

# OCR实战

## 第一课 OCR技术概览

# 课程介绍

## 第一阶段 一览OCR技术和字符识别

### 第一课 OCR技术概览



知识点1：什么是OCR

知识点2：常见应用概览，比如印刷体识别：文档识别、车牌识别、证照/名片/票据识别、视频/图像理解、拍照搜题等，以及手写识别：在线识别、离线识别

知识点3：基本流程：预处理、版面分析、文本行定位、字符识别、后处理

知识点4：常见OCR工具：Tesseract、Abbyy、百度OCR API、科大讯飞OCR API

**实战项目：**Tesseract 搭建第一个OCR引擎

### 第二课 单字符分割与识别



知识点1：单字符识别（模板匹配、特征提取 + 传统分类器、CNN模型）

知识点2：字符分割1：定位候选分割位置（基于投影规则的候选位置提取、基于模型的候选位置提取）

知识点3：字体分割2：分割路径选择（Viterbi算法 / Beam Search、PCFG / 2D-PCFG）

**实战项目：**基于CNN的单字符识别与过分割

### 第三课 字符序列识别



知识点1：概览RNN以及LSTM + CTC、CRNN、RARE

知识点2：概览注意力机制Attention，以及DRAM / DRAW、Transformer

**实战项目：**实战CNN和RNN的综合体CRNN

# 课程介绍

## 第二阶段：掌握文本定位与文本检测

### 第四课 文本行定位



知识点1：传统方法（基于投影的文本行定位、基于最小生成树的文本行定位）

知识点2：深度学习（Full-Page Text Recognition: Learning Where to Start and When to Stop, Learning Text-Line Localization with Shared and Local Regression Neural Networks, TextSnake）

**实战项目：** 实战Text Line Extraction Based on MST

### 第五课 自然场景中的文本检测



知识点1：Reading Text in the Wild with Convolutional Neural Networks

知识点2：CTPN、RRPN、FTSN、DMPNet、EAST

知识点3：SegLink、PixelLink

知识点4：Textboxes、WordSup、FOTS

**实战项目：** 实战场景文本检测之CTPN算法

# 课程介绍

## 第三阶段 掌握其他图像问题

### 第六课 图像质量增强和预处理

知识点1：图像增强（去模糊、超分辨率重建）

知识点2：二值化（全局阈值、局部阈值：克服局部光照和噪声、基于学习的方法）

知识点3：旋转 / 扭曲变形（角度估计与矫正、扭曲复原与DocUNet）

**实战项目：**实战生成对抗网络GAN的变体SRGAN



# 第一课 OCR技术概览

- 什么是OCR
  - OCR技术的使用场景
  - 目前OCR的技术难点
- OCR基本流程
  - 实例一：通用文本识别
  - 实例二：车牌检测与识别
- 第一个OCR小程序
  - TESSERACT

# 什么是OCR

- OCR，光学字符识别（OPTICAL CHARACTER RECOGNITION），作为计算机视觉领域的经典问题之一，指对图像中的文字进行识别，并获取文本结果。常见于拍照检查、文档识别、证照票据识别、车牌识别、自然场景文本定位识别等，相关技术在数字时代得到了广泛应用。
- 近年，随着深度学习技术在视觉领域的不断发展，OCR也迅速突破了传统技术框架的瓶颈，作为新的研究热点展现出更多活力和广阔的应用场景

# 第一课 OCR技术概览

- 什么是OCR
  - OCR技术的使用场景
  - 目前OCR的技术难点
- OCR基本流程
  - 实例一：通用文本识别
  - 实例二：车牌检测与识别
- 第一个OCR小程序
  - TESSERACT

# OCR的典型应用场景

## 第一章 食品与化学

食品通常是指经过加工制作可以供人食用的物质。食品的发展有着悠久而丰富的内涵,它深深植根于人们的日常饮食生活中。人类的生存离不开食品,它是人类进行物质联系并赖以生存的基础,是人类维持生命活动的重要物质。

社会发展到今天,人类对食品有了更全面、更深层的认识。人们开始从健康养生、科学的角度注重饮食生活,对因饮食不当等原因而造成心脏病、糖尿病等各病的现象已引起充分重视。

### 第一节 概 述

《食品安全法》第九十九条对“食品”的定义:食品,指各种供人食用或者饮用的原料以及按照传统既是食品又是药品的物品,但是不包括以治疗为目的的物品。  
《工业基本术语》对食品的定义:可供人类食用或饮用的物质,包括加工食品、半成品和未加工食品,不包括烟草或只作药品用的物质。

从食品卫生立法和管理的角度看,广义的食品概念还涉及:所生产食品的原料,食品原料种植、养殖过程接触的物质和环境,食品的添加物质,所有直接或间接接触食品的包装材料、设施以及影响食品原有品质的环境。

### 不同功效组分的功能饮料抗疲劳作用的研究

蔡愈杭<sup>1</sup>, 荣耀中<sup>1</sup>, 罗殷<sup>1</sup>, 程能能<sup>2</sup>, 陈伟<sup>1</sup>

1. 上海东锦食品集团有限公司(上海 201812); 2. 复旦大学药学院(上海 201203)

**摘要** 比较不同功效组分的功能饮料抗疲劳作用。将雄性昆明小鼠按体重随机分组, 分别灌胃不同剂量不同功效组分的饮料30 d, 之后观察各组小鼠体重的变化和负重游泳情况。三款功能饮料均能够延长小鼠负重游泳时间、持续游泳生存率。三款功能饮料均具有抗疲劳作用, 但其中以氨基酸、瓜拉纳为主要功效组分的氨基酸功能饮料抗疲劳效果最佳。

**关键词** 功能饮料; 游泳时间; 生存率; 抗疲劳

### Study on the Anti-fatigue Effect of Functional Drink Contained Different Effectiveness

Cai Yu-hang<sup>1</sup>, Rong Yao-zhong<sup>1</sup>, Luo Yin<sup>1</sup>, Cheng Neng-neng<sup>2</sup>, Chen Wei<sup>1</sup>

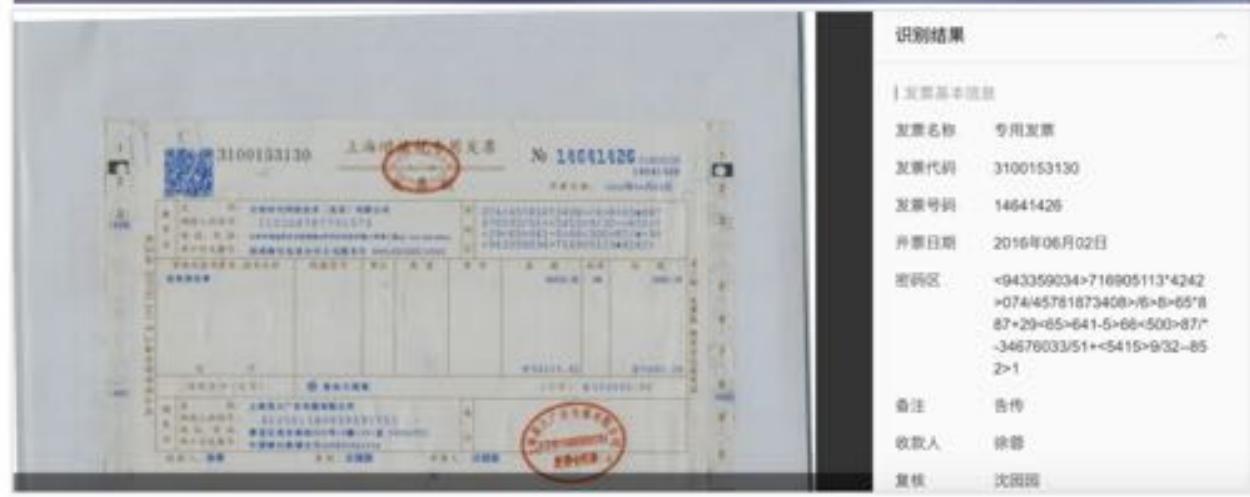
1. Shanghai Tohkin Food Group Co. Ltd (Shanghai 201812); 2. School of Pharmacy, Fudan University (Shanghai 201203)

**Abstract** To study the anti-fatigue effect of functional drink contained different effectiveness, the KM male mice were randomly divided into groups according to their body weight. The mice were administered respectively with water and different functional drink within different dosage. After 30 d, the increase of mice weight and the forced swimming test were determined. All of the functional drink prolonged the swimming time of mice and improved their' survival rate. All of the functional drink had an anti-fatigue effect, which contain amino acid and guarana functional drink was the best.

**Keywords** functional drink; swimming time; survival rate; anti-fatigue

# OCR的典型应用场景

证照票据识别



# OCR的典型应用场景

车牌识别



# OCR的典型应用场景

拍照搜题



## 我的題目

题目详情	(1) 3x
<p>Wang Lin is a good <u>1</u>. She is twenty-one years old. She drives a car in a <u>2</u>. She <u>3</u> from Sunday to Friday. Her home isn't <u>4</u> the factory. She gets up <u>5</u> at six. She goes to work at 6:20. <u>6</u> 7:30 she must get there. She <u>7</u> in the factory. She has lunch there. She <u>8</u> the factory at 1:00 in the afternoon. She cooks and then does housework in the evening. She likes <u>9</u> very much. She <u>10</u> to bed at about 10:30 pm.</p>	(2) 11
2013年01月04日 06:30	(3) x
	搜索结果
	题目

### 完形填空。

Wang Lin is a good 1. She is twenty-one years old. She drives a car in a 2. She 3 from Sunday to Friday. Her home is not 4 the factory. She gets up 5 at six.

(7)  $m^2 + m^2 \times m_1$

6. 计算:

(1)  $(2x^2)^3 - 6x^2(x^4 + 2x^2 + x)$

(3)  $[(x+y)^2 - (x-y)^2] + 2xy$

7. 求下列各式的值:

(1)  $3x^2 + (-\frac{3}{2}x + \frac{1}{3}y^2)(2x - \frac{2}{3}y)$ , 其中  $x = -\frac{1}{3}$ ,  $y = \frac{2}{3}$ ;

(2)  $[(xy+2)(xy-2) - 2x^2y^2 + 4] + xy$ , 其中  $x = 10$ ,  $y = -\frac{1}{25}$ ;

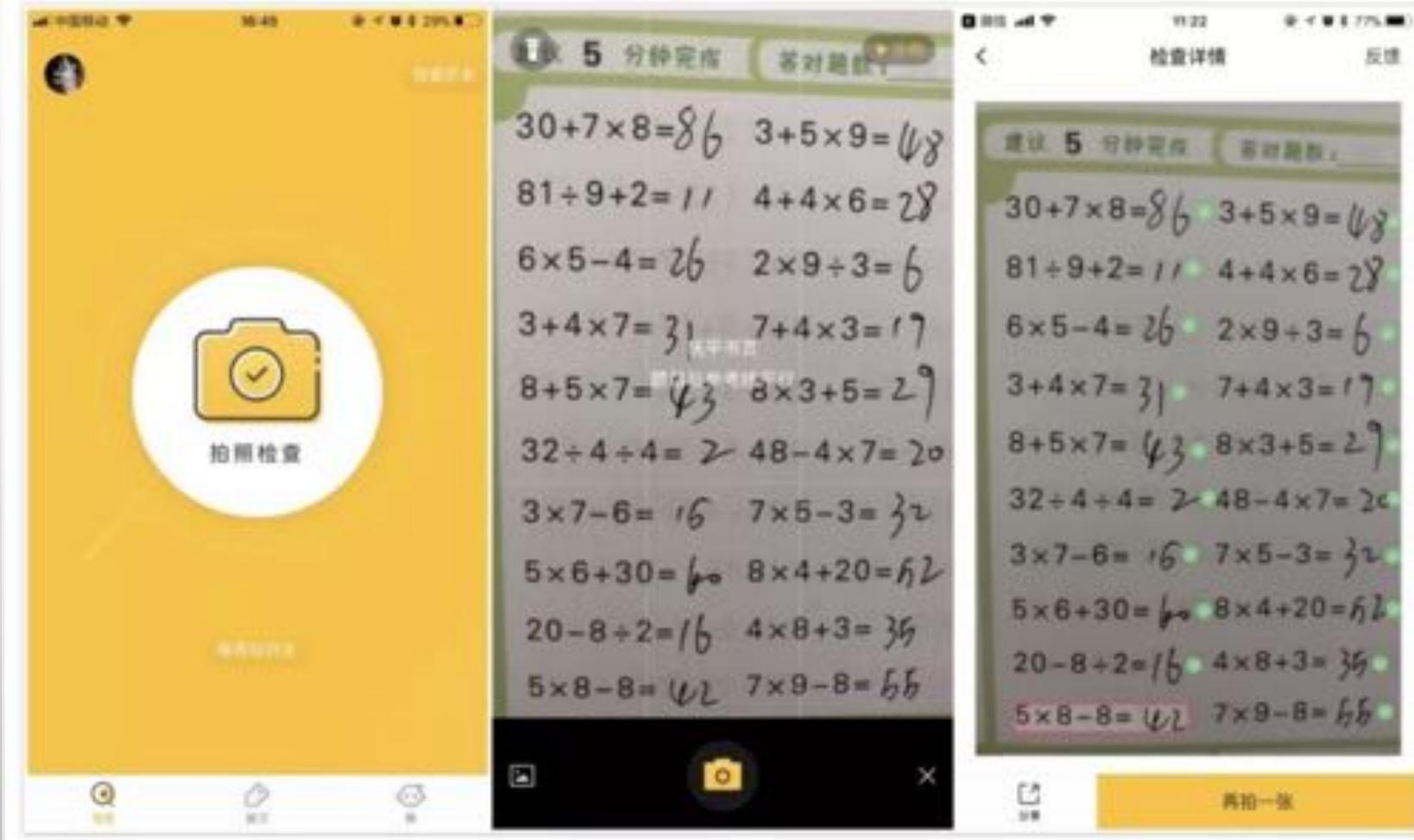
(3)  $x(x+2y) - (x+1)^2 + 2x$ , 其中  $x = \frac{1}{25}$ ,  $y = -25$ .

下列词语中加红词语运用不恰当的是

A: 不知道什么时候，出现了一个神通广大的女神，叫做女娲。

# OCR的典型应用场景

拍照检查



# OCR的典型应用场景

自然场景文字识别

image



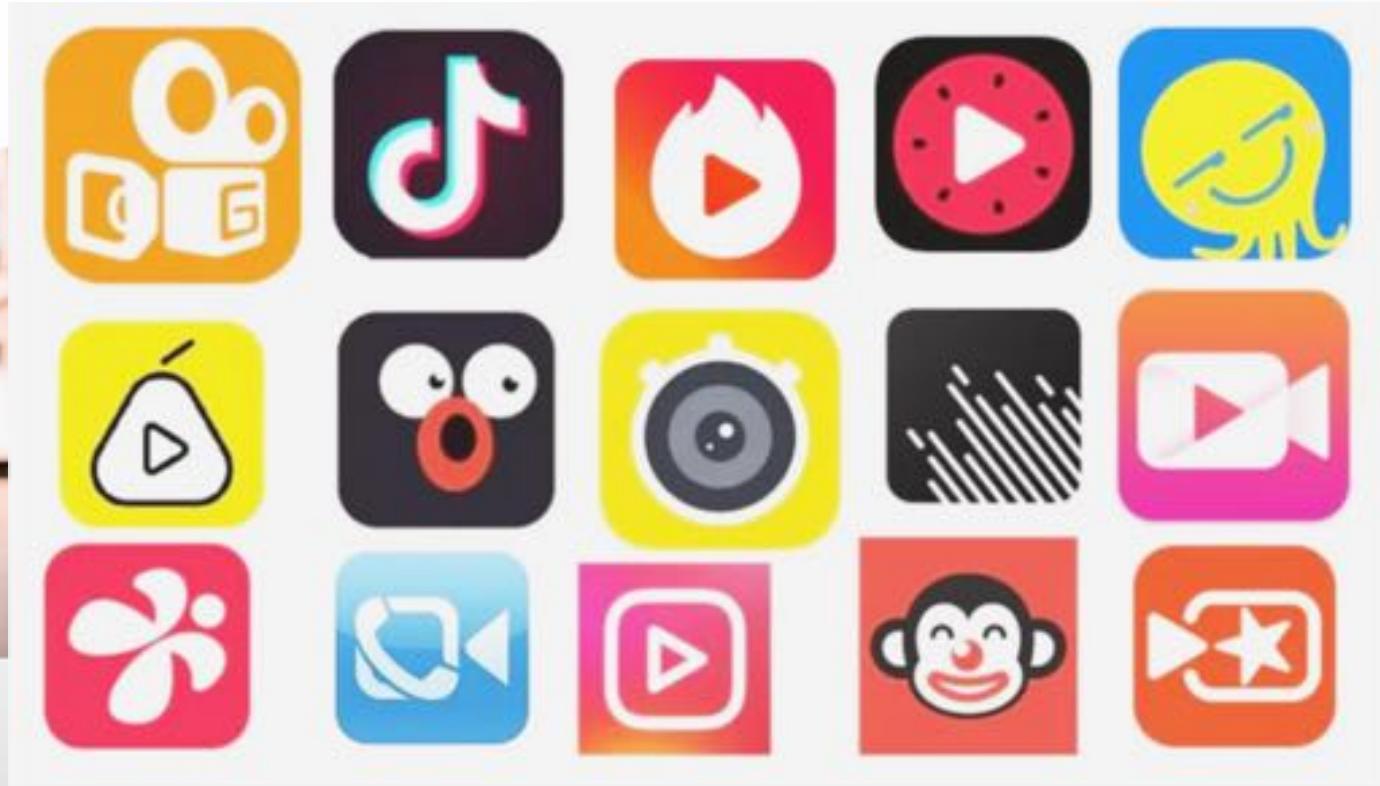
annotation

公牛	烟酒百货超市
公牛安全插座	送货电话
烟酒	烟
数字提取技术	批发雪糕
店	公用电话
槟榔	手机
水饺	充值
虾丸	超市
鱼丸	
购物中心	大卖场
手机	
购物精彩	



