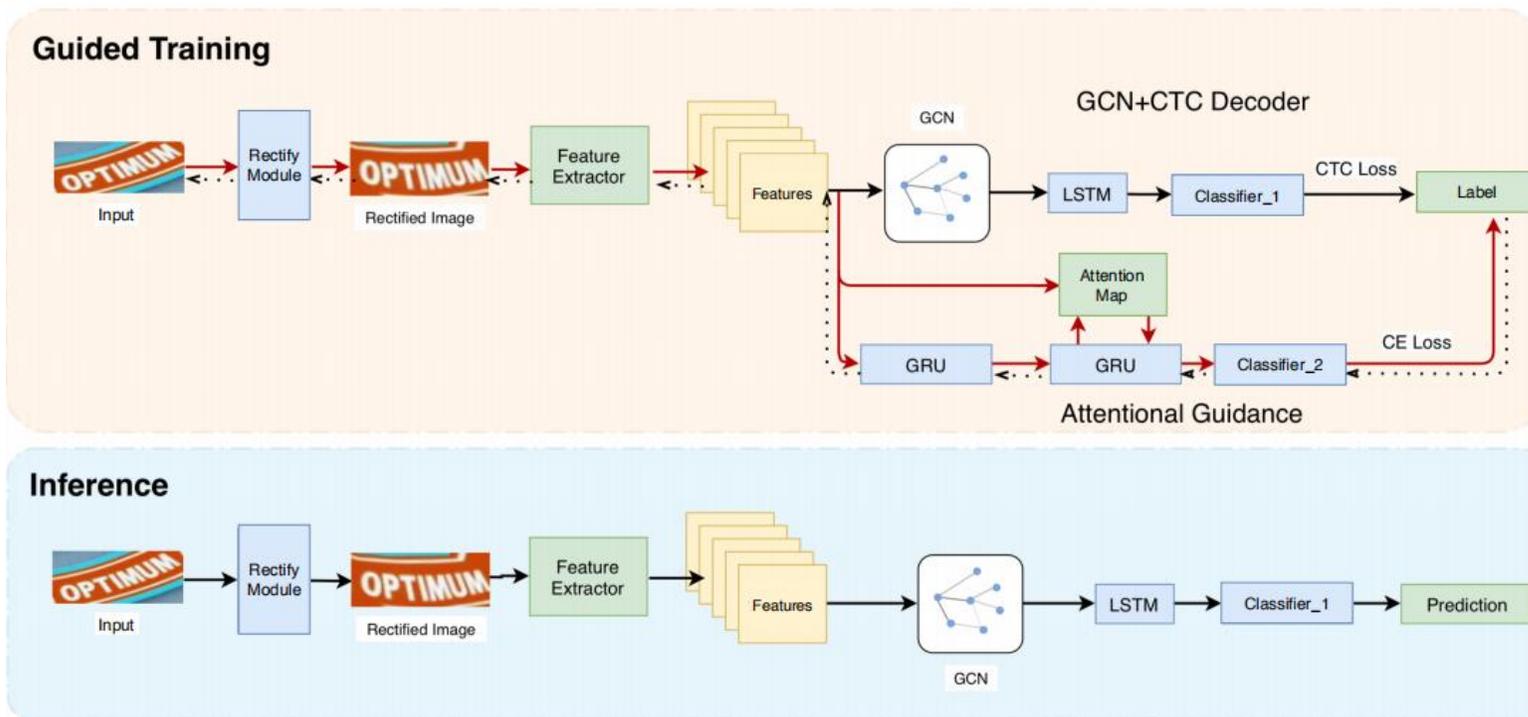
CTC loss（转录层），使用CTC损失，把从循环层获取的一系列标签分布转换成最终的标签序列。

- 引入注意力机制，以CNN特征作为输入，通过注意力模型对RNN的状态和上一状态的注意力权重计算出新一状态的注意力权重。之后将CNN特征和权重输入RNN，通过编码和解码得到结果。

- 基于注意力模型的卷积循环神经网络（Attention-CRNN）的整体框架如图所示，它主要由三部分组成：卷积神经网络、循环神经网络和注意力模型。
- 卷积神经网络从图片中提取特征,注意力模型计算注意力权重，二者结合解码后得到字符集的概率分布，提取概率最高字符作为识别结果。



- 将attention和ctc两种方式进行融合，利用attention对ctc的对齐进行监督和引导，同时，在ctc分支中加入GCN图卷积神经网络提高模型表达能力。



