基于小猿搜题的视觉应用落地

朱珊珊

猿辅导公司 基础研究团队研究员

zhuss@fenbi.com

AiCon
全球人工智能与机器学习技术大会

主办方: Geekbang 极客邦科技 InfoQ

场景

- 小猿搜题
  ➡ 小猿搜题

Step 1: Open ApeSearch Application

Step 2: Take a Picture of a Question

Step 3: Crop and Submit

Step 4: Get the Answer!

场景

- 小猿搜题

  ➡ 小猿搜题

  ➡ 英语作文拍照批改   难点1：手写英文作文的转化   拍照识别?

  ➡ 口算检查   难点2：批改打分的算法

  ➡ 古诗文背诵

  ➡ ……

场景
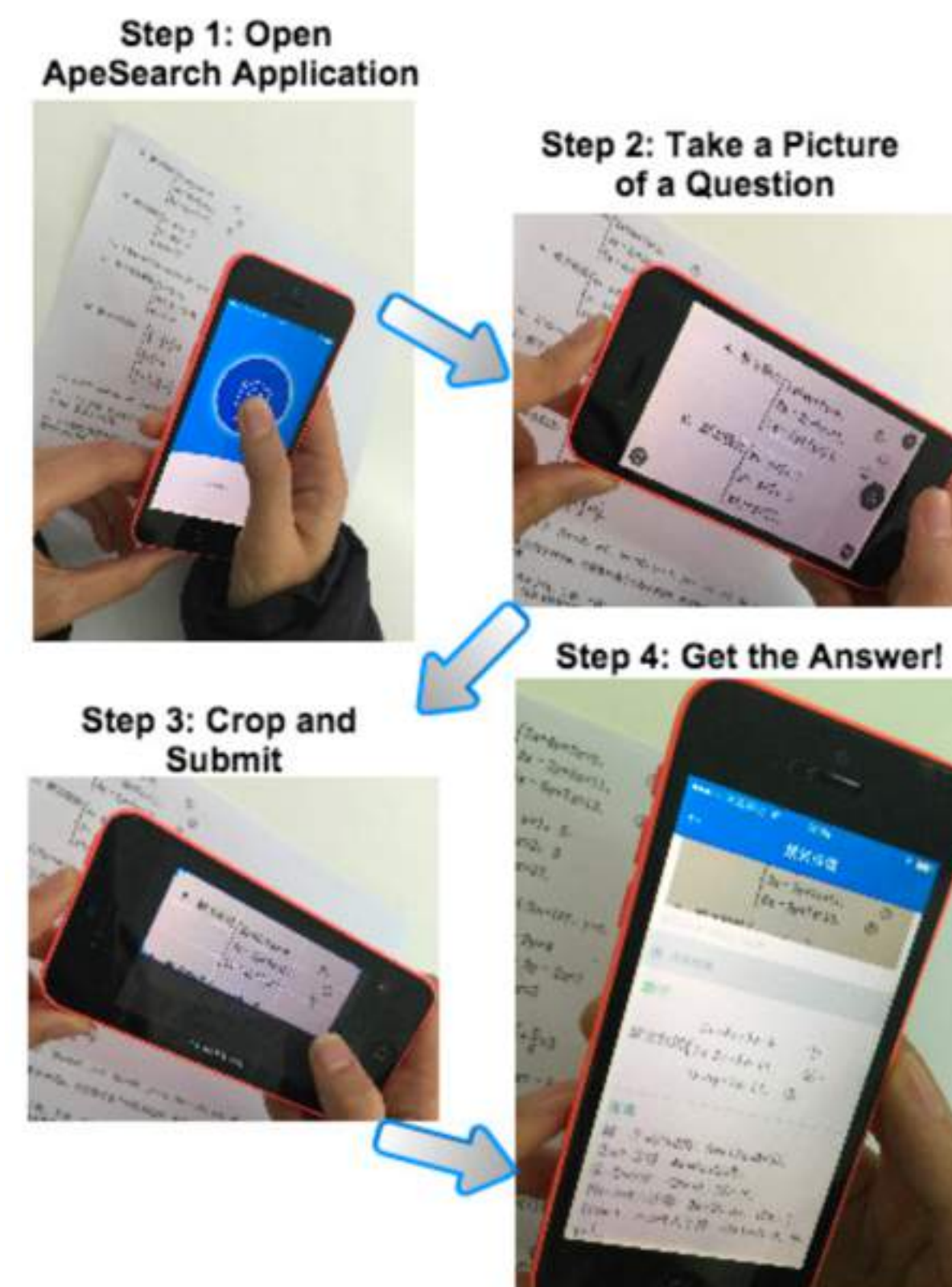
- 小猿搜题
  - ➡小猿搜题
  - ➡英语作文拍照批改
  - ➡口算检查
  - ➡古诗文背诵
  - ➡……