# OCR的典型应用场景

内容审核



# OCR的典型应用场景

- 内容理解



# 第一课 OCR技术概览

- 什么是OCR
  - OCR技术的使用场景
  - 目前OCR的技术难点
- OCR基本流程
  - 实例一：通用文本识别
  - 实例二：车牌检测与识别
- 第一个OCR小程序
  - TESSERACT

## 常见挑战

- 长期以来，丰富的排版样式、复杂背景、多变的自然场景、恶劣的光照和分辨率环境、文本扭曲变形、手写识别、公式识别、多语言混排等问题，仍然是学界和工业界试图攻克的难点。

# 常见挑战

- 复杂版式

DST Model

Delexicalisation-Based Model

Delexicalisation-Based Model + Semantic Dictionary

NEURAL BELIEF TRACKER: NBT-DNN

NEURAL BELIEF TRACKER: NBT-CNN

DSTC2		WOZ 2.0	
Goals	Requests	Goals	Requests
69.1	95.7	70.8	87.1
72.9*	95.7	83.7*	87.6
72.6*	96.4	84.4*	91.2*
73.4*	96.5	84.2*	91.6*

Table 1: DSTC2 and WOZ 2.0 test set accuracies for: a) joint goals; and b) turn-level requests. The asterisk indicates statistically significant improvement over the baseline trackers (paired *t*-test;  $p < 0.05$ ).

indicates that future work should focus on ASR compensation if the model is to perform well in environments with challenging

## CONTENTS 目录

### 第一章 食品与化学

第一节 概述	001
第二节 食物中的营养成分	012
第三节 常见食物的化学成分及贮存	049
第四节 饮品与化学	052
第五节 食品添加剂	075
第六节 毒物与化学	088
第七节 健康与化学	110



## Leaders mark China's progress

Top officials affirm commitment to nation's openness and engagement with the world

By SHI YUQING (CHINA DAILY)

The leaders of the People's Republic of China are making their annual visit to the Great Hall of the People in Beijing to mark the 70th anniversary of the founding of the People's Republic of China. The meeting is a traditional state occasion, featuring speeches by Chinese leaders and foreign guests, and the leaders will issue an address to the nation.



## Telecom scammers will face harsher punitive measures

By CHEN JIANGHUA

china@chinadaily.com.cn

“Starting from this year, telecom scammers will face harsher punitive measures,” said Wang Jun, a spokesman of the Ministry of Public Security.

Wang told reporters he plans to banish the practice of “telecom fraud” from China, and vowed to crack down on such illegal activities.

“We will impose heavier penalties on telecom scammers and punish them more strictly than before,” said Wang.

He also called for a ban on illegal telecommunications, which have been rampant across the country.

“We will prohibit the illegal sale of mobile phones and illegal sales of personal information, which can easily lead to telecom fraud,” said Wang.

“We will implement strict laws and regulations to prevent telecom scammers from using mobile phones to commit crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

“We will also ban the use of mobile phones to commit crimes, such as telecom fraud, and we will increase the punishment for such crimes,” said Wang.

# 常见挑战

## • 扭曲形变

We all have a teacher we love. For Markus, that teacher is Mr. J. When Mr. J walked into his classroom earlier this week, he reportedly discovered a letter, addressed to him, on his desk. "Thank you for being a kind teacher and for being amazing! ... — Markus"

In his short message, Markus explained that he saw Mr. J as a father. "I look on you as my dad because you treat me as if I were your son. You always feed me when I am hungry and hug me when I am sad." While the boy couldn't stay in the fifth grade forever, Markus didn't want Mr. J to think that he would forget his teacher. "I'll never forget you, Mr. J," he wrote.

Mr. J shared the photo of Markus' letter on the Internet. "I tried so hard to fight back the

people living in Nuevo Saposoa, a remote village in Peru. However, things went from bad to worse in March 2015 41 a flood damaged the power station in the area. The villagers were 42 to turn to kerosene (煤油) lamps, which are not only expensive but also 43 because of the poisonous matter they give off.

44, the researchers and students at the Universidad de Ingeniería y Tecnología (UTEC) in Lima, Peru heard about their 45 situation and came

2 下列各句中，没有语病的一项是

- A. 由于莫言获得诺贝尔奖，必然会使得他的作品在短期内销量大增，这毫无疑问，但文学想要再造昔日辉煌，已几乎没有可能。
- B. 国庆期间高速公路免费通行让许多高速公路变成了巨型停车场，虽然，政策制定者事先没能料到这样的结果，他们的初衷并非是这样的。
- C. 最新人口普查结果显示，河南省常住人口总数名列全国第三，达到了9 402.4万人，其中南阳、周口、郑州三城市人口最多，均超过了800万。
- D. 伦敦奥运闭幕式上的“里约8分钟”十分精彩，那一刻，和着欢快

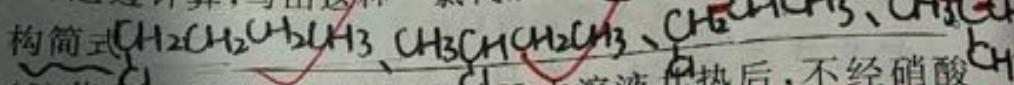
# 常见挑战

- 笔迹干扰 / 手写 / 涂改

4. 某一氯代烷 1.85 g, 与足量  $\text{NaOH}$  小心地共热，硝酸酸化，再加入足量的  $\text{AgNO}_3$  溶液，生成白色沉淀。

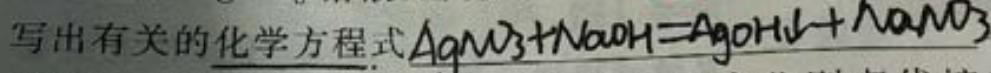
$$2.87 \text{ g.} n(\text{AgI}) = \frac{2.87}{143.5} = 0.02 \text{ mol}$$

(1) 通过计算, 写出这种一氯代烷的各种同分异构体的结构。

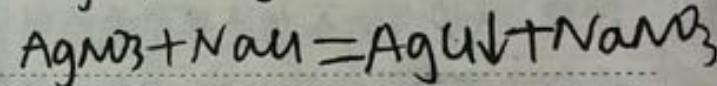
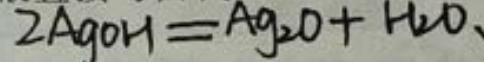


(2) 若此二氯代烷与足量  $\text{NaOH}$  水溶液共热后，不经硝酸

酸化后加  $\text{AgNO}_3$  溶液，出现的现象为 将产生白色沉淀。



(3)能否用硝酸银溶液直接与卤代烃反应来鉴别卤代烷?



不能，古代水钟的计时均是以自然的形式与破坏。

#### 七、四大洲八洋行取回承古用

44. 进食可刺激胃腺细胞分泌胃液，胃液中含有胃酸及胃蛋白酶有利于消化。 胃酸 胃蛋白酶

(1) 胃酸可以杀灭进入胃内的细菌，这属于机体的非特异性免疫。

免疫；胃蛋白酶仅对食物中的蛋白质成分有消化作用，这体现了酶的专一性。

(2) 哺乳动物进食时, 食物尚未进入胃内就可引起胃液分泌。称为头期胃液分泌, 该过程受神经调节和神经-体液调节的共同调控, 如图所示。

①胃泌素通过体液运输到达胃腺细胞，促进胃液分泌。若胃酸分泌过多，又可抑制胃泌素的分泌，这种调节方式叫做反馈调节。

②促进胃腺细胞分泌胃液的信号物质除胃泌素外还有神经递质

(3) 为探究两种调节机制对头期胃液分泌的影响，有人用同一只狗连续进行了以下实验

步骤	实验条件	胃液的分泌量(毫升)	4小时胃液分泌量
1	假饲	++	+++
2	切除胃窦，再假饲	+	+
3	假饲，并注射一定量的胃泌素 (控制胃泌素的注射剂量，使胃液分泌量与步骤1相当)	++	+++
4	只注射与步骤3相同剂量的胃泌素，但不假饲	++	神经体液

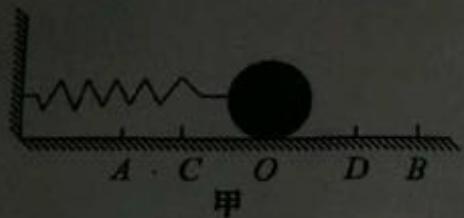
(假饲是让狗吃肉，但吞下的肉从食道上的人工瘘口排出，不进入胃；“+”数量表示胃液分泌量多少)

# 常见挑战

- 不均匀光照 / 反光 / 弱光
- 失焦 / 运动模糊 / 摩尔纹



10. 如图甲所示,一根细长的弹性绳子的一端固定在光滑的桌面上,手握小球把小球拉到C点放手,小球便在A、B之间来回运动,从O开始计时,其有关时刻的数据如下表。已知得 $OD = OC = 7 \text{ cm}$ ,  $DB = 3 \text{ cm}$



(1) 0.2 s 内小球发生的位移为  $\frac{7\pi}{2} \text{ cm}$ , 方向  $\rightarrow$ , 上升运动的末速度为  $v = \frac{\pi}{2} \text{ m/s}$ , 方向  $\uparrow$

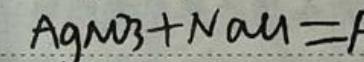
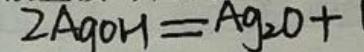
4. 某一氯代烷  $1.85 \text{ g}$ , 与足量  $\text{NaOH}$  水溶液共热, 硝酸酸化, 再加入足量的  $\text{AgNO}_3$  溶液, 生产物质量为  $2.87 \text{ g}$ .  $n(\text{AgCl}) = \frac{2.87}{143.5} = 0.02 \text{ mol}$

(1) 通过计算, 写出这种一氯代烷的各种同分异构简式  $\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_3$ 、 $\text{CH}_3\text{CHCH}_2\text{CH}_3$ 、 $\text{CH}_3\text{CH}_2\text{CH}_2\text{CH}_3$

(2) 若此一氯代烷与足量  $\text{NaOH}$  水溶液共热, 硝酸酸化就加  $\text{AgNO}_3$  溶液, 出现的现象为 不会

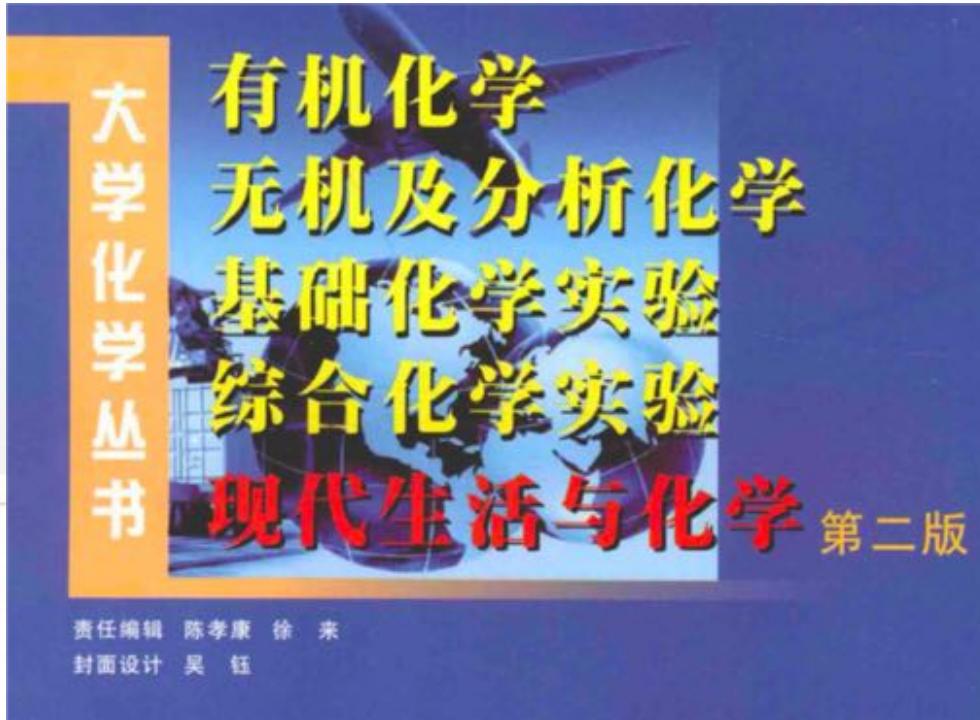
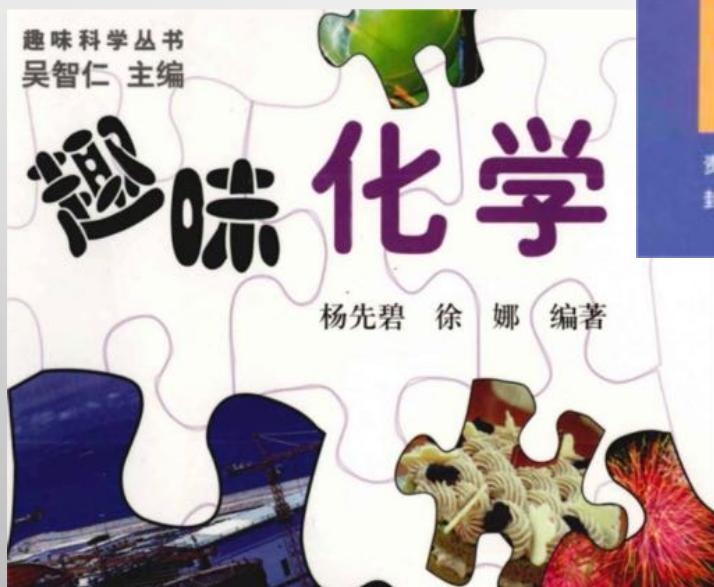
写出有关的化学方程式  $\text{AgNO}_3 + \text{NaOH} \rightarrow \text{AgOH} + \text{NaNO}_3$

(3) 能否用硝酸银溶液直接与卤代烃反应来检验卤代烃? 为什么?