- 利用attention分支对矫正网络和特征提取网络进行一定的监督作用。
- 在CRNN分支中加入GCN，加强序列特征局部的学习。



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

5

领域优化

5. OCR领域优化

1. OCR领域优化目标

- 1) 识别效果 (准确率、识别精度)
- 2) 识别速度 (耗时)
- 3) 模型大小

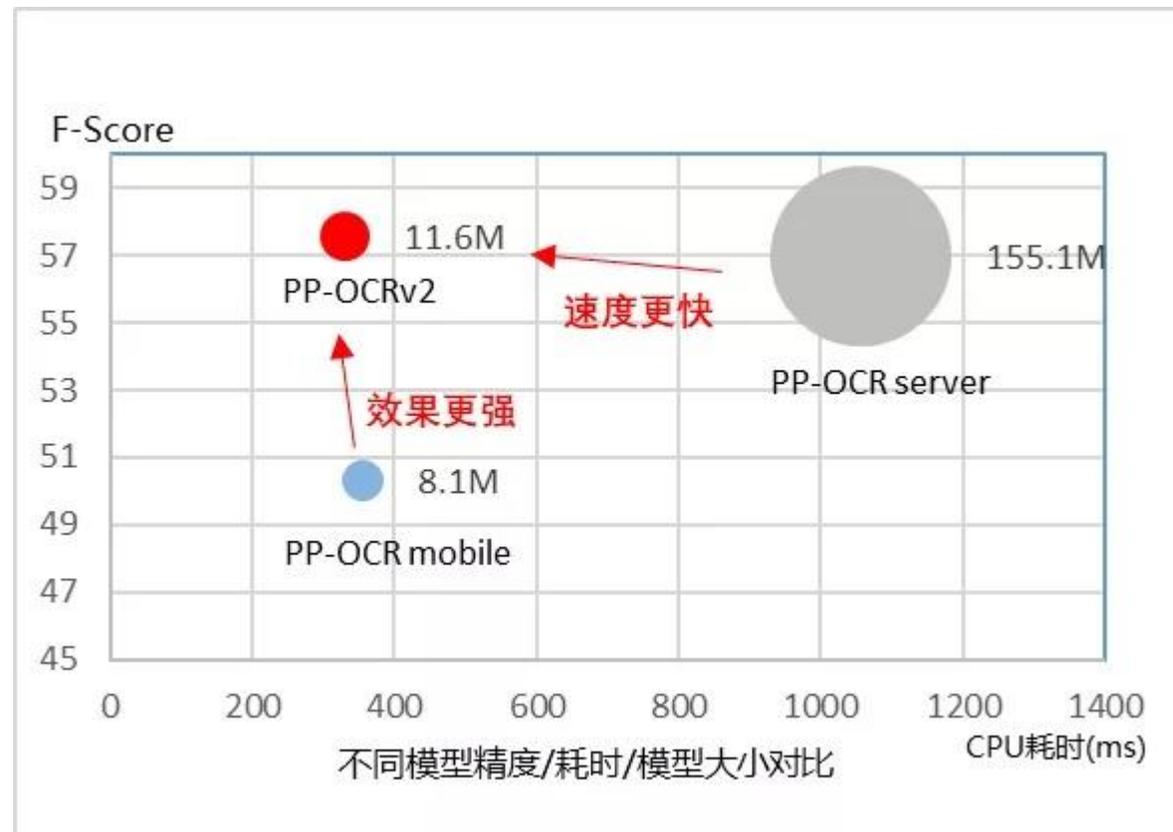
2. OCR领域优化思路 (算法)

OCR路线:

数据 → **文字检测** → **文字识别** → 输出

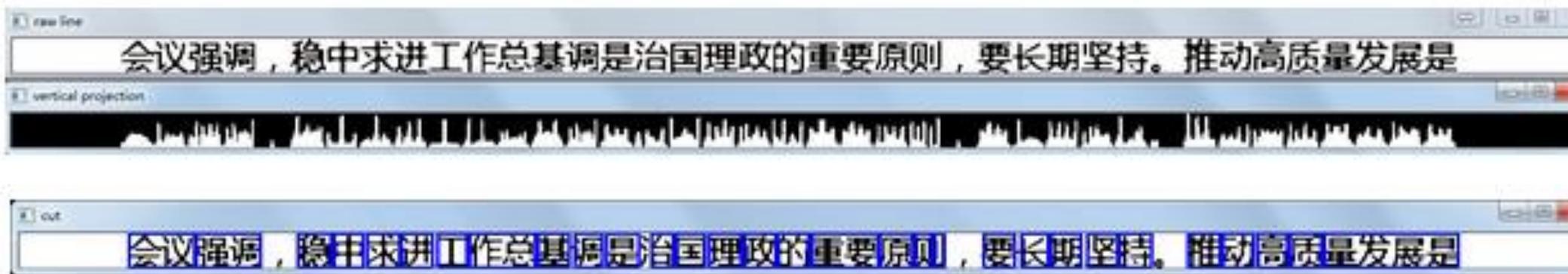
3. 优化工作的重点:

- 1) 检测优化
- 2) 识别优化



5.1 文本检测优化 (处理方法优化)

1. 文本识别-投影切割法



三种常见问题:

业注册号 业名
1.jpg 22.jpg

多个文字粘连

卩 艮
40.jpg 41.jpg

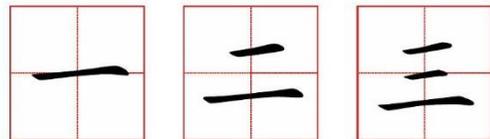
左右结构分开

江 比长兴路
58.jpg 59.jpg

粘连+分开

5.1 文本检测优化 (处理方法优化)

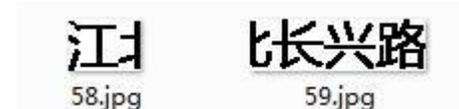
解决方法：通过长宽尺寸比进行判断。（默认汉字的长宽比在1附近）



多个文字粘连



左右结构分开



粘连+分开



长/宽 \approx 正整数

长/宽 < 1

长/宽 > 1 && 长/宽 \neq 整数

(成对出现，左图从左侧切割，右图从右侧切割。)