共同点?

图像/音频 ➡ 识别结果 ➡ high-level 的评判

场景
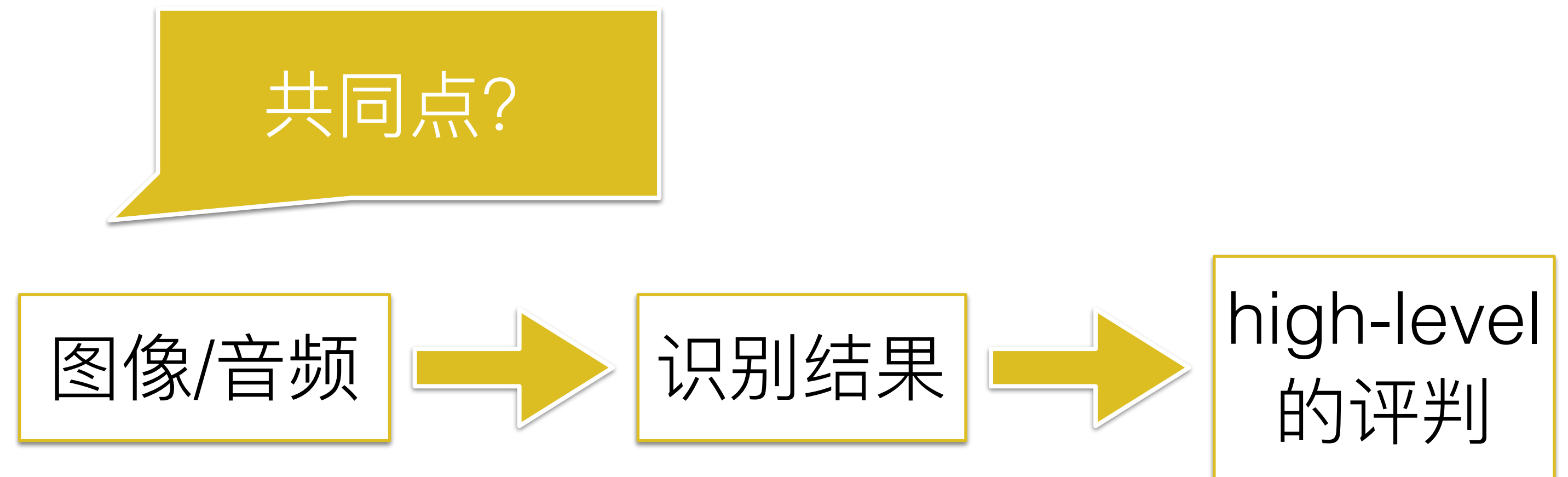
- 小猿搜题

  ➡小猿搜题