# 常见挑战

- 复杂背景



# 常见挑战

- 多字体 / 多语言混排

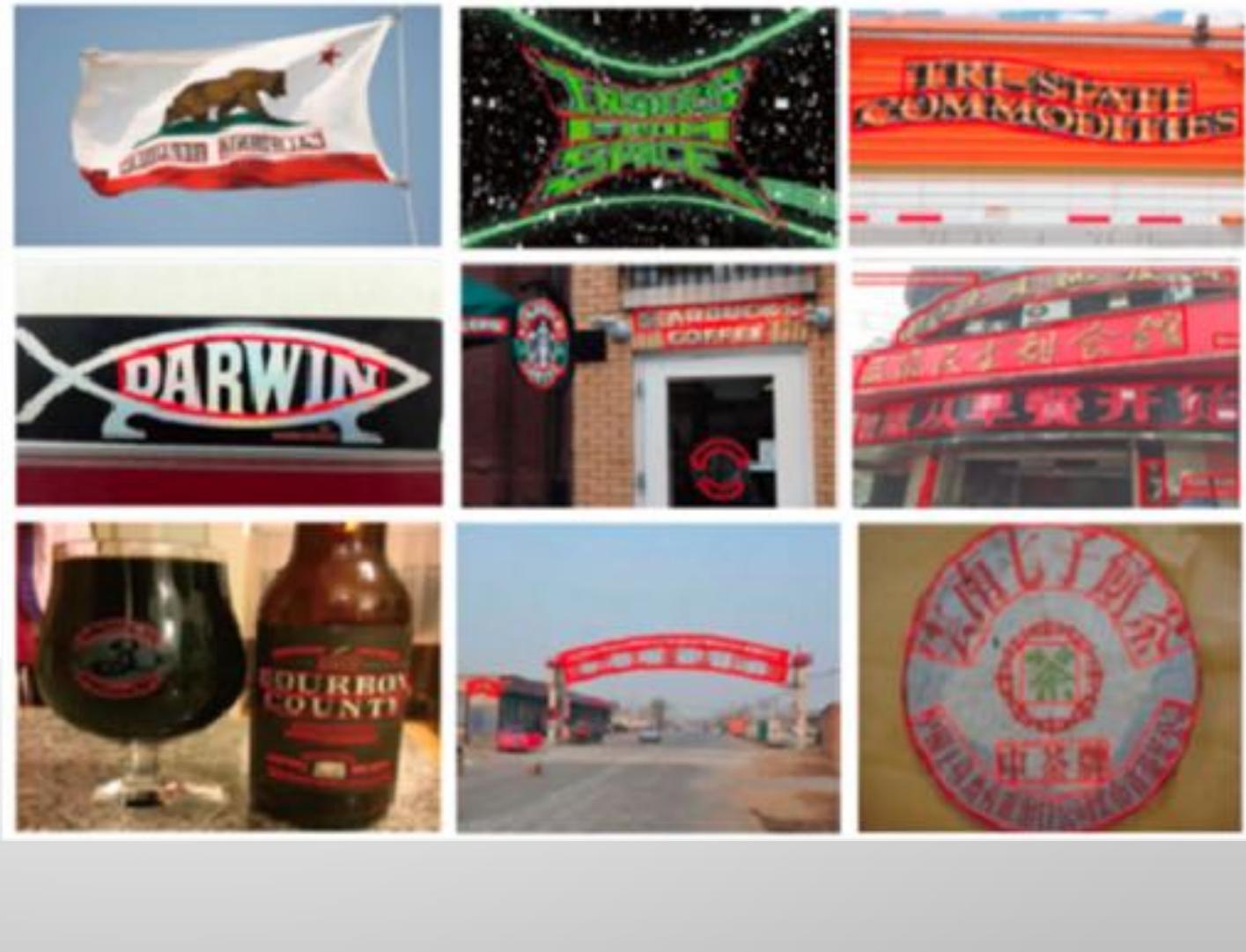


15. 如图 12-2-18,  $AB, CD$  相交于点  $O$ ,  
 $\triangle ACO \cong \triangle BDO$ ,  $CE // DF$ 。试说明:  $CE = DF$ 。



# 常见挑战

- 角度 / 弯曲 / 变形



# ICDAR2019 CHALLENGE

- ROBUST READING CHALLENGE ON ARBITRARY-SHAPED TEXT

- <HTTP://RRC.CVC.UAB.ES/?CH=14&COM=TASKS>

## Awards

The prize for ICDAR 2019-ArT is \$8,700 in total, sponsored by Baidu.

Task 1. Scene Text Detection, \$1,700/\$800/\$400 for top 3 winners.

Task 2. Scene Text Recognition, \$1,700/\$800/\$400 for top 3 winners.

Task 3. Scene Text Spotting, \$1,700/\$800/\$400 for top 3 winners.

- ROBUST READING CHALLENGE ON LARGE-SCALE STREET VIEW TEXT WITH PARTIAL LABELING

- <HTTP://RRC.CVC.UAB.ES/?CH=16&COM=TASKS>

## Awards

The prize for ICDAR 2019-LSVT is \$8,700 in total, sponsored by Baidu.

Task 1. Text detection, \$2,500/\$1,250/\$600 for top 3 winners.

Task 2. End-to-end text spotting, \$2,500/\$1,250/\$600 for top 3 winners.

# 第一课 OCR技术概览

- 什么是OCR
  - OCR技术的使用场景
  - 目前OCR的技术难点
- OCR基本流程
  - 实例一：通用文本识别
  - 实例二：车牌检测与识别
- 第一个OCR小程序
  - TESSERACT

# OCR基本流程

- 预处理
- 版面分析
- 文本行定位
- 字符分割识别
- 后处理

# OCR基本流程

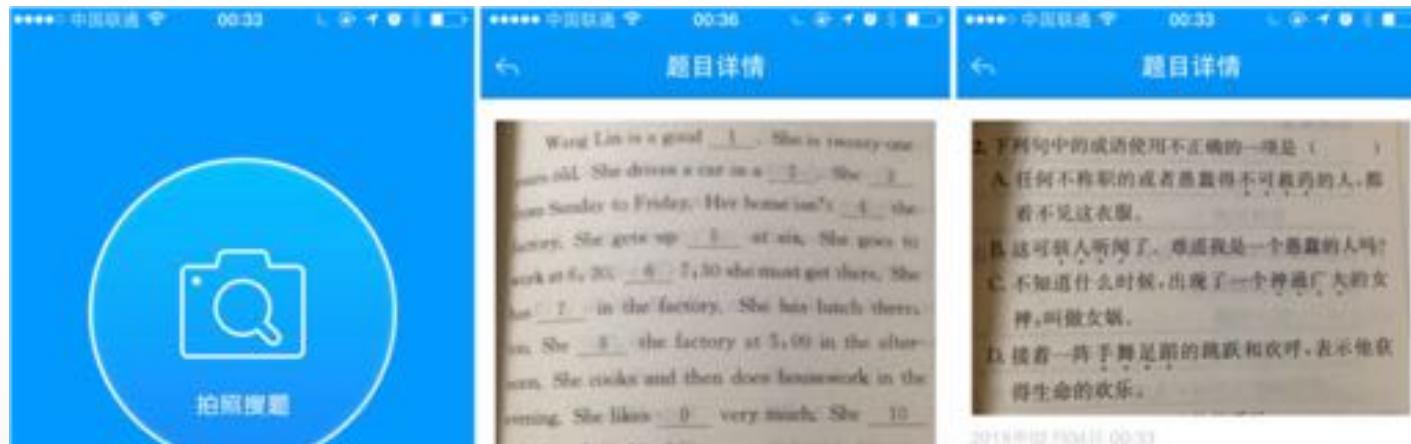
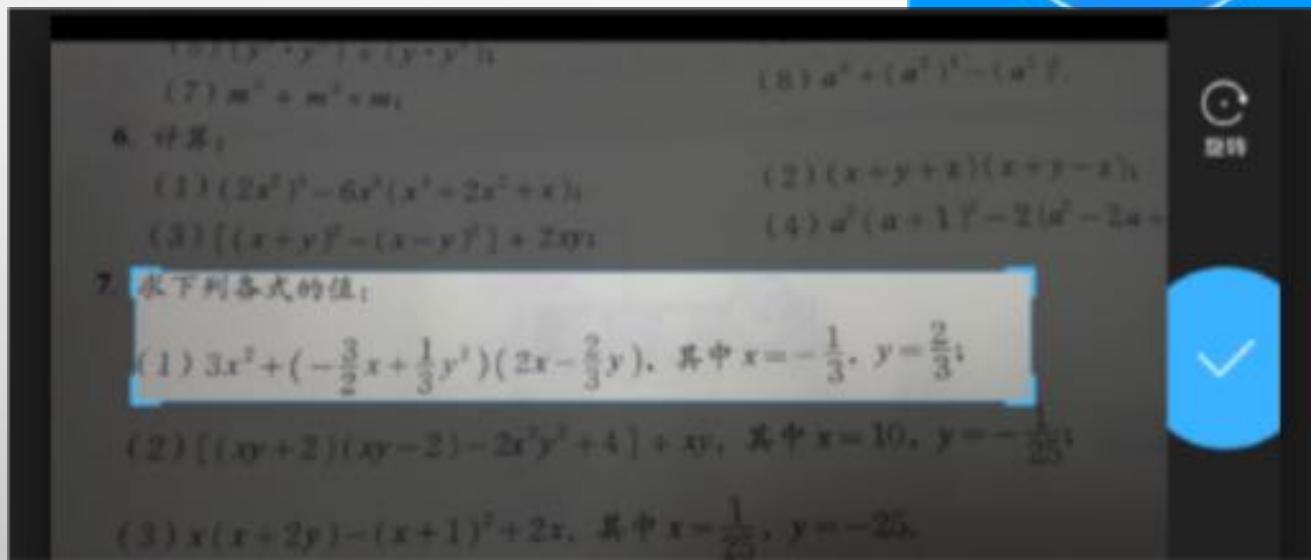
- 预处理
  - 降噪  
滤波、光照处理
  - 增强  
灰度拉伸
  - 二值化  
由灰度图像变成二值图像
  - 倾斜校正  
HOUGH变换、投影法

# 第一课 OCR技术概览

- 什么是OCR
  - OCR技术的使用场景
  - 目前OCR的技术难点
- OCR基本流程
  - 实例一：通用文本识别
  - 实例二：车牌检测与识别
- 第一个OCR小程序
  - TESSERACT

# 实例一

## • 通用文本识别



2018/03/21 10:26

④ 拍照结果

完形填空。

Wang Lin is a good 1. She is twenty-one years old. She drives a car in a 2. She 3 from Sunday to Friday. Her home isn't 4 the factory. She gets up 5 at six. She goes to work at 6:30; 6 7:30 she must get there. She 7 in the factory. She has lunch there, too. She 8 the factory at 5:00 in the afternoon. She cooks and then does housework in the evening. She likes 9 very much. She 10 to bed at about 10:30 pm.

A: 不知道什么时候，出现了一个神通广大的女神，叫做女娲。

B: 任何不称职的或者愚笨得不可救药的人，都看不起这衣服。

C: 不知道什么时候，出现了一个神通广大的女神，叫做女娲。

D: 接着一阵手舞足蹈的跳跃和欢呼，表示他获得了生命的欢乐。

⑤ 答案结果

题干

下列词语中加红词语运用不恰当的是

A: 不知道什么时候，出现了一个神通广大的女神，叫做女娲。

# Binarization & Connected Components Analysis

## 八年级数学上(六三制)

已知：点 C 是  $\angle MAN$  平分线上一点， $\angle BCD$  的两边  $CB, CD$  分别与射线  $AM$  交于 B, D 两点，且  $\angle BCD + \angle MAN = 180^\circ$ . 过点 C 作  $CE \perp AB$ , 垂足为 E.

当点 E 在线段 AB 上时(如图 1), 求证:  $\angle NDC + \angle BCE = 90^\circ$ ;

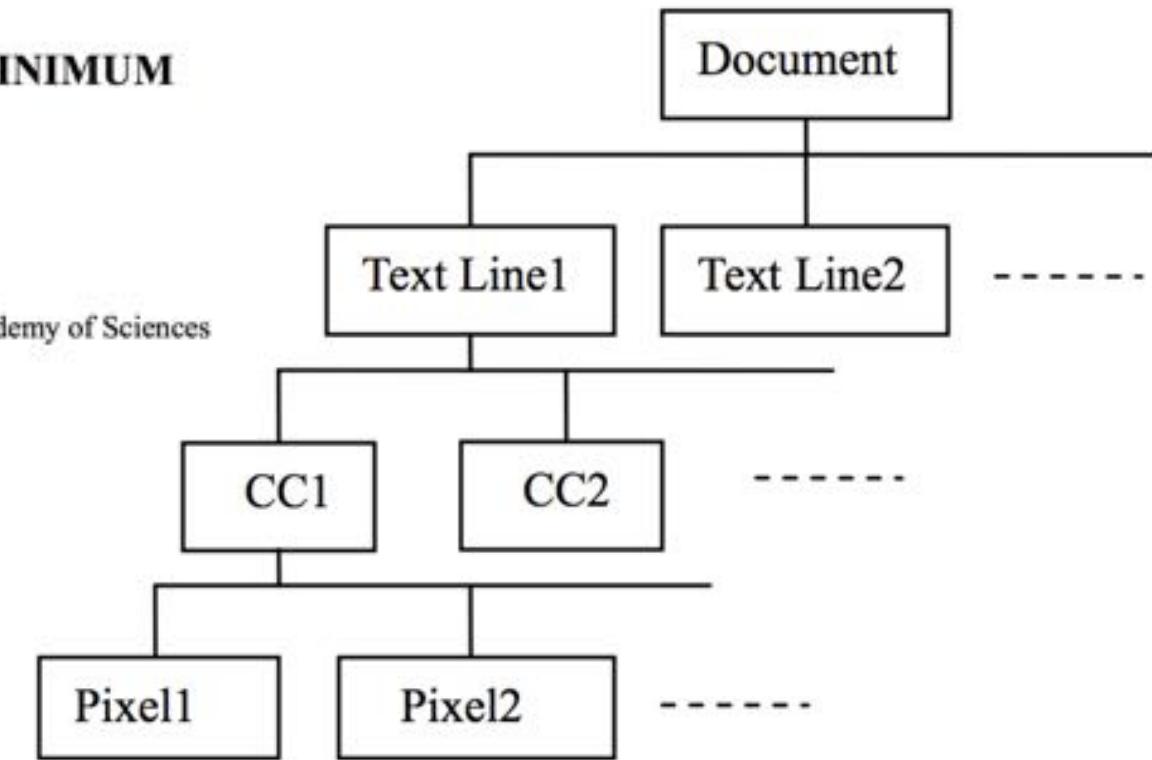
(1) 当点 E 在线段 AB 的延长线上时(如图 2), 求证:  $AB - AD = 2BE$ ;

(2) 在(2)的条件下, 若  $\angle MAN = 60^\circ$ , 连接 BD, 作  $\angle ABD$  的平分线 BF 交 AD 于点 O, 连接 DO 并延长交 AB 于点 G.  $BG = 2, DF = 5$ , 求线段 BD.

## HANDWRITTEN TEXT LINE EXTRACTION BASED ON MINIMUM SPANNING TREE CLUSTERING

FEI YIN, CHENG-LIN LIU

National Laboratory of Pattern Recognition (NLPR), Institute of Automation, Chinese Academy of Sciences  
P.O. Box 2728, Beijing 100080, P.R. China  
E- MAIL: {fyin, liucl}@nlpr.ia.ac.cn



## Distance Measure between Connected Components

- MINIMUM RUN-LENGTH (MRL) DISTANCE
- EUCLIDEAN DISTANCE BETWEEN GRAVITY CENTERS (ECC)

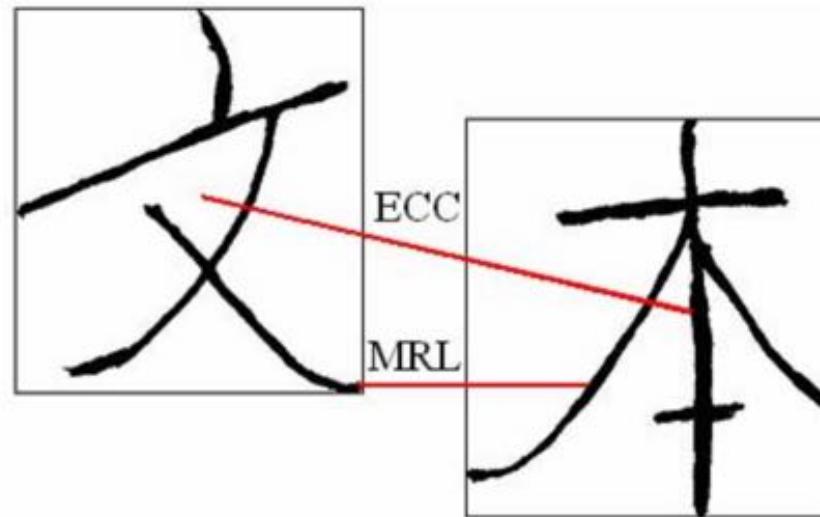


Fig. 3 Example of MRL and ECC

## Distance Measure between Connected Components

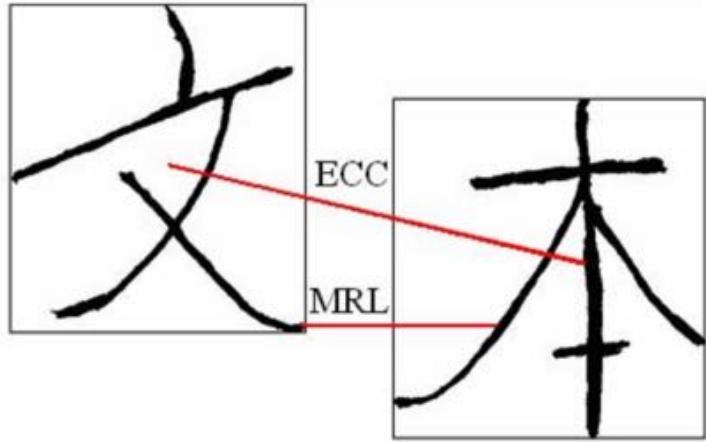


Fig. 3 Example of MRL and ECC

Considering the fact that connected components in the same text line mostly overlap vertically (align horizontally), we combine the two distances MRL and ECC incorporating the degree of vertical overlap. Following [18], the normalized overlap degree is computed as

$$novlp = \frac{1}{2} \left( \frac{ovlp}{h_1} + \frac{ovlp}{h_2} \right) - \frac{vdc}{span},$$

where  $ovlp$  is the overlapping height of two bounding boxes,  $h_1$  and  $h_2$  are the heights of bounding boxes,  $vdc$  is the vertical distance between the centers of bounding boxes, and  $span$  is the spanning height of two bounding boxes. We use  $w=(novlp+1)/2$  as a weight for combining MRL and ECC. Denote the MRL and ECC between two connected components  $CC_i$  and  $CC_j$  as  $MRL(i,j)$  and  $ECC(i,j)$ , the combined distance is computed as

$$D(i, j) = w * MRL(i, j) + (1 - w) * ECC(i, j).$$

$D(i,j)$  is used as the weight of edges in the document graph for computing the MST.