其他问题及解决方法：

在实际的应用中，图像来源和图像质量错综复杂，来源方面有屏幕截图、拍照和手写等，文字可能存在模糊、断裂、折叠等情况，目前有如下处理手段：

增强对比度，加入模糊算子手段处理图片模糊问题，预处理解决倾斜、噪声等问题。

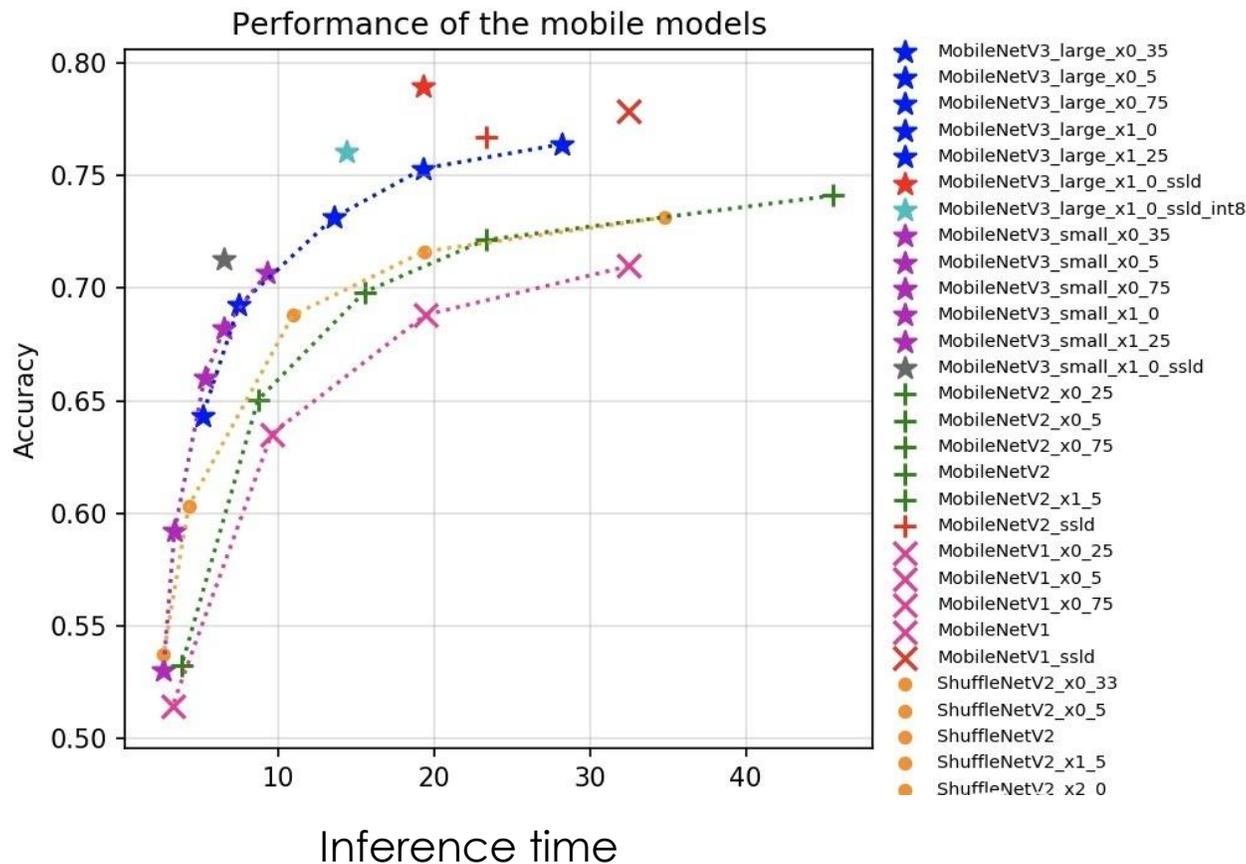
5.1 文本检测优化 (模型优化)

- 目前PP-OCR模型中主要使用以下办法进行优化，包括超轻量骨干网络选择、头部轻量化、SE模块的移除、余弦学习率、预热学习率以及使用FPGM模型裁剪器。

1. 超轻量骨干网络选择

影响检测器模型大小的一个主要因素是骨干网络结构的选择。超轻量文本检测器应该选用超轻量的骨干网络。

右图对比了不同规模超轻量模型在预测速度和准确度的情况。



5.1 文本检测优化 (模型优化)