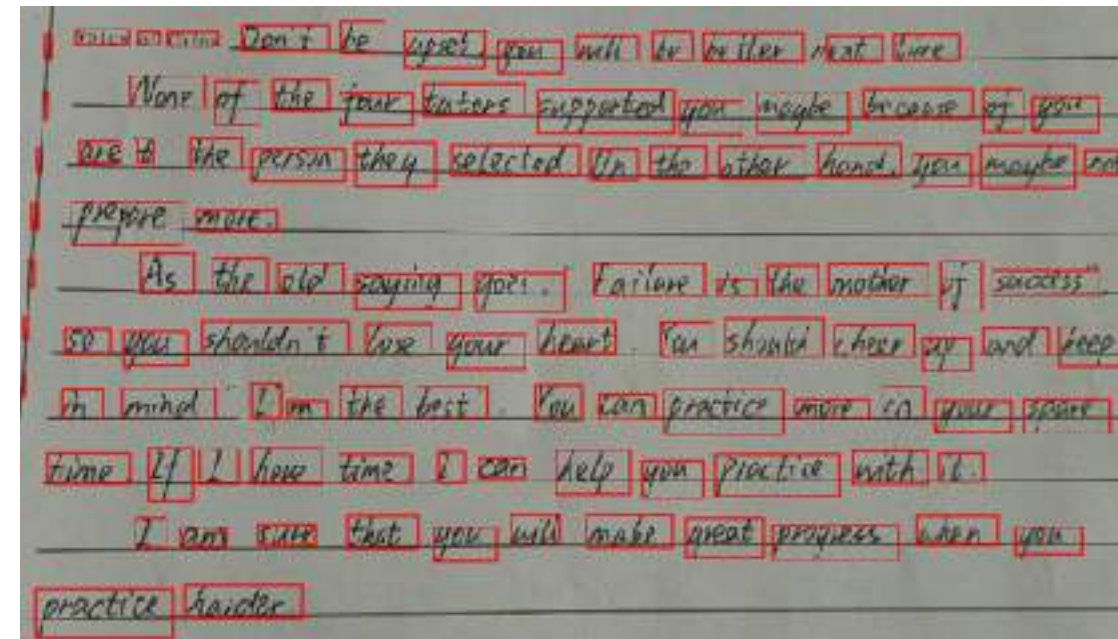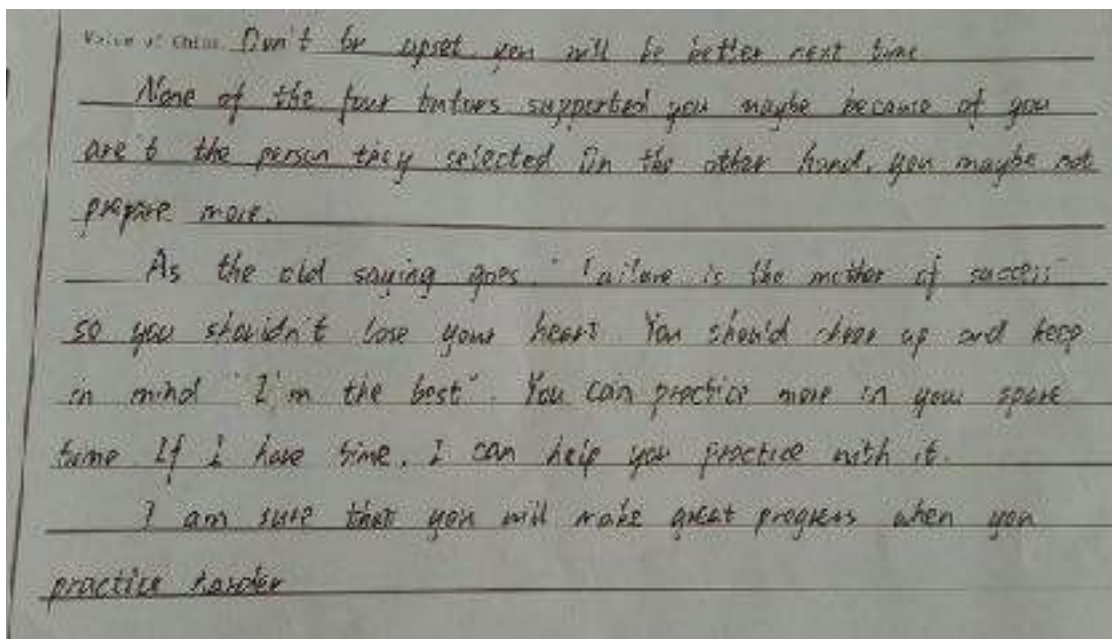 ➡英语作文拍照批改 ⟵——今天的落地实例！