## Distance Measure between Connected Components

八年级数学上(第3课时)

如图，在等腰 $\triangle ABC$ 中， $AB=AC$ ， $BC=4$ ， $AD$ 是底边的中线， $\angle BAC=130^\circ$ ，过点 D 作  $DE \perp AB$ ，垂足为 E，求  $DE$  的长。

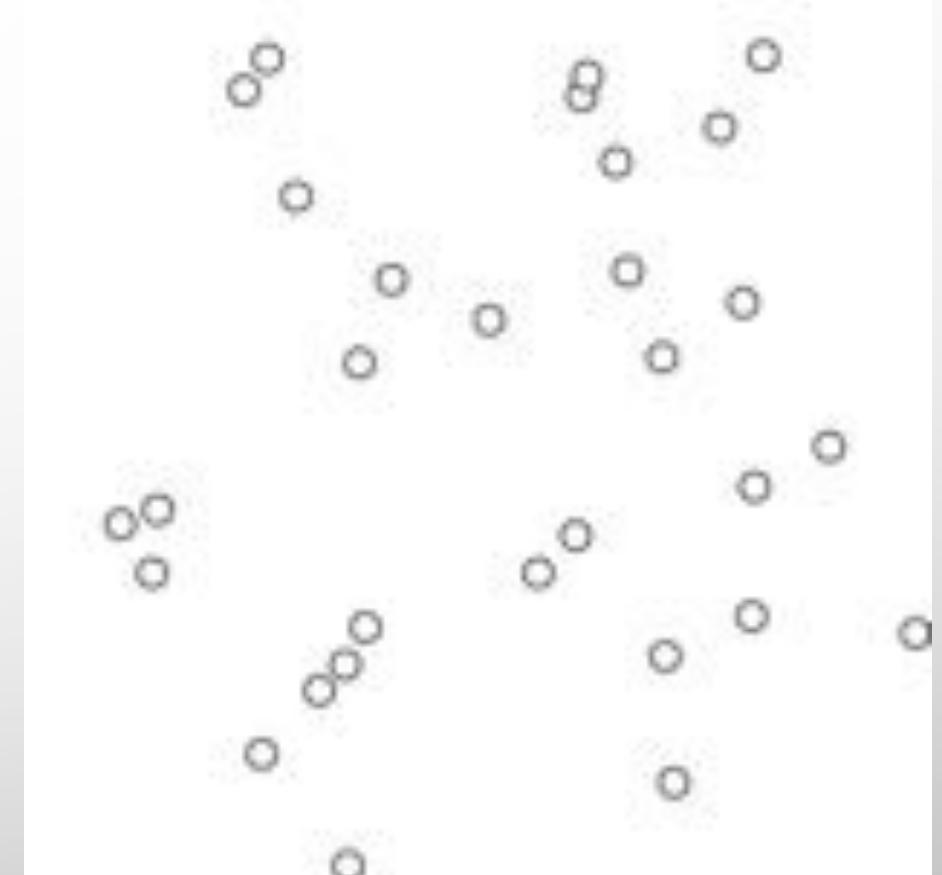
分析：当点 D 在线段 AD 上时(如图 1)，求证： $\angle AED = \angle BCD = 90^\circ$ ；

当点 D 在线段 AD 延长线上时(如图 2)，求证： $AB - AD = 2BD$ ；

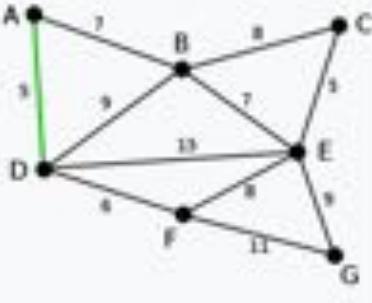
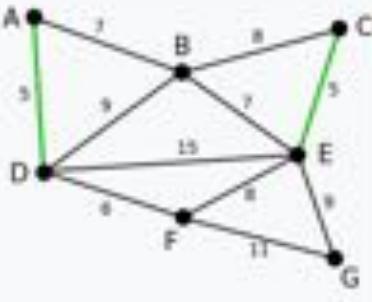
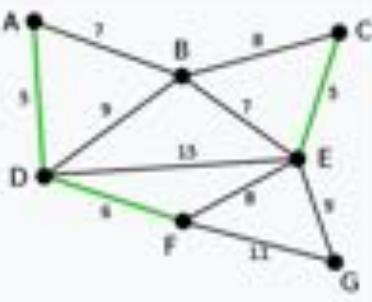
当点 D 在线段 AD 延长线上时(如图 3)，若  $\angle BAD = 60^\circ$ ，连接 BD，作  $\angle ABC$  的平分线  $BE$  交 AC 于点 O，连接 BO 并延长交 AB 于点 C.  $BC = 2$ ,  $CD = 3$ ，求线段 BD

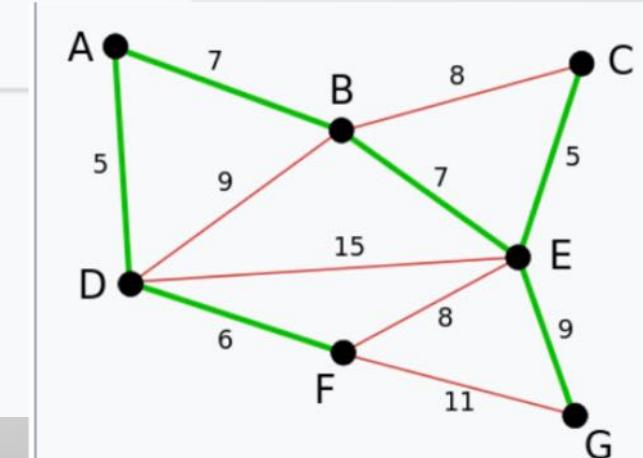
# Text Line Extraction based on Minimum Spanning Tree Clustering

```
KRUSKAL(G):
1 A = ∅
2 foreach v ∈ G.V:
3     MAKE-SET(v)
4 foreach (u, v) in G.E ordered by weight(u, v), increasing:
5     if FIND-SET(u) ≠ FIND-SET(v):
6         A = A ∪ {(u, v)}
7         UNION(u, v)
8 return A
```

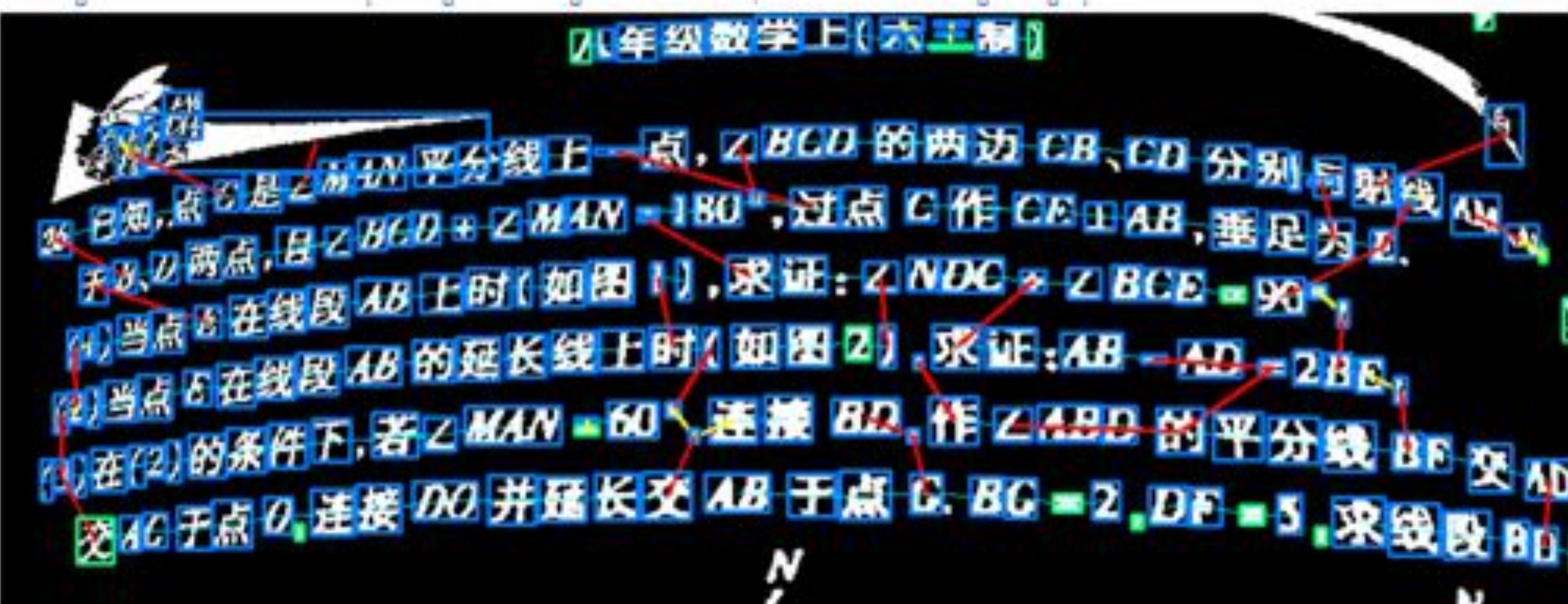


## Text Line Extraction based on Minimum Spanning Tree Clustering

Image	Description
	AD and CE are the shortest edges, with length 5, and AD has been arbitrarily chosen, so it is highlighted.
	CE is now the shortest edge that does not form a cycle, with length 5, so it is highlighted as the second edge.
	The next edge, DF with length 6, is highlighted using much the same method.



# Text Line Extraction based on Minimum Spanning Tree Clustering



## (Sub)Text Line Extraction

### 八年级数学上(苏科版)

如图，点 C 是  $\angle MAN$  平分线上一点， $\square BCD$  的两边  $CB$ 、 $CD$  分别与射线  $AM$  于 B、D 两点，且  $\angle BCD + \angle MAN = 180^\circ$ . 过点 C 作  $CE \perp AB$ ，垂足为 E.

(1) 当点 D 在线段 AB 上时(如图 1)，求证： $\square NDX + \square BCE = 90^\circ$ ；

(2) 当点 D 在线段 AB 的延长线上时(如图 2)，求证： $AB - AD = 2BE$ ；

(3) 在(2)的条件下，若  $\angle MAN = 60^\circ$ ，连接 BD，作  $\square ABD$  的平分线 BE 交  $AD$  于点 O，连接 DO 并延长交 AB 于点 C.  $BC = 2$ ,  $DE = 3$ . 求线段 BD.

## Text Line Extraction

系貴臣。遠小人。比先漢所以興隆也。亲十八。远贤。近。此后汉所以倾颓也。先帝去时。每于臣论此事。臣未尝不叹息痛恨于桓、灵也。侍中、尚书、长史、参军。此尤大臣无尺寸过。原陛下亲之信之。则汉室之隆。可计日而待也。

In consideration of your use of the service, you represent that you are of legal age to form a binding contract and are not a person barred from receiving services under the laws of the United States or other applicable jurisdiction. You also agree to: (a) provide true, accurate, current and complete information about yourself as prompted by the service's registration form and (b) maintain and promptly update the registration data to keep it true, accurate, current and complete. For this reason, parents of children under the age of 13 who wish to allow their children to access the service must create a family account.