2. 头部轻量化

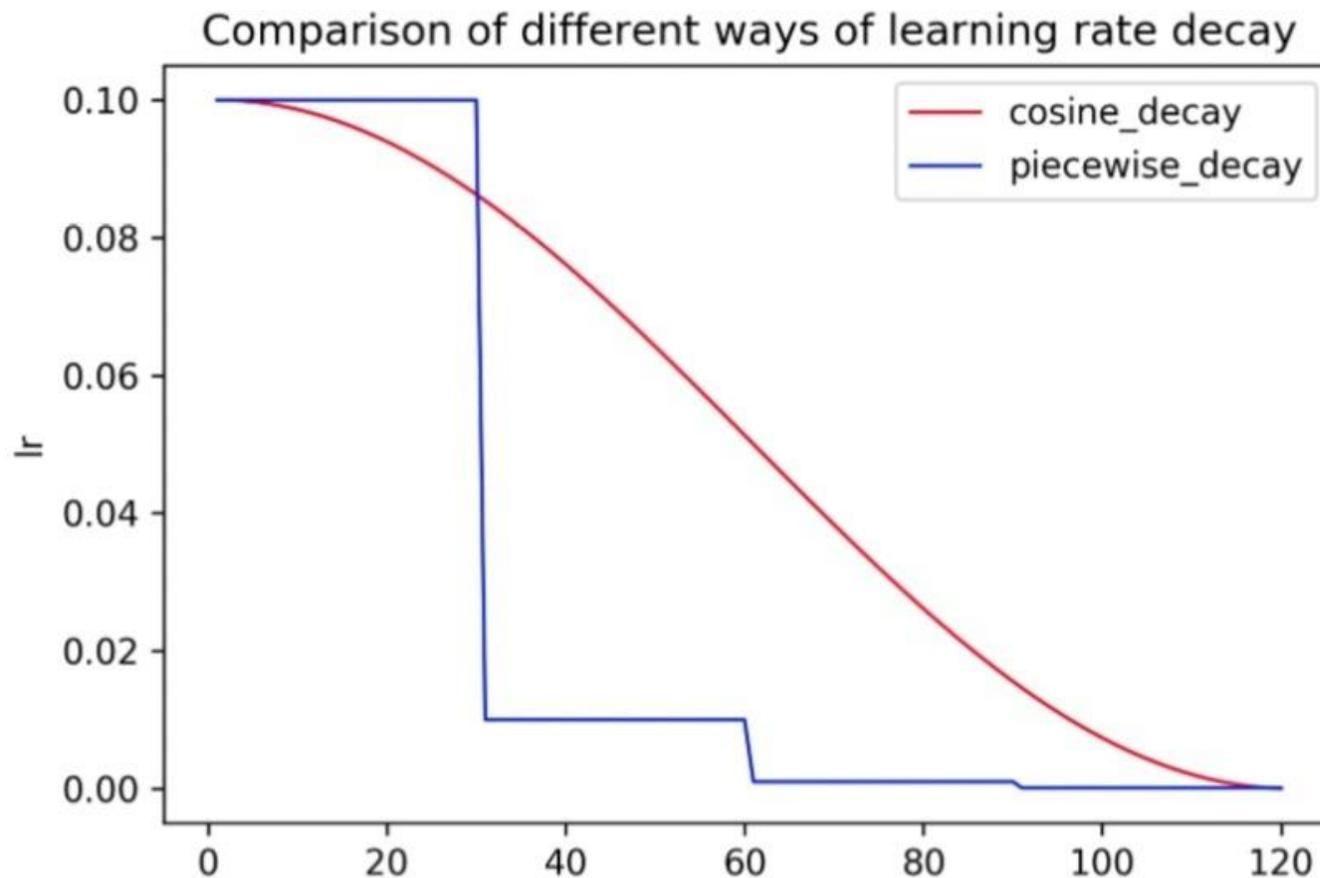
DB文本检测的头部是融合了多个分辨率的特征图，提升对不同尺度目标的检测效果。

Inner channel of the head	Remove SE	Cosine learning rate decay	Learning rate Warm-up	Precision	Recall	HMean	Model Size(M)	Inference Time(CPU, ms)
256				0.6821	0.5560	0.6127	7	406
96				0.6677	0.5524	0.6046	4.1	213
96	√			0.6952	0.5413	0.6087	2.6	173
96	√	√		0.7034	0.5404	0.6112	2.6	173
96	√	√	√	0.7349	0.5420	0.6239	2.6	173

5.1 文本检测优化 (模型优化)

3. 余弦学习率

采用余弦学习率替代分段调整学习率，整个过程中学习率都比较大，因此收敛比较慢，但是最终收敛效果更好。



5.1 文本检测优化（模型优化）

4. 预热学习率

很多研究表明，在训练刚开始，使用太大的学习率，会导致学习过程中数值上不稳定。建议刚开始使用较小的学习率，逐步增加到初始学习率，这样有助于效果提升。

5. SE模块的移除

MobileNetV3结构中有多个SE模块。对于文本检测任务，图中往往有多个文本目标和嘈杂的背景信息，输入分辨率一般都很大，例如640*640，通过SE挖掘特征之间的信息比较困难，同时又增加了计算量。