  ➡口算检查

  ➡古诗文背诵

  ➡……

英语作文拍照批改

清晰整洁的图片

准确的单词切分

英文字符识别

批改

- ✦ 单词拼写纠错
- ✦ 单词搭配纠错
- ✦ 语法纠错
- ✦ 高级用法加分
- ✦ 评分

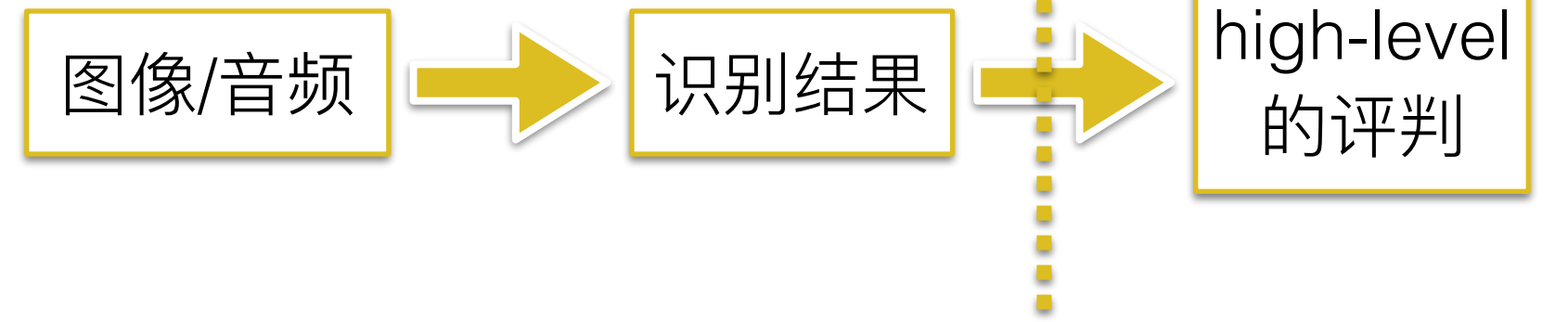Don't be upset, you will be better next time. ……

评分: **67.4**

评语: 全文结构较为严谨，再增加词汇量积累就更好了，高级词汇使用不是很多，建议增加衔接词汇的积累，注意不要过多使用简单句。

Do not be upset ,you will be better next time .None of the four tutors supported you maybe because of you are not the person they selected On the other

[语法错误]: because后面跟句子不需要加of。

As the old saying goes, Failure is the mother of success so you should not lose your heart .You should cheer up and keep in mind, I am the best. You can practice more in your spare time. If I have time, I can help you practice with it. I am sure that you will make great progress when you practice harder

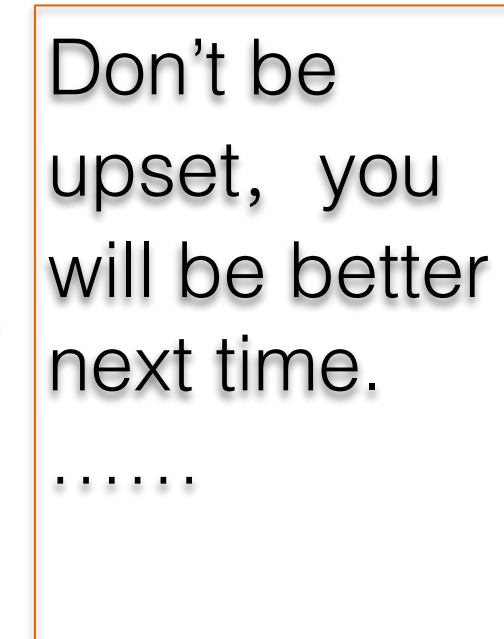