## Character Segmentation

基于行片段拟合边界，再次对连通域进行划分



## Character Segmentation

计算分割点置信度，在粗分割后，递归地以置信度最大点进行分割



投影法

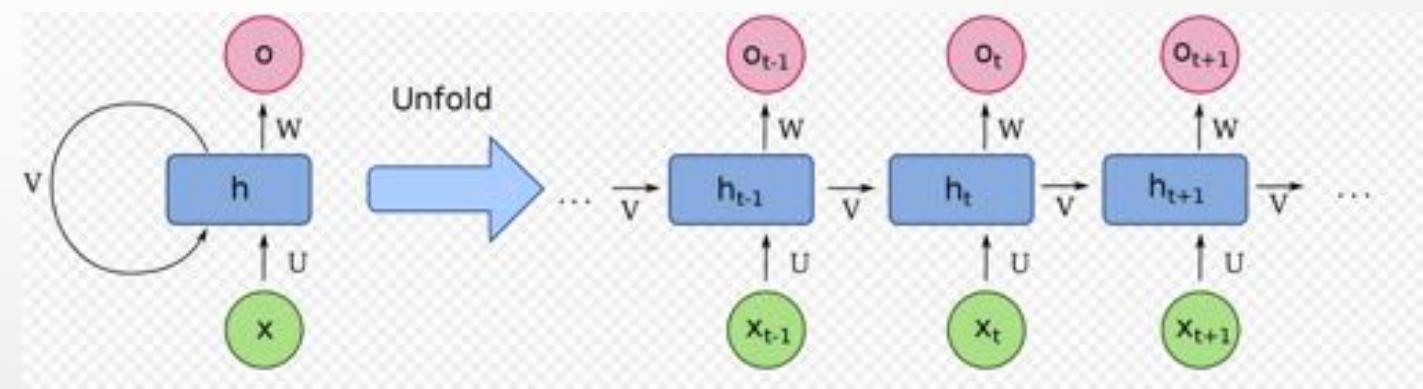
# Character Segmentation

计算分割点置信度，在粗分割后，递归地以置信度最大点进行分割



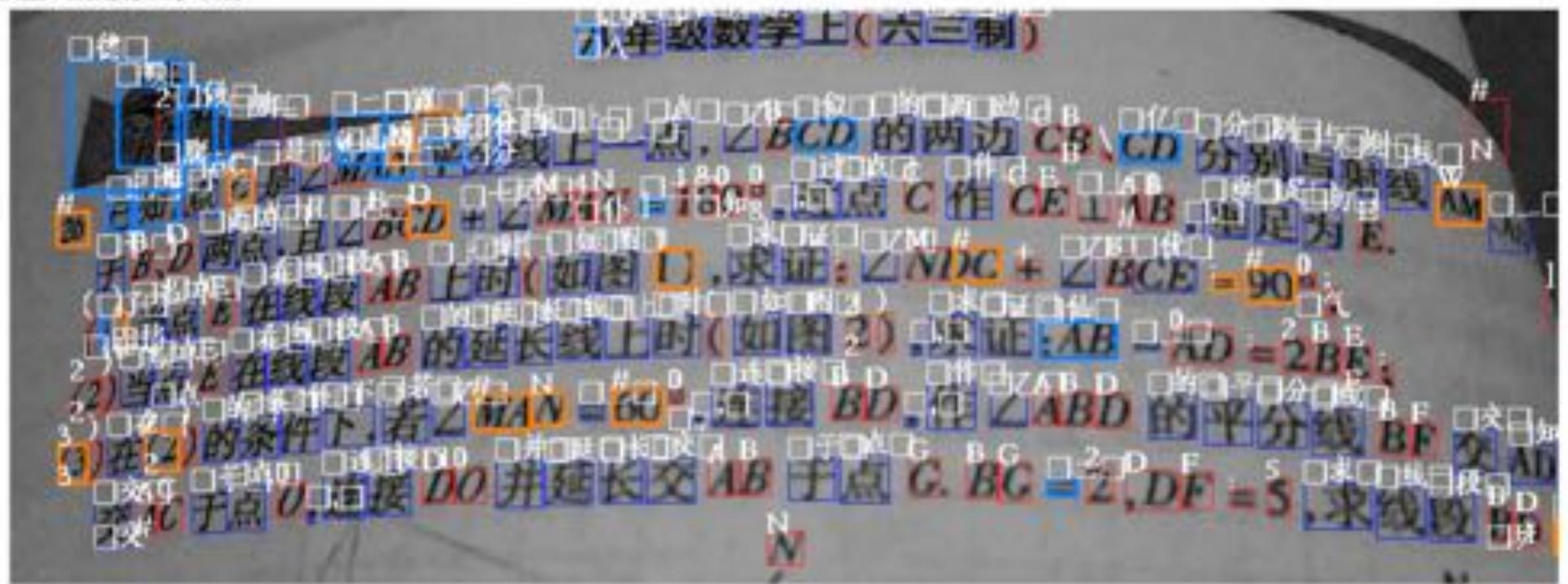
# Character Segmentation

计算分割点置信度，在粗分割后，递归地以置信度最大点进行分割



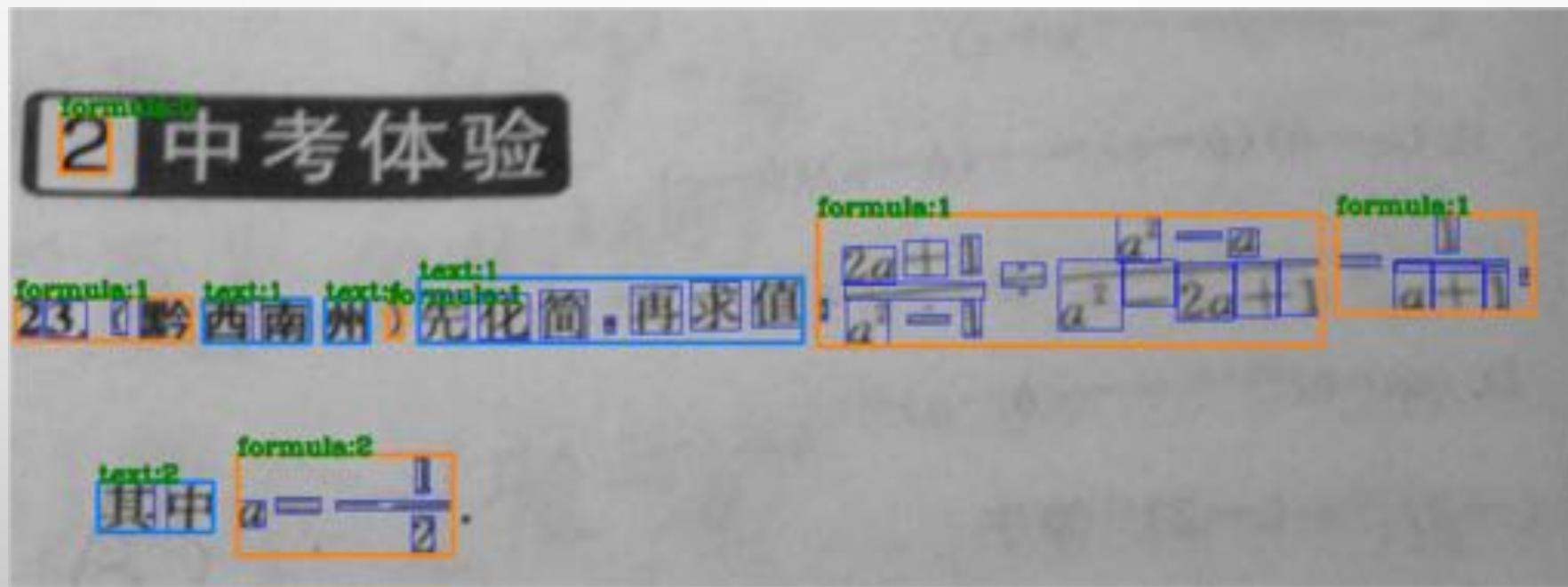
# Result (Round1)

调用模型识别进行仅字识别



# 遗留问题

- 中文 / 公式混排



## 遗留问题

- 字符分割

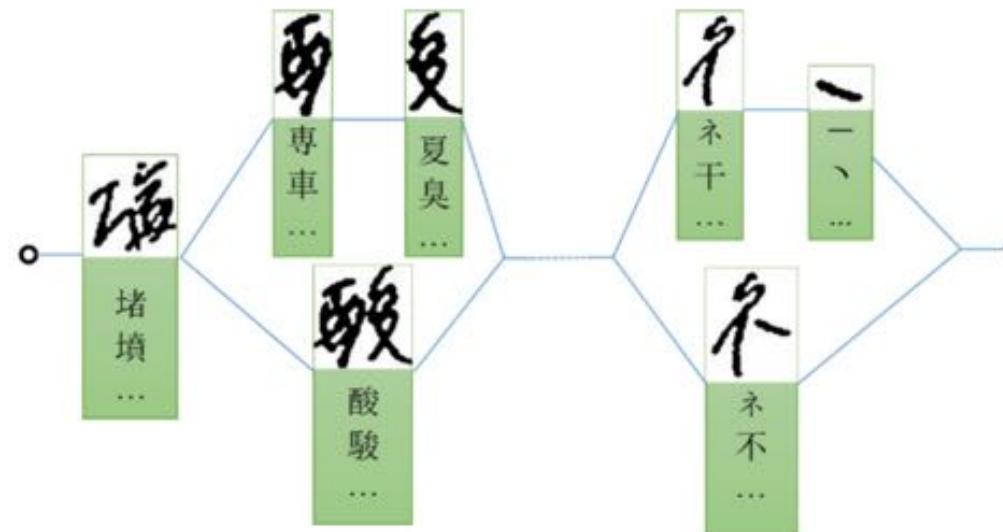
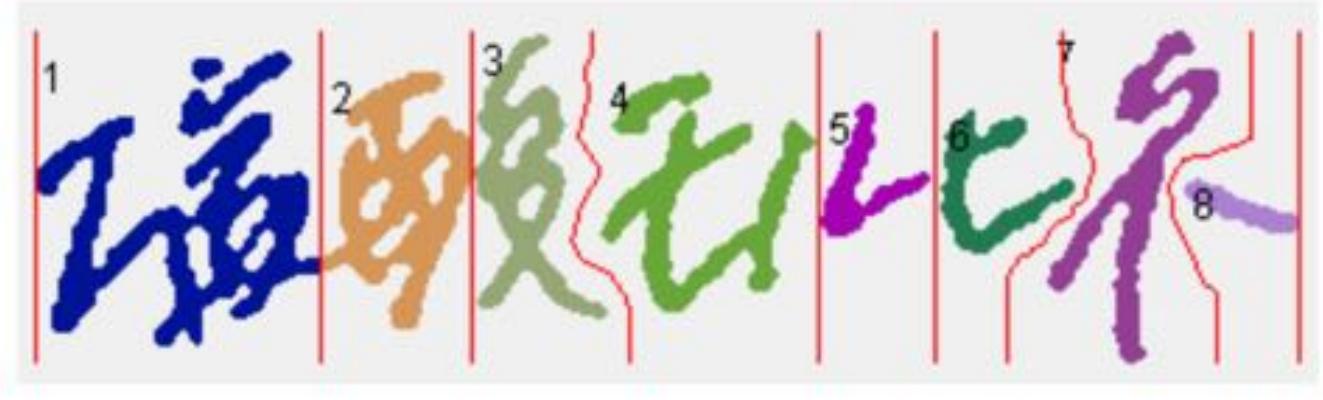


Fig. 11: Candidate lattice diagram.

# 遗留问题

- 检查中文 / 公式是否符合语法规则 ?

## 第二课 单字符分割与识别

知识点1：单字符识别（模版匹配、特征提取 + 传统分类器、CNN模型）

知识点2：字符分割1：定位候选分割位置（基于投影规则的候选位置提取、基于模型的候选位置提取）

知识点3：字体分割2：分割路径选择（Viterbi算法 / Beam Search、PCFG / 2D-PCFG）

**实战项目：** 基于CNN的单字符识别与过分割

# 遗留问题

- 二值化怎么做
  - OTSU
  - MSER (MAXIMALLY STABLE EXTERNAL REGIONS)
  - ...

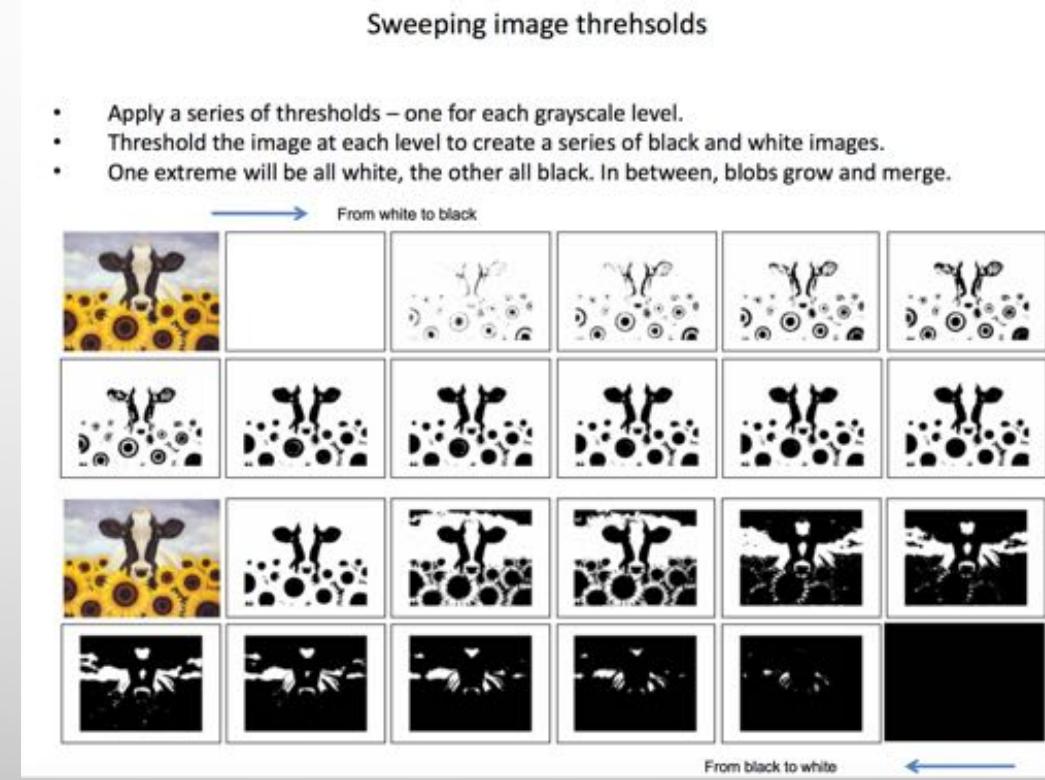
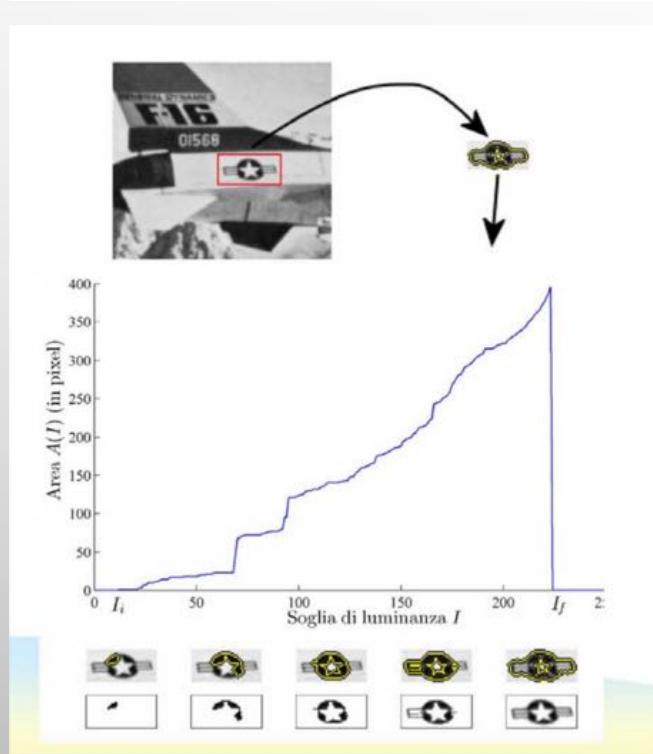
# OTSU

在计算机视觉和图像处理中，大津二值化法用来自动对基于聚类的图像进行二值化，<sup>[1]</sup>或者说，将一个灰度图像退化为二值图像。该算法以大津展之命名。算法假定该图像根据双模直方图（前景像素和背景像素）把包含两类像素，于是它要计算能将两类分开的最佳阈值，使得它们的类内方差最小；由于两两平方距离恒定，所以即它们的类间方差最大。<sup>[2]</sup>因此，大津二值化法粗略的来说就是一维 Fisher判别分析的离散化模拟。



# MSER (MAXIMALLY STABLE EXTERNAL REGIONS)

MSER算子的主要思想为：设定一系列的阈值参数 $\Delta$ 的范围[0 255]进行分割操作，高于阈值参数 $\Delta$ 的图像像素设置为白色，相反低于 $\Delta$ 的设置为黑色。在阈值参数 $\Delta$ 不断改变过程中，会逐渐出现闭合区域。最后，在设置阈值参数 $\Delta$ 的浮动范围内的极值点区域面积变化最小的判定为最大稳定极值区域。由于MSER只能单方面计算最大稳定极值区域，所以Matas提出通过图像反相来求取最小极值稳定区域。



# MSER (MAXIMALLY STABLE EXTERNAL REGIONS)

MSER算子公式如下：

$$Q_{ratio}(i) = \frac{|Q_{i+\Delta} - Q_{i-\Delta}|}{|Q_i|}$$

上式参数 $Q_i$ 代表阈值为*i*时连通区域大小， $\Delta$ 为阈值变化系数， $Q_{ratio}$ 是区域 $Q_i$ 在 $2\Delta$ 范围内的变换率。当 $Q_{ratio}$ 的比率值很小时，可以得出 $Q_i$ 在变化范围 $2\Delta$ 为最大稳定极值区域。MSER采取分水岭区域增长算法进行斑块极值轮廓的提取，同时MSER在区域提取阈值参数增长的过程中类似多尺度检测，不同阈值检测出不同的大小结构。由于MSER算法效率较低，2008年David Nister等提出改进MSER算法，借助于改进分水岭算法的理论，加快MSER算法检测效率。

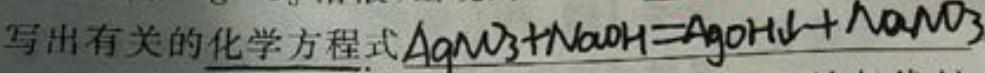
Br

4. 某一氯代烷 1.85 g, 与足量 NaOH 水溶液共热后, 不经硝酸酸化, 再加入足量的 AgNO<sub>3</sub> 溶液, 生成白色沉淀。

$$2.87 \text{ g} \cdot n(\text{AgCl}) = \frac{2.87}{143.5} = 0.02 \text{ mol}$$

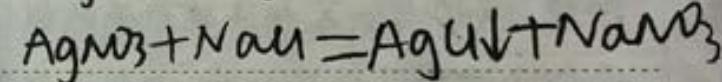
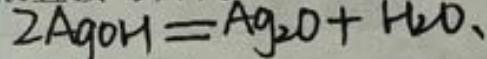
(1) 通过计算, 写出这种一氯代烷的各种同分异构体的结构简式:  $\text{CH}_2\text{CH}_2\text{CH}_2\text{CH}_3$ ,  $\text{CH}_3\text{CH}(\text{CH}_3)\text{CH}_2\text{CH}_3$ ,  $\text{CH}_3\text{CH}_2\text{CH}(\text{CH}_3)_2$

(2) 若此一氯代烷与足量 NaOH 水溶液共热后, 不经硝酸酸化就加 AgNO<sub>3</sub> 溶液, 出现的现象为 将产生褐色沉淀



(3) 能否用硝酸银溶液直接与卤代烃反应来鉴别卤代烷?

为什么?