将SE模块移除，模型大小有所减小，计算时间缩减，同时对Hmean影响微小，因此将SE模块移除是一项不错的优化方法。

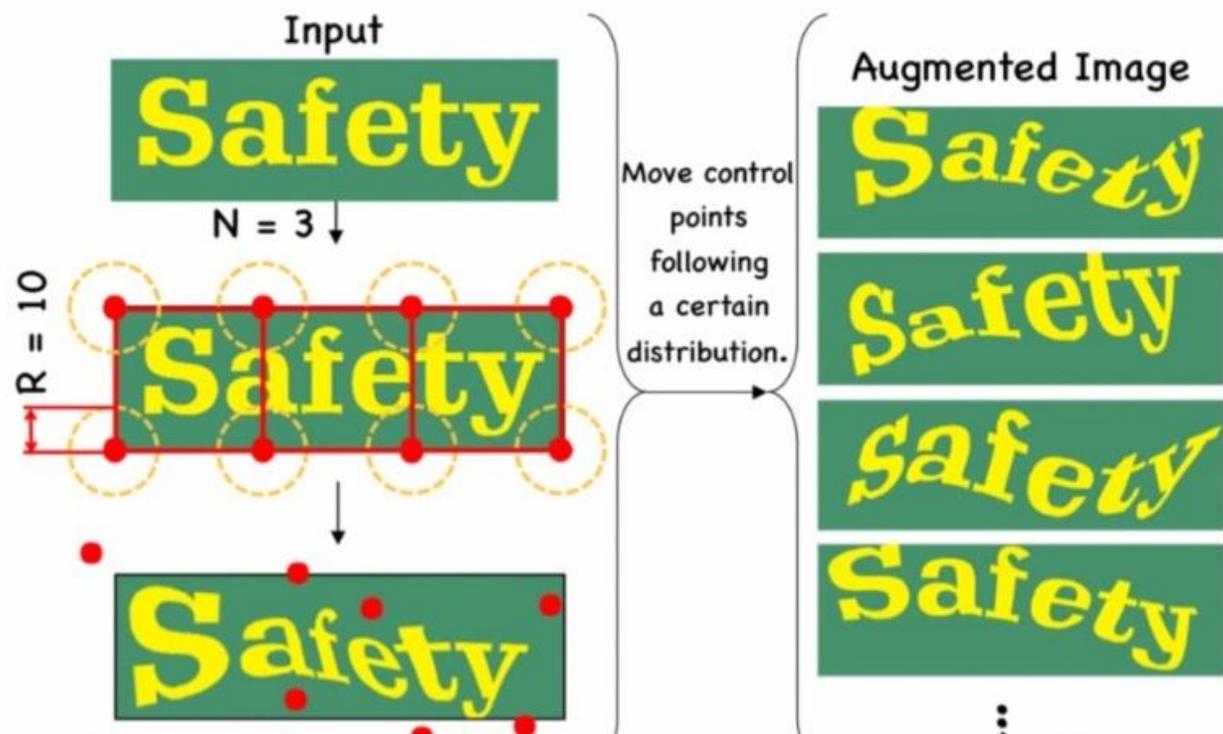
除此之外，另有FPGM模型裁剪器方法，PP-OCR采用FPGM模型裁剪器去进一步减小模型大小。

5.2 文本识别优化

- 在文本识别方面优化，主要有以下策略超轻量骨干网络选择，头部轻量化，余弦学习率，预热学习率，数据增强，增大特征图分辨率，正则化参数，预训练大模型和PACT量化等。

1. 数据增强

数据增强也叫数据扩增，意思是在不实质性的增加数据的情况下，让有限的数
据产生等价于更多数据的价值。



5.2 文本识别优化

2. 增大特征图分辨率

在PP-OCR中，为了保留更多的水平方向和垂直方向图像的信息，将stride变换都从 (2, 2) 改为 (1, 1)，整个网络的特征图的分辨率都变大。实验表明，增大特征图的分辨率，识别准确率有一定提升。

Input	Operator	exp size	#out	SE	NL	s	Stride	Feature Map Resolution
$224^2 \times 3$	conv2d, 3x3	-	16	-	HS	2		
$112^2 \times 16$	bneck, 3x3	16	16	✓	RE	2	→ (2,1)(8*160)	→ (1,1)(16*160)
$56^2 \times 16$	bneck, 3x3	72	24	-	RE	2	→ (2,1)(4*160)	→ (2,1)(8*160)
$28^2 \times 24$	bneck, 3x3	88	24	-	RE	1		
$28^2 \times 24$	bneck, 5x5	96	40	✓	HS	2	→ (2,1)(2*160)	→ (2,1)(4*160)
$14^2 \times 40$	bneck, 5x5	240	40	✓	HS	1		
$14^2 \times 40$	bneck, 5x5	240	40	✓	HS	1		
$14^2 \times 40$	bneck, 5x5	120	48	✓	HS	1		
$14^2 \times 48$	bneck, 5x5	144	48	✓	HS	1		
$14^2 \times 48$	bneck, 5x5	288	96	✓	HS	2	→ (2,1)(1*160)	→ (2,1)(2*160)
$7^2 \times 96$	bneck, 5x5	576	96	✓	HS	1		
$7^2 \times 96$	bneck, 5x5	576	96	✓	HS	1		
$7^2 \times 96$	conv2d, 1x1	-	576	✓	HS	1		
$7^2 \times 576$	pool, 7x7	-	-	-	-	1	→ (2,2)(1*80)	→ (2,2)(1*80)



3. 正则化参数

正则化一般是为了防止过拟合，添加到损失函数中，使用正则化使整个网络权重值倾向变小，从而提升模型的泛化能力。（识别准确率）

4. 预训练大模型

在图像分类、目标检测和图像分割中，使用ImageNet 1000训练的图像分类预训练模型，有助于模型快速收敛和效果提升。

5. PACT量化

PACT 量化可以减小模型中的冗余，可以减小模型大小，缩短识别时间，从而提升模型效果。



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY



Demo展示



Paddle OCR是一款轻量的OCR工具库

- 支持中英文数字组合识别、竖排文本识别、长文本识别
- 支持多语言识别丰富易用的OCR相关工具组件
- 支持用户自定义训练，提供丰富的预测推理部署方案
- 支持PIP快速安装使用
- 可运行于Linux、Windows、MacOS等多种系统



Paddle OCR自带轻量级预训练模型

- PP-OCRv2系列

检测 (3.1M) + 方向分类器 (1.4M) + 识别 (8.5M) = 13.0M

- PP-OCR mobile移动端系列

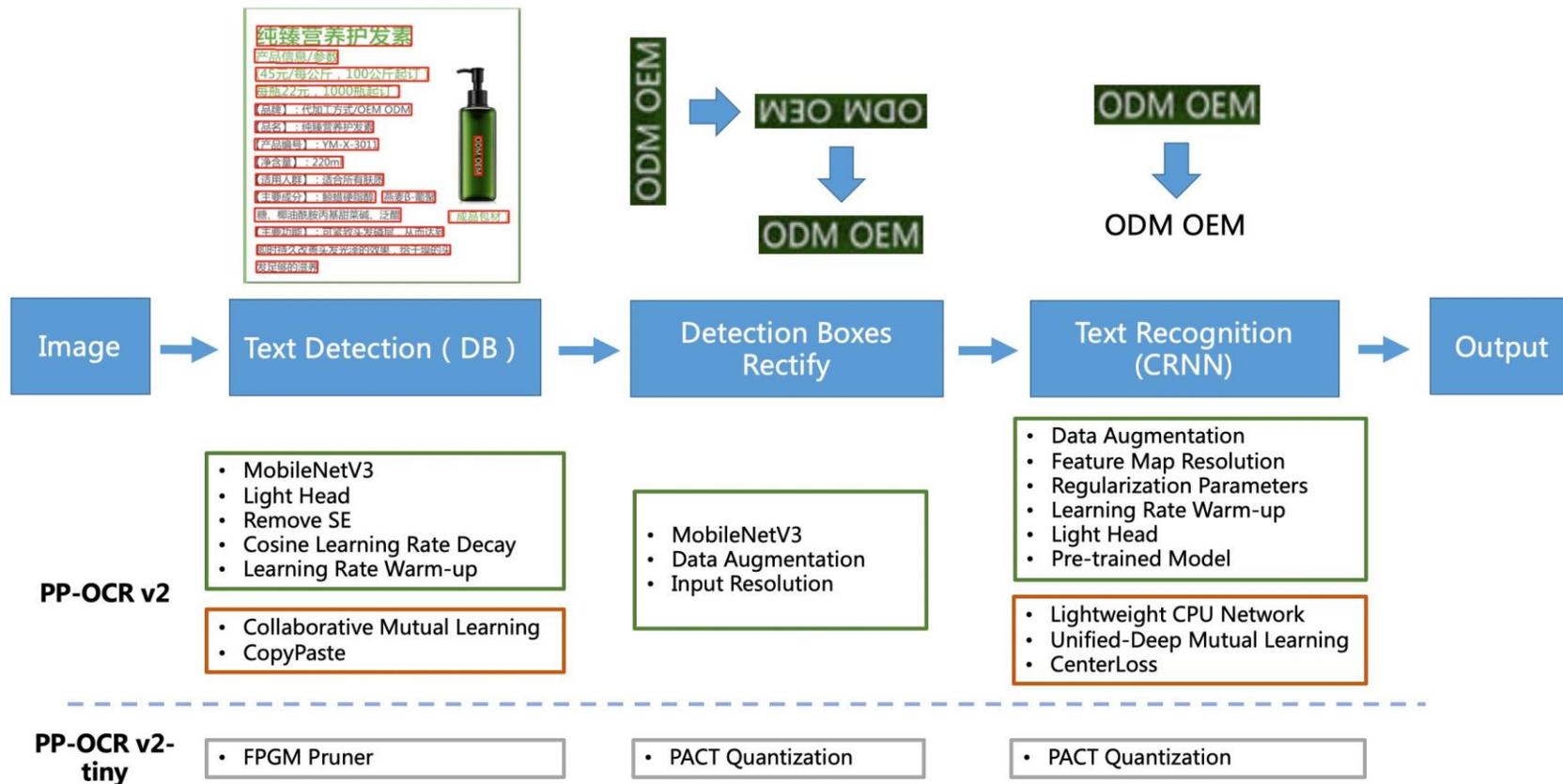
检测 (3.0M) + 方向分类器 (1.4M) + 识别 (5.0M) = 9.4M