不能, 卤代烃中的卤素是以离子的形式析出的

2. 下列各句中, 没有语病的一项是

- A. 由于莫言获得诺贝尔奖, 必然会使得他的作品在短期内销量大增, 这毫无疑问, 但文学想要再造昔日辉煌, 已几乎没有可能。
- B. 国庆期间高速公路免费通行让许多高速公路变成了巨型停车场, 显然, 政策制定者事先没能料到这样的结果, 他们的初衷并非是这样的。
- C. 最新人口普查结果显示, 河南省常住人口总数名列全国第三, 达到了 9 402.4 万人, 其中南阳、周口、郑州三城市人口最多, 均超过了 800 万。
- D. 伦敦奥运闭幕式上的“里约 8 分钟”十分精彩, 那一刻, 和着欢快

# 遗留问题

- 二值化得不到好结果怎么办

## 第二阶段：掌握文本定位与文本检测

### 第四课 文本行定位

知识点1：传统方法（基于投影的文本行定位、基于最小生成树的文本行定位）

知识点2：深度学习（Full-Page Text Recognition: Learning Where to Start and When to Stop, Learning Text-Line Localization with Shared and Local Regression Neural Networks, TextSnake）

实战项目：实战Text Line Extraction Based on MST

### 第五课 自然场景中的文本检测

知识点1：Reading Text in the Wild with Convolutional Neural Networks

知识点2：CTPN, RRPN, FTSN, DMPNet, EAST

知识点3：SegLink, PixelLink

知识点4：Textboxes, WordSup, FOTS

实战项目：实战场景文本检测之CTPN算法

4. 某一氯代烷 1.85 g, 与足量 NaOH 小心  
硝酸酸化, 再加入足量的 AgNO<sub>3</sub> 溶液, 生成白色沉淀

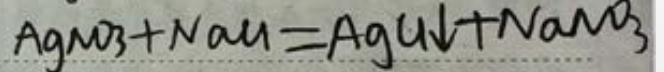
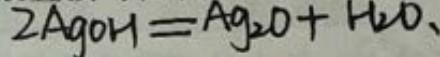
$$2.87 \text{ g} \cdot n(\text{AgCl}) = \frac{2.87}{143.5} = 0.02 \text{ mol}$$

(1) 通过计算, 写出这种一氯代烷的各种同分异构体的结  
构简式 ~~CH<sub>2</sub>CH<sub>2</sub>CH<sub>2</sub>CH<sub>3</sub>, CH<sub>3</sub>CH(CH<sub>3</sub>)<sub>2</sub>, CH<sub>3</sub>CH<sub>2</sub>CH<sub>2</sub>Cl, CH<sub>3</sub>CH(Cl)CH<sub>3</sub>~~

(2) 若此一氯代烷与足量 NaOH 水溶液共热后, 不经硝酸  
酸化就加 AgNO<sub>3</sub> 溶液, 出现的现象为 将会产生褐色沉淀

写出有关的化学方程式 AgNO<sub>3</sub> + NaOH = AgOH↓ + NaNO<sub>3</sub>

(3) 能否用硝酸银溶液直接与卤代烃反应来鉴别卤代烷?  
为什么?



不能, 卤代烃中的卤素均是以离子的形式存在

# 第一课 OCR技术概览

- 什么是OCR
  - OCR技术的使用场景
  - 目前OCR的技术难点
- OCR基本流程
  - 实例一：通用文本识别
  - 实例二：车牌检测与识别
- 第一个OCR小程序
  - TESSERACT

## 实例二

- 车牌检测与识别



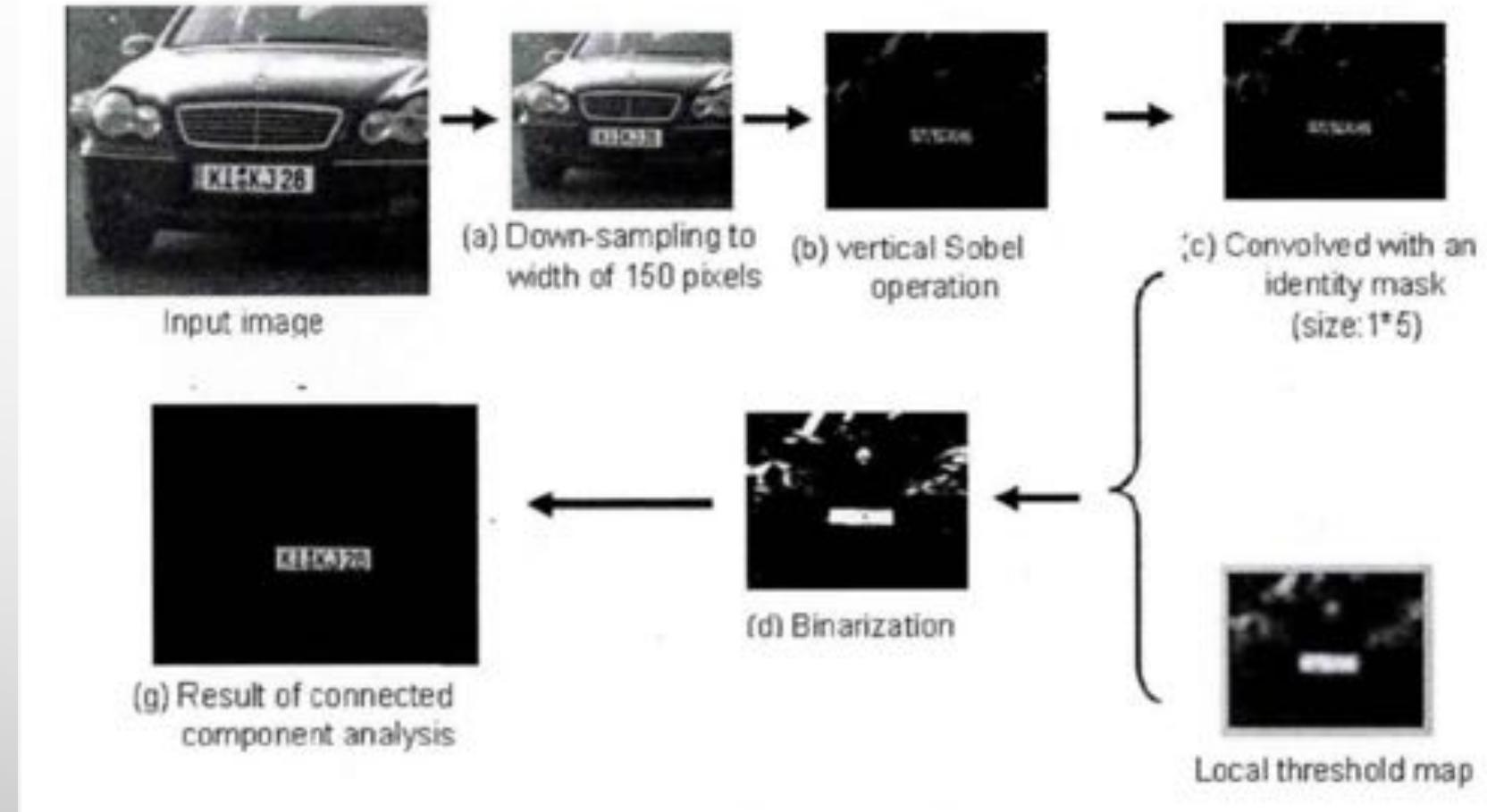
## Vehicle Plate Detection

### 竖直纹理卷积法

- 1) 车牌区域内的边缘灰度直方图统计特征，车牌区域内的边缘灰度直方图具有两个明显且分离的分布中心；
- 2) 车牌的几何特征，即车牌的高、宽和高宽比在一定范围内；
- 3) 车牌区域水平或垂直投影特征，车牌区域水平或垂直投影呈现连续的峰、谷、峰的分布；
- 4) 车牌与周边区域纹理密度分布对比<sup>[33]</sup>；

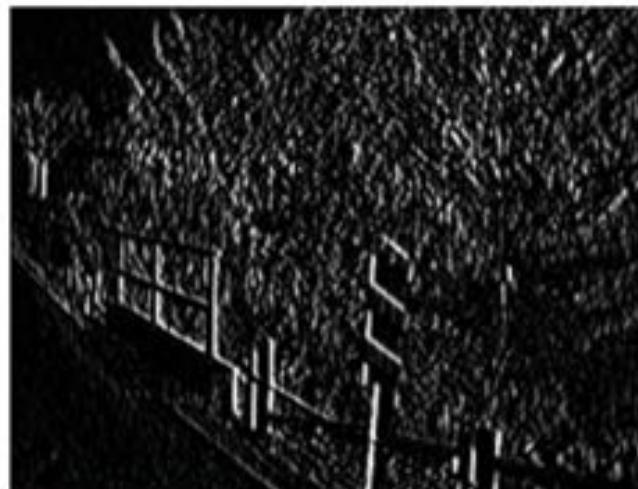
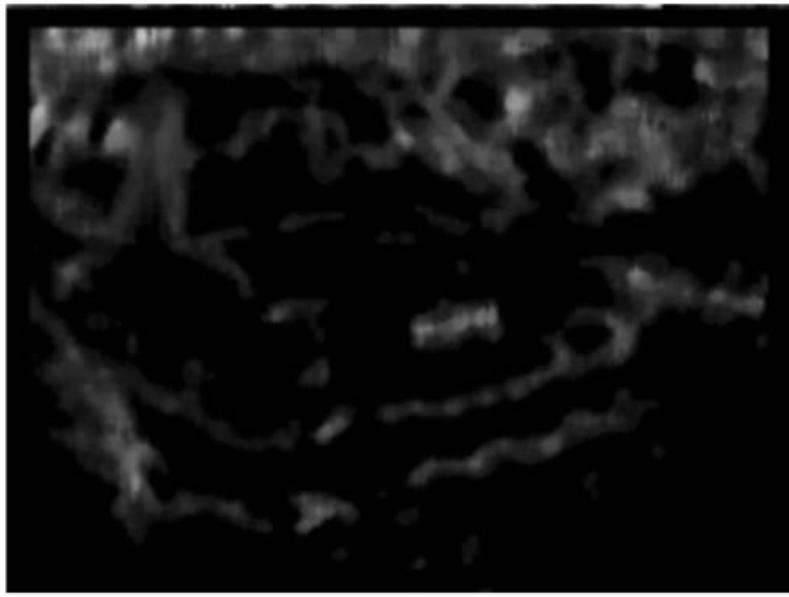
# Vehicle Plate Detection

竖直纹理卷积法



## Vehicle Plate Detection

竖直纹理卷积法



1	0	-1
2	0	-2
1	0	-1

Horizontal diff.

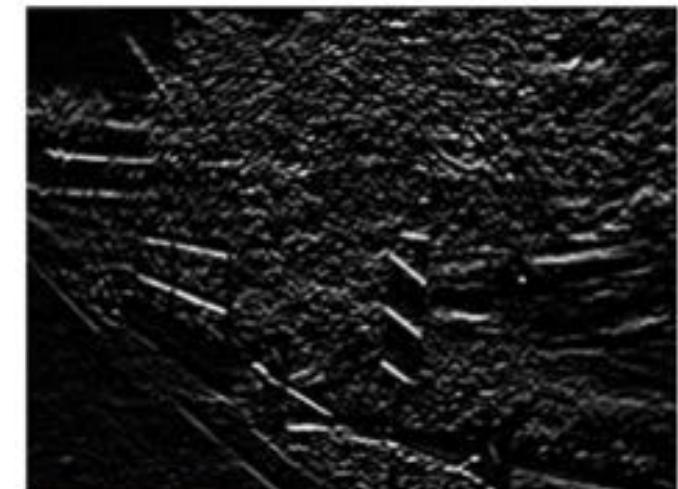
Computer Vision :

1	2	1
0	0	0
-1	-2	-1

Vertical diff.

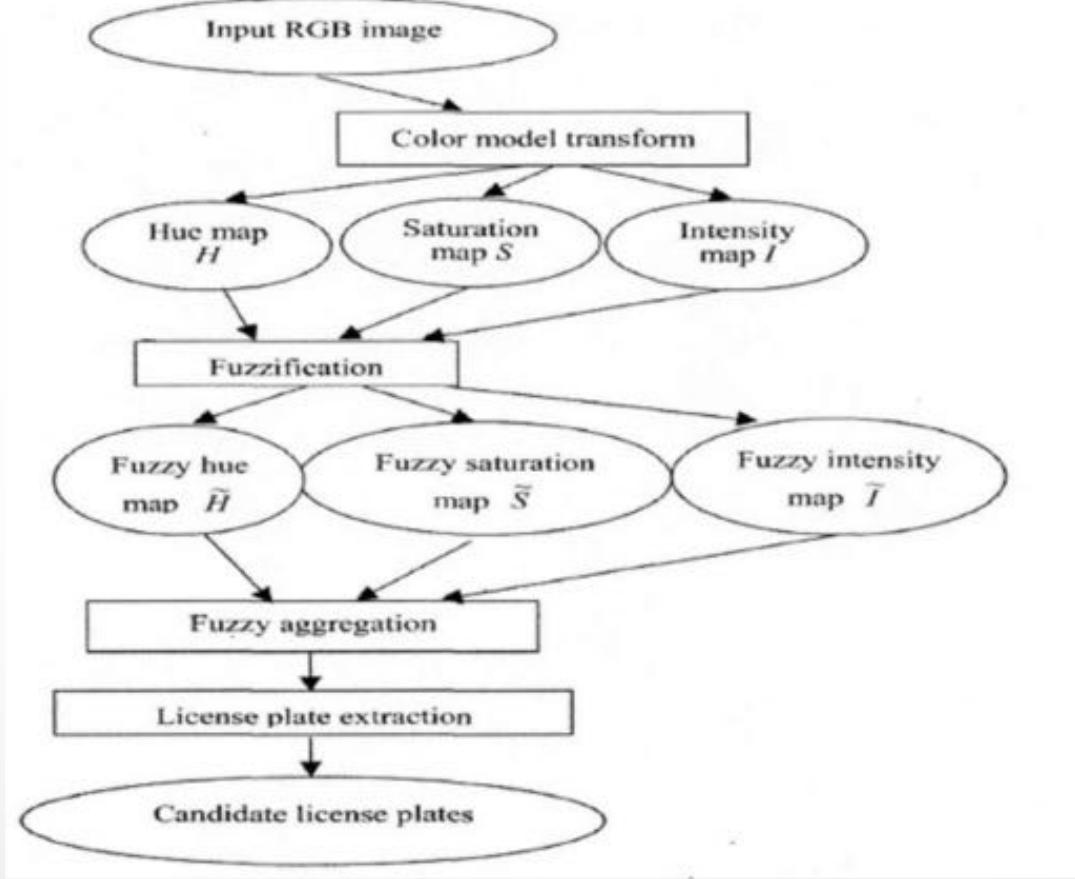
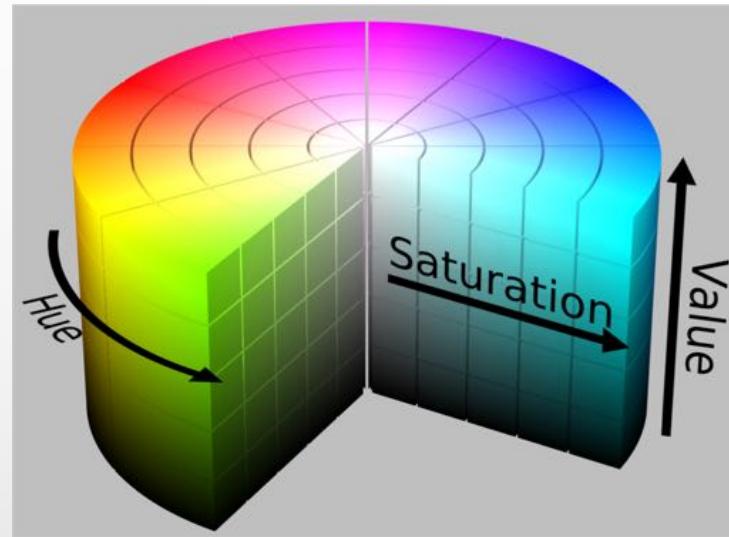
## Sobel Edge Detection: Gradient Approximation

*Note anisotropy of edge finding*



# Vehicle Plate Detection

## 色彩空间模糊集



目前我国的汽车牌照有黄底黑字、蓝底白字、黑底白字、白底黑字红字、黑底红字等 5 种颜色的车牌，常见的是民用的黄底黑字和蓝底白字两种车牌可以根据车牌的色彩特征将车牌定位。

## Vehicle Plate Detection

区域分割法

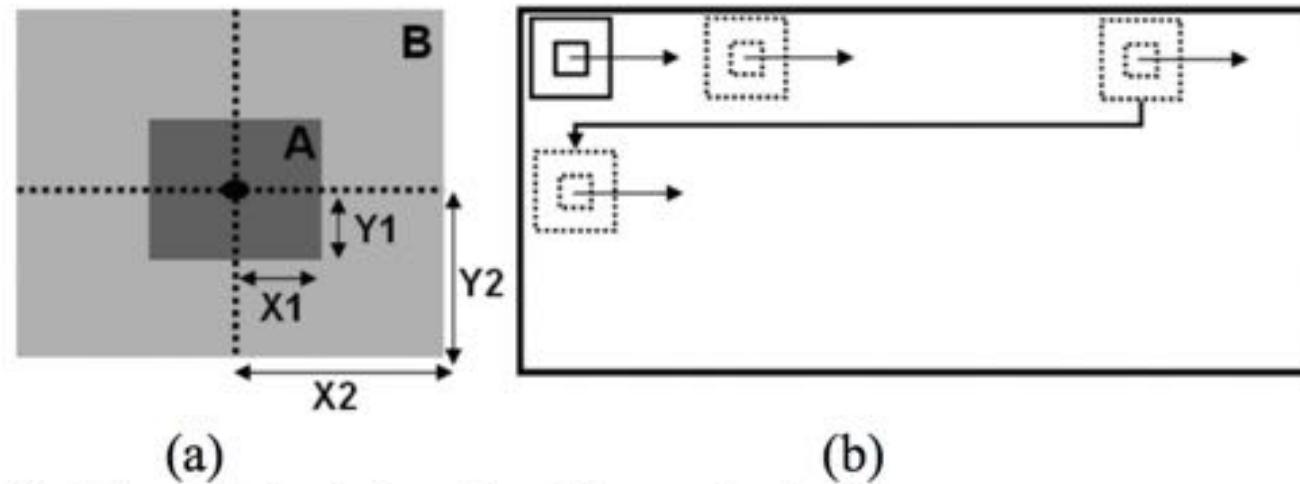


Fig.1 Concentric windows A and B scanning the image.