- PPOCR server系列

检测 (47.1M) + 方向分类器 (1.4M) + 识别 (94.9M) = 143.4M

- PP-OCR 主要由 DB 文本检测、检测框矫正和CRNN文本识别三部分组成。
- 该系统从骨干网络选择和调整、预测头部的的设计、数据增强、学习率变换策略、正则化参数选择、预训练模型使用以及模型自动裁剪量化8个方面，采用19个有效策略，对各个模块的模型进行效果调优和瘦身，最终得到整体大小为3.5M的超轻量中英文OCR和2.8M的英文数字OCR。
- PP-OCRv2在PP-OCR的基础上，进一步在5个方面重点优化，检测模型采用CML协同互学习知识蒸馏策略和CopyPaste数据增广策略；
- 识别模型采用LCNet轻量级骨干网络、UDML改进知识蒸馏策略和Enhanced CTC loss损失函数改进，进一步在推理速度和预测效果上取得明显提升。

PP-OCRv2 Pipeline



```
ppocr_test.py ×
ppocr_test.py > ...
1  from paddleocr import PaddleOCR, draw_ocr
2
3  filename='germanad.png'
4  # You can set the parameter `lang` as `ch`, `en`, `fr`, `german`, `korean`, `japan`
5  ocr = PaddleOCR(use_angle_cls=True, lang='german', use_gpu=False)
6  img_path = './ppocr_img/my_imgs/' + filename
7  result = ocr.ocr(img_path, cls=True)
8  for line in result:
9      | print(line)
10
11
12  # draw result
13  from PIL import Image
14  image = Image.open(img_path).convert('RGB')
15  boxes = [line[0] for line in result]
16  txts = [line[1][0] for line in result]
17  scores = [line[1][1] for line in result]
18  im_show = draw_ocr(image, boxes, txts, scores, font_path='./fonts/simfang.ttf')
19  im_show = Image.fromarray(im_show)
20  im_show.save('./ppocr_img/results/' + filename)
```

中文OCR检测与识别

学院新闻

- 北理工计算机学院组织开展“北理工精神校友谈”纪念活动
- 北理工计算机学院2021级研究生新生开学典礼暨入学教育顺利召
- 北理工计算机学院(唐山研究院)2
- 北理工计算机学院顺利完成2021级研究生新生迎新工作
- 北理工王国仁、刘驰团队获CCF-A类顶会KDD 2021最佳

通知公告

- 北京理工大学计算机学院2021年研究生学业奖学金评审细则
- 计算机学院2021年研究生国家奖学金拟推荐人选公示通知
- 北京理工大学计算机学院2022年接收优秀应届本科毕业生推荐免
- 北京理工大学计算机学院2021年研究生国家奖学金评审细则
- 北京理工大学计算机学院2022年推荐优秀应届本科毕业生免试

- 1: 学院新闻 0.992
- 2: 《北理工计算机学院组织开展“北理工精神校友谈”纪念活动》 0.956
- 3: 《北理工计算机学院2021级研究生新生开学典礼暨入学教育顺利召》 0.968
- 4: 《北理工计算机学院(唐山研究院)2》 0.974
- 5: 《北理工计算机学院顺利完成2021级研究生新生迎新工作》 0.94
- 6: 《北理工王国仁、刘驰团队获CCF-A类顶会KDD 2021最佳》 0.935
- 7: 通知公告 0.999
- 8: 《北京理工大学计算机学院2021年研究生学业奖学金评审细则》 0.976
- 9: 《计算机学院2021年研究生国家奖学金拟推荐人选公示通知》 0.972
- 10: 《北京理工大学计算机学院2022年接收优秀应届本科毕业生推荐免》 0.923
- 11: 北京理工大学计算机学院2021年研究生国家奖学金评审细则 0.9
- 12: 《北京理工大学计算机学院2022年推荐优秀应届本科毕业生免试》 0.900

中文OCR检测与识别

热点新闻

习近平心中的美丽中国

“我们的初心和梦想历久弥坚”告台湾同胞书

延乔路，又被鲜花包围了重新吹响集结号

【祖国颂·节日市场探访记】休闲体育成时尚

全国铁路预计5日发送旅客1050万人次祖国颂专题

【中国共产党人的精神谱系】关键时刻不退缩

因为他们，这座“冰城”变“暖城”