## Vehicle Plate Detection

### 区域分割法

算出 A 和 B 的均值和标准差等统计特征值，如果 A 和 B 框的统计特征值的比值超过定义的阈值，则当前点被认为是感兴趣点，被赋值为 1，否则为 0。A 和 B 对整幅图像逐点滑动，直到整个图像扫描完为止。接着使用连通域分析(CCA)把相邻的兴趣点连结起来，根据 CCA 可以找出面积、方向、宽高比等特征，若连结好的区域满足欧拉数大于 3、宽高比大于 2 小于 6 等特征，即为车牌的候选位置储存在矩阵中。

# Vehicle Plate Detection

Deep Learning

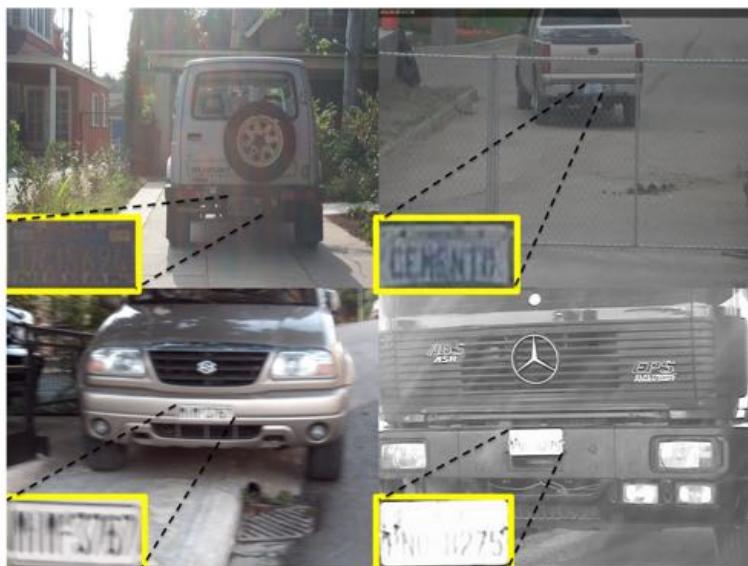
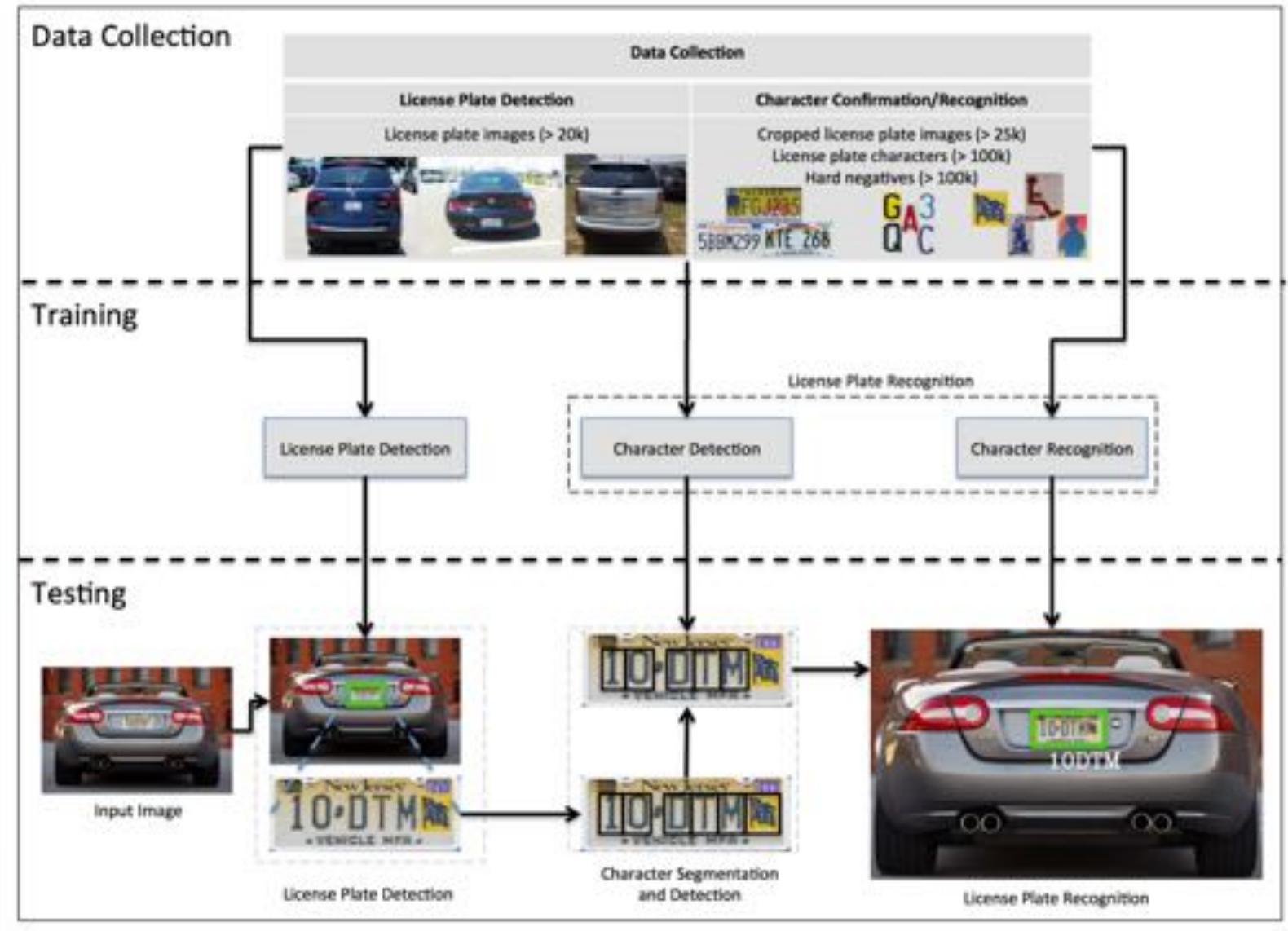
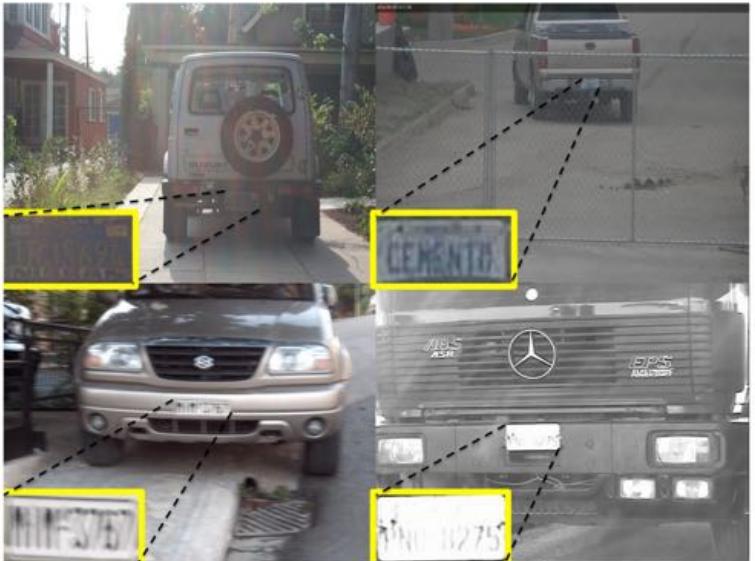


Fig. 3: Figure shows some of the hard examples (lighting, blurriness, size, angle, etc) removed before running benchmarks. The cropped license plate used for recognition processing is shown in the bottom-left corner for each sample.

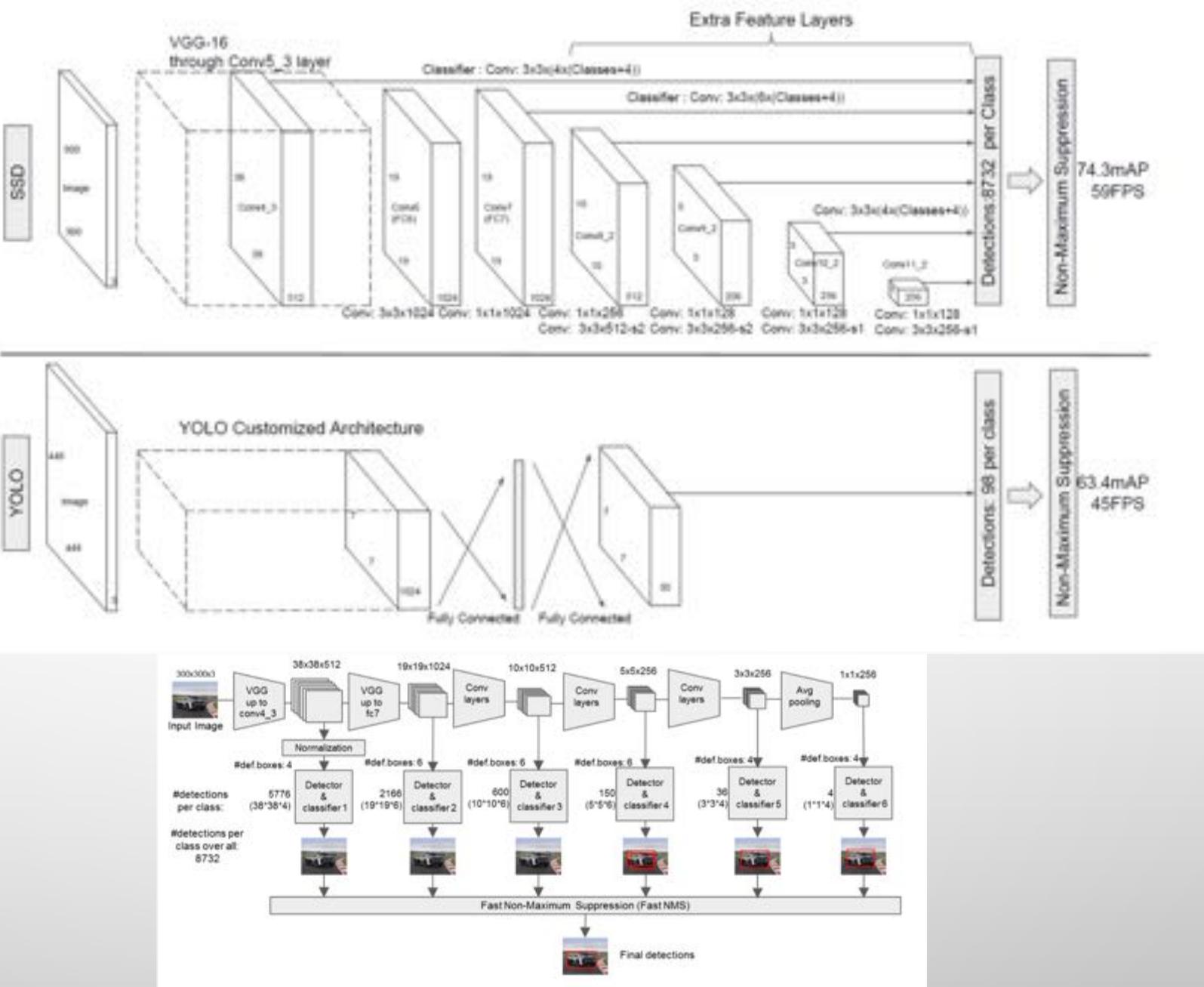


# Vehicle Plate Detection

## Deep Learning



**Fig. 3:** Figure shows some of the hard examples (lighting, blurriness, size, angle, etc) removed before running benchmarks. The cropped license plate used for recognition processing is shown in the bottom-left corner for each sample.



License Plate Detection and Recognition Using Deeply Learned Convolutional Neural Networks

# Vehicle Plate Recognition

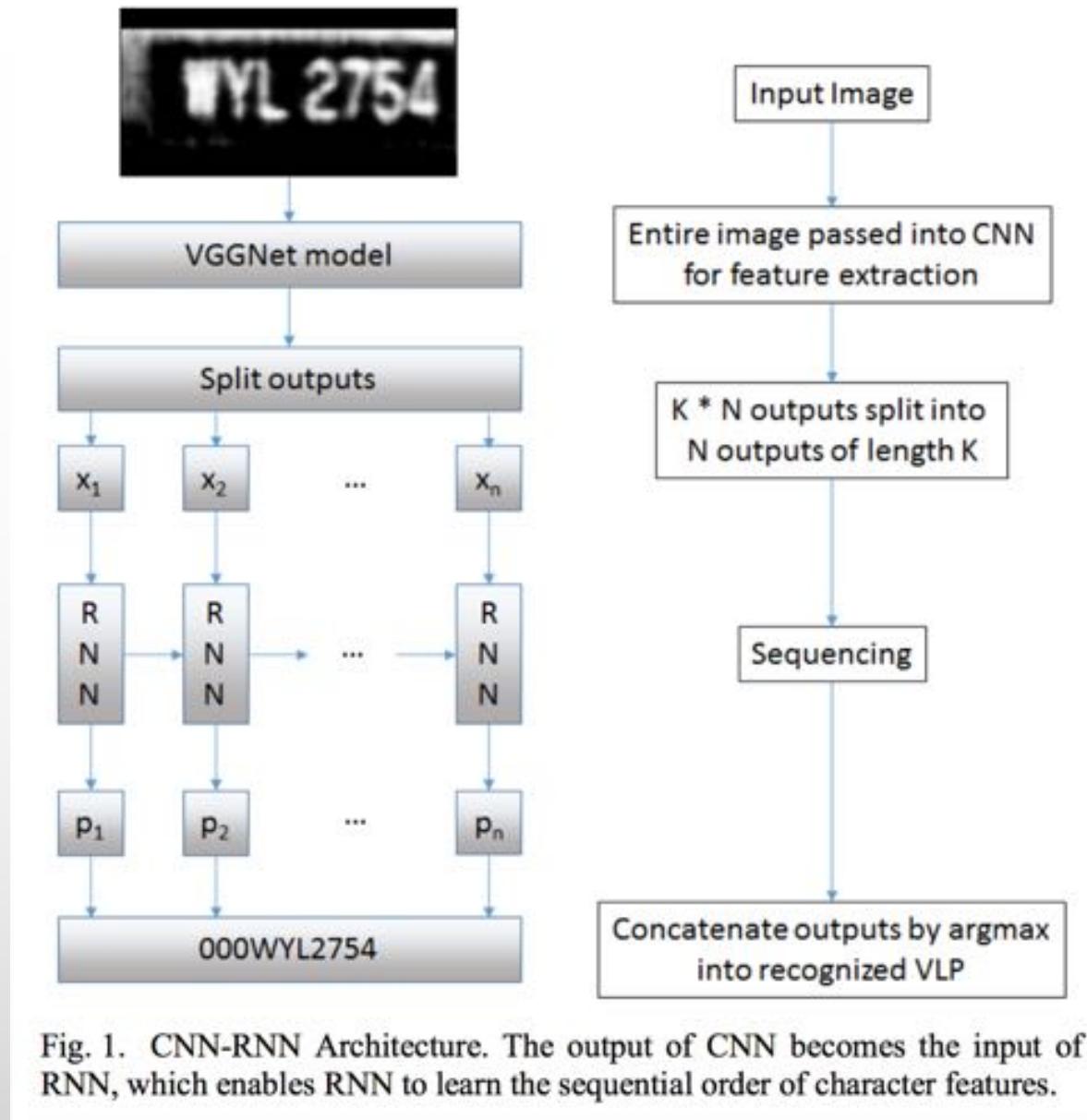


Fig. 1. CNN-RNN Architecture. The output of CNN becomes the input of RNN, which enables RNN to learn the sequential order of character features.

# 第一课 OCR技术概览

- 什么是OCR
  - OCR技术的使用场景
  - 目前OCR的技术难点
- OCR基本流程
  - 实例一：通用文本识别
  - 实例二：车牌检测与识别
- 第一个OCR小程序
  - TESSERACT

# TESSERACT OCR

## Brief history

Tesseract was originally developed at Hewlett-Packard Laboratories Bristol and at Hewlett-Packard Co, Greeley Colorado between 1985 and 1994, with some more changes made in 1996 to port to Windows, and some C++izing in 1998. In 2005 Tesseract was open sourced by HP. Since 2006 it is developed by Google.

The latest (LSTM based) stable version is [4.0.0](#), released on October 29, 2018. Latest source code for 4.0 is available from [master branch on GitHub](#). Open issues can be found in [issue tracker](#), and [Planning](#) wiki.

The latest 3.5 version is [3.05.02](#), released on June 19, 2018. Latest source code for 3.05 is available from [3.05 branch on GitHub](#). There is no development for this version, but it can be used for special cases (e.g. see [Regression of features from 3.0x](#)).

See [Release Notes](#) and [Change Log](#) for more details of the releases.

# TESSERACT OCR

- GITHUB
  - [HTTPS://GITHUB.COM/TESSERACT-OCR/TESSERACT](https://github.com/tesseract-ocr/tesseract)
- COMPILING
  - [HTTPS://GITHUB.COM/TESSERACT-OCR/TESSERACT/WIKI/COMPILING](https://github.com/tesseract-ocr/tesseract/wiki/compiling)
- TRAINING
  - [HTTPS://GITHUB.COM/TESSERACT-OCR/TESSERACT/WIKI/TRAININGTESSERACT](https://github.com/tesseract-ocr/tesseract/wiki/trainingtesseract)
  - [HTTPS://GITHUB.COM/TESSERACT-OCR/TESSERACT/WIKI/TRAININGTESSERACT-4.00](https://github.com/tesseract-ocr/tesseract/wiki/trainingtesseract-4.00)

# TESSERACT OCR

- COMMAND LINE USAGE

- <https://github.com/tesseract-ocr/tesseract/wiki/Command-Line-Usage#simplest-invocation-to-ocr-an-image>

```
tesseract --help-extra
```

Usage:

```
tesseract --help | --help-extra | --help-psm | --help-oem | --version  
tesseract --list-langs [--tessdata-dir PATH]  
tesseract --print-parameters [options...] [configfile...]  
tesseract imagename|imagelist|stdin outputbase|stdout [options...] [configfile...]
```

OCR options:

--tessdata-dir PATH	Specify the location of tessdata path.
--user-words PATH	Specify the location of user words file.
--user-patterns PATH	Specify the location of user patterns file.
-l LANG[+LANG]	Specify language(s) used for OCR.
-c VAR=VALUE	Set value for config variables. Multiple -c arguments are allowed.
--psm NUM	Specify page segmentation mode.
--oem NUM	Specify OCR Engine mode.

NOTE: These options must occur before any configfile.

# TESSERACT OCR

Page segmentation modes:

- 0 Orientation and script detection (OSD) only.
- 1 Automatic page segmentation with OSD.
- 2 Automatic page segmentation, but no OSD, or OCR.
- 3 Fully automatic page segmentation, but no OSD. (Default)
- 4 Assume a single column of text of variable sizes.
- 5 Assume a single uniform block of vertically aligned text.
- 6 Assume a single uniform block of text.
- 7 Treat the image as a single text line.
- 8 Treat the image as a single word.
- 9 Treat the image as a single word in a circle.
- 10 Treat the image as a single character.
- 11 Sparse text. Find as much text as possible in no particular order.
- 12 Sparse text with OSD.
- 13 Raw line. Treat the image as a single text line, bypassing hacks that are Tesseract-specific.

## tesseract --help-extra

### Usage:

```
tesseract --help | --help-extra | --help-psm | --help-oem | --version  
tesseract --list-langs [--tessdata-dir PATH]  
tesseract --print-parameters [options...] [configfile...]  
tesseract imagename|imagelist|stdin outputbase|stdout [options...] [configfile...]
```

### OCR options:

--tessdata-dir PATH	Specify the location of tessdata path.
--user-words PATH	Specify the location of user words file.
--user-patterns PATH	Specify the location of user patterns file.
-l LANG[+LANG]	Specify language(s) used for OCR.
-c VAR=VALUE	Set value for config variables. Multiple -c arguments are allowed.
--psm NUM	Specify page segmentation mode.
--oem NUM	Specify OCR Engine mode.

NOTE: These options must occur before any configfile.

### OCR Engine modes:

- 0 Legacy engine only.
- 1 Neural nets LSTM engine only.
- 2 Legacy + LSTM engines.
- 3 Default, based on what is available.

# TESSERACT OCR

- TESSERACT TEST.PNG TEST.TXT -L ENG --PSM 1 --OEM 1

## Installing Tesseract

You can either [Install Tesseract via pre-built binary package](#) or [build it from source](#).

Supported Compilers are:

- GCC 4.8 and above
- Clang 3.4 and above
- MSVC 2015, 2017

Other compilers might work, but are not officially supported.

```
→ ocr cat test.txt.txt  
Installing Tesseract
```

You can either [Install Tesseract via pre-built binary package](#) or [build it from source](#).

Supported Compilers are:

- ↳ GCC 4.8 and above
- ↳ Clang 3.4 and above
- ↳ MSVC 2015, 2017

Other compilers might work, but are not officially supported.

# TESSERACT OCR

- API EXAMPLE

## Basic example

Code:

```
#include <tesseract/baseapi.h>
#include <leptonica/allheaders.h>

int main()
{
    char *outText;

    tesseract::TessBaseAPI *api = new tesseract::TessBaseAPI();
    // Initialize tesseract-ocr with English, without specifying tessdata path
    if (api->Init(NULL, "eng")) {
        fprintf(stderr, "Could not initialize tesseract.\n");
        exit(1);
    }

    // Open input image with leptonica library
    Pix *image = pixRead("/usr/src/tesseract/testing/phototest.tif");
    api->SetImage(image);
    // Get OCR result
    outText = api->GetUTF8Text();
    printf("OCR output:\n%s", outText);

    // Destroy used object and release memory
    api->End();
    delete [] outText;
    pixDestroy(&image);

    return 0;
}
```

```
* Press F for Help  
  (up = left)
```

- dev/
- devlib/
- drivers/
- examples/
- include/
- lib/
- libexec/
- man/
- modules/
- net/
- scripts/
- security/
- tools/
- uapi/
- userspace/
- vhost/
- vhost-user/
- vhost-virtio/
- vhost-win/

```

124     const char* lang = multilang;
125     const char* image = nullimage;
126     const char* outputbase = nullbase;
127     const char* datapath = nullpath;
128     bool first_lang = false;
129     bool print_parameters = false;
130     t1_tmr32 dns = 0;
131     int arg_i = 1;
132     tessarch::PageBreakMode pagebreakmode = tessarch::PSM_AUTO;
133     #if defined(TESSARCH_LEGACY_FUNCTIONS)
134     auto enginemode = tessarch::GER_LISTONLY;
135     #endif
136     tessarch::ScanningMode enginemode = tessarch::SMR_DEFAULT;
137     Resultf
138     /* main() calls function-like parameters which call entry()
139      * Tella metadati in maniera simile a <math>\text{char}^*\text{argv}</math> e <math>\text{char}^*\text{envp}</math> per
140      * decodare le loro variabili. Distruttore di <math>\text{main}</math> non è chiamato. */
141     static GenericVector<string> vars_vec;
142     static GenericVector<string> vars_values;
143
144     #if !defined(TESSARCH_LEGACY_FUNCTIONS)
145     // Associate debugging and informational messages from Logarithma
146     setMsgSeverity(L_SEVERITY_ERROR);
147     #endif
148
149     #if defined(TESSARCH_LEGACY_FUNCTIONS)
150     // Show classify errors and warnings on console (not in log)
151     TIFFSetErrorHandler(Win32ErrorHandler);
152     TIFFSetWarningHandler(Win32WarningHandler);
153     #endif
154     #if HAVE_TIFF32K & !defined(XTIOS2)
155
156     ParseArgs(argc, argv, &lang, &image, &outputbase, &datapath, &dns,
157               &first_lang, &print_parameters, &vars_vec, &vars_values, &arg_i,
158               &pagebreakmode, &enginemode);
159
160     if (lang == multilang) {
161         SetDefaultLanguageIfNoneAreGiven();

```

Prints off file, P for help  
functions  
-FLxPageSegMode  
-PonSettings  
-PrvLoadRenderers  
-PrintBanner  
-PrintHelpOrErrors  
-PrintHelpOnPSK  
-PrintHelpOnPSH  
-PrintHelpMessage  
-PrintLangList  
-PrintVersionInfo  
-SetIconSizeFromQIcon  
-Win32ErrorHandler  
-Win32WarningHandler  
-Checking Pictures

[www.uptodate.com](http://www.uptodate.com)

www.english-test.net

中華書局影印

```
1080 }
1081
1082 // Master ProcessPages calls ProcessPagesInternal and then does any post-
1083 // processing required due to being in a training mode.
1084 bool TessBaseAPI::ProcessPages(const char* filename, const char* retry_config,
1085                                 int timeout_millisec,
1086                                 TessResultRenderer* renderer) {
1087     bool result =
1088         ProcessPagesInternal(filename, retry_config, timeout_millisec, renderer);
1089 #ifndef DISABLED_LEGACY_ENGINE
1090     if (result) {
1091         if (tesseract_->tessedit_train_from_boxes &&
1092             !tesseract_->WriteTRFile(*output_file_)) {
1093             tprintf("Write of TR file failed: %s\n", output_file_->string());
1094             return false;
1095         }
1096     }
1097 #endif // !defined DISABLED_LEGACY_ENGINE
1098     return result;
1099 }
1100
1101 // In the ideal scenario, Tesseract will start working on data as soon
1102 // as it can. For example, if you stream a filelist through stdin, we
1103 // should start the OCR process as soon as the first filename is
```

```
|1212  
|1213 bool TessBaseAPI::ProcessPage(Pix* pix, int page_index, const char* filename,  
|1214                               const char* retry_config, int timeout_millisec,  
|1215                               TessResultRenderer* renderer) {  
|1216     PERF_COUNT_START("ProcessPage")  
|1217     SetInputName(filename);  
|1218     SetImage(pix);  
|1219     bool failed = false;  
|1220  
|1221     if (tesseract_->tessedit_pageseg_mode == PSM_AUTO_ONLY) {  
|1222         // Disabled character recognition  
|1223         PageIterator* it = AnalyseLayout();  
|1224  
|1225         if (it == nullptr) {  
|1226             failed = true;  
|1227         } else {  
|1228             delete it;  
|1229         }  
|1230     } else if (tesseract_->tessedit_pageseg_mode == PSM OSD ONLY) {  
|1231         failed = FindLines() != 0;  
|1232     } else if (timeout_millisec > 0) {  
|1233         // Running with a timeout.  
|1234         ETEXT_DESC monitor;  
|1235         monitor.cancel = nullptr;  
|1236         monitor.cancel_this = nullptr;  
|1237         monitor.set_deadline_msecs(timeout_millisec);  
|1238  
|1239         // Now run the main recognition.  
|1240         failed = Recognize(&monitor) < 0;  
|1241     } else {  
|1242         // Normal layout and character recognition with no timeout.
```

```
839
840 /**
841 * Recognize the tesseract global image and return the result as Tesseract
842 * internal structures.
843 */
844 int TessBaseAPI::Recognize(ETEXT_DESC* monitor) {
845     if (tesseract_ == nullptr)
846         return -1;
847     if (FindLines() != 0)
848         return -1;
849     delete page_res_;
850     if (block_list_->empty()) {
851         page_res_ = new PAGE_RES(false, block_list_,
852                               &tesseract_->prev_word_best_choice_);
853         return 0; // Empty page.
854     }
855
856     tesseract_->SetBlackAndWhitelist();
857     recognition_done_ = true;
858 #ifndef DISABLED_LEGACY_ENGINE
859     if (tesseract_->tessedit_resegment_from_line_boxes) {
860         page_res_ = tesseract_->ApplyBoxes(*input_file_, true, block_list_);
861     } else if (tesseract_->tessedit_resegment_from_boxes) {
862         page_res_ = tesseract_->ApplyBoxes(*input_file_, false, block_list_);
863     } else
```

```
289 *
290 * Walk the page_res, recognizing all the words.
291 * If monitor is not null, it is used as a progress monitor/timeout/cancel.
292 * If dopasses is 0, all recognition passes are run,
293 * i just pass 1, 2 passes2 and higher.
294 * If target_word_box is not null, special things are done to words that
295 * overlap the target_word_box:
296 * if word_config is not null, the word config file is read for just the
297 * target word(s), otherwise, on pass 2 and beyond ONLY the target words
298 * are processed (Jetsoft modification.)
299 * Returns false if we cancelled prematurely.
300 *
301 * @param page_res page structure
302 * @param monitor progress monitor
303 * @param word_config word_config file
304 * @param target_word_box specifies just to extract a rectangle
305 * @param dopasses 0 - all, 1 just pass 1, 2 passes 2 and higher
306 */
307
308 bool Tesseract::recog_all_words(PAGE_RES* page_res,
309                                 ETEXT_DESC* monitor,
310                                 const TBOX* target_word_box,
311                                 const char* word_config,
312                                 int dopasses) {
313     PAGE_RES_IT page_res_it(page_res);
314
315     if (tessedit_minimal_rej_pass1) {
316         tessedit_test_adoption.set_value (TRUE);
317         tessedit_minimal_rejection.set_value (TRUE);
318     }
319
320     if (dopasses==0 || dopasses==1) {
321         page_res_it.restart_page();
322         // ***** Pass 1 *****
323
324 #ifndef DISABLED_LEGACY_ENGINE
325         // If the adoptive classifier is full switch to one we prepared earlier,
326         // ie on the previous page. If the current adoptive classifier is non-empty,
327         // prepare a backup starting at this page, in case it fills up. Do all this
```

# TESSERACT OCR

- TESSEROCR
  - [HTTPS://GITHUB.COM/SIRFZ/TESSEROCR](https://github.com/sirfz/TESSEROCR)
- PYTESSERACT
  - [HTTPS://GITHUB.COM/MADMAZE/PYTESSERACT](https://github.com/madmaze/PYTESSERACT)

# BAIDU OCR API

- OCR-PYTHON SDK 文档

- [HTTPS://CLOUD.BAIDU.COM/DOC/OCR/OCR-PYTHON-SDK.HTML](https://cloud.baidu.com/doc/OCR/OCR-PYTHON-SDK.html)
- [HTTPS://CLOUD.BAIDU.COM/DOC/OCR/OCR-PYTHON-SDK/24.5.C.E6.8E.A5.E5.8F.A3.E8.AF.B4.E6.98.8E.HTML](https://cloud.baidu.com/doc/OCR/OCR-PYTHON-SDK/24.5.C.E6.8E.A5.E5.8F.A3.E8.AF.B4.E6.98.8E.HTML)

## 新建AipOcr

AipOcr是OCR的Python SDK客户端。为使用OCR的开发人员提供了一系列的交互方法。

参考如下代码新建一个AipOcr：

```
from aip import AipOcr

""" 你的 APPID AK SK """
APP_ID = '你的 App ID'
API_KEY = '你的 Api Key'
SECRET_KEY = '你的 Secret Key'

client = AipOcr(APP_ID, API_KEY, SECRET_KEY)
```

在上面代码中，常量 APP\_ID 在百度云控制台中创建，常量 API\_KEY 与 SECRET\_KEY 是在创建完毕应用后，系统分配给用户的，均为字符串。用于标识用户，为访问做签名验证，可在AI服务控制台中的应用列表中查看。

# 作业

- 尝试使用TESSERACT
- 如果你来实现证照票据识别，会怎么设计吗，会遇到什么问题吗



# THANKS