中国馆亮相迪拜世博会

长征七号遥四运载火箭末级残骸已再入大气层

这场“东方之约”，等你来

“中国梦”主题新创作歌曲展专题：“十一”黄金周

国庆假期第五天全国道路交通平稳有序6日迎来客流返程高峰

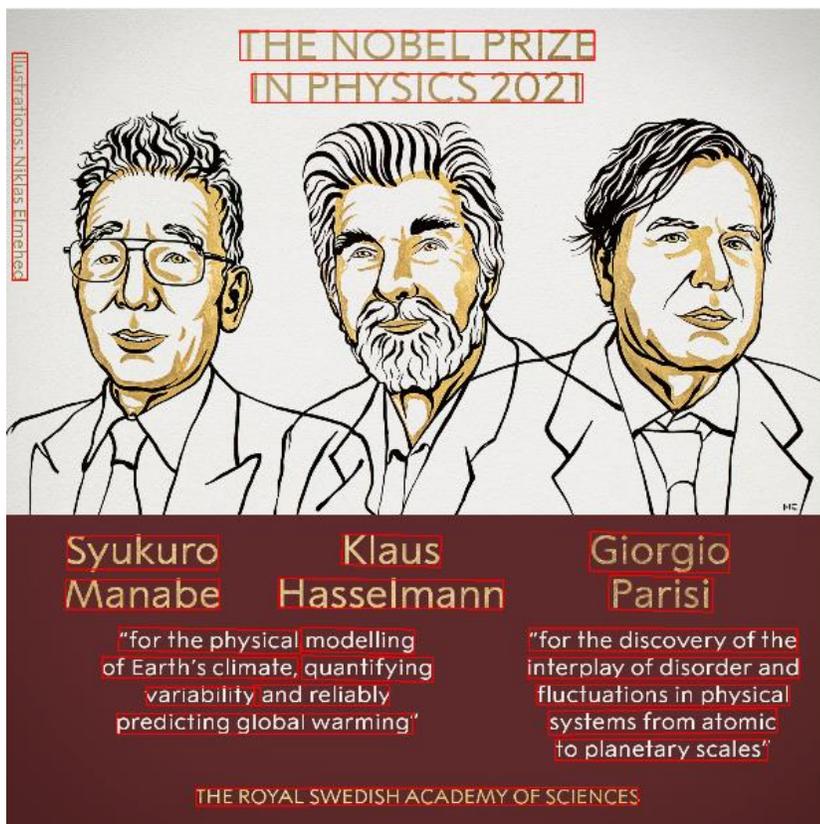
- 1: 热点新闻 0.993
- 2: ·习近平心中的美丽中国 0.949
- 3: “我们的初心和梦想历久弥坚”告台湾同胞书 0.952
- 4: ■延乔路，又被鲜花包围了重新吹响集结号 0.888
- 5: ! (祖国颂·节日市场探访记) 休闲体育成时尚 0.916
- 6: 1全国铁路预计5日发送旅客1050万人次祖国颂专题 0.939
- 7: ! (中国共产党人的精神谱系) 关键时刻不退缩 0.908
- 8: ■因为他们，这座“冰城”变“暖城” 0.914
- 9: ·中国馆亮相迪拜世博会 0.958
- 10: ·长征七号遥四运载火箭末级残骸已再入大气层 0.919
- 11: 这场“东方之约”，等你来 0.885
- 12: ““中国梦”主题新创作歌曲展专题：“十一”黄金周 0.835
- 13: 国庆假期第五天全国道路交通平稳有序6日迎来客流返程高峰 0.978

■ 英文OCR检测与识别



1: THAT'S 0.923
2: WHAT 0.985
3: pepsi 0.970
4: I LIKE 0.846

英文OCR检测与识别

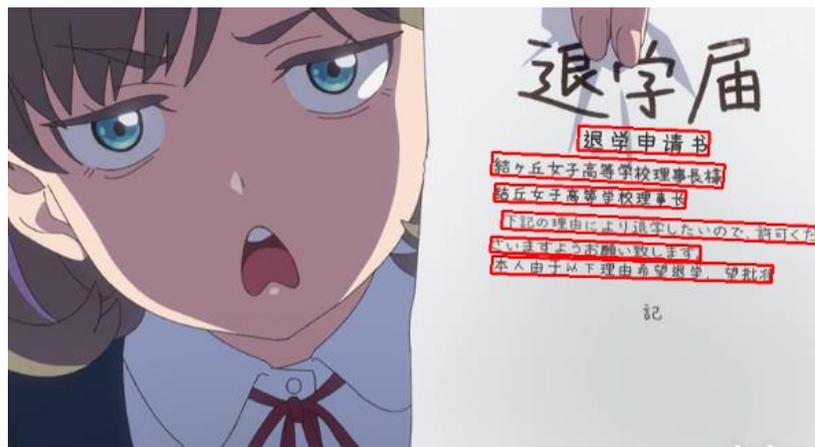


- 1: THE NOBEL PRIZE 0.944
- 2: Illustrations:Niklas Elmehed 0.923
- 3: INPHYSICS2021 0.992
- 4: Klaus 0.958
- 5: Syukuro 0.965
- 6: Giorgio 0.987
- 7: Hasselmann 0.926
- 8: Manabe 0.994
- 9: Parisi 0.973
- 10: "for the physical 0.928
- 11: modelling 0.992
- 12: "for the discovery of the 0.968
- 13: of Earth's climate, 0.969
- 14: quantifying 0.968
- 15: interplay of disorder and 0.971
- 16: 'and reliably 0.903
- 17: variability 0.955
- 18: fluctuations in physical 0.977
- 19: predicting global warming 0.948
- 20: systems from atomic 0.981
- 21: fo planetary scales' 0.947
- 22: THE ROYAL SWEDISH ACADEMY OF SCIENCES 0.962

■ 其他语言OCR检测与识别



- 1: ZvlHalloxeenmvtt 0.983
- 2: Keuchvp□fflie à en 0.997
- 3: m 0.824



- 1: 銀学申估も 0.545
- 2: 第ヶ丘女子高号学枚理専様 0.674
- 3: 丘安寺高野单叔理中状 0.563
- 4: 下記の理由により乗学したいので、古守くた 0.839
- 5: 当いますようお願いいたします。 0.828
- 6: 本人由丁込下理由希製銀学配北刑 0.621



北京理工大学
BEIJING INSTITUTE OF TECHNOLOGY

OCR及领域优化

谢谢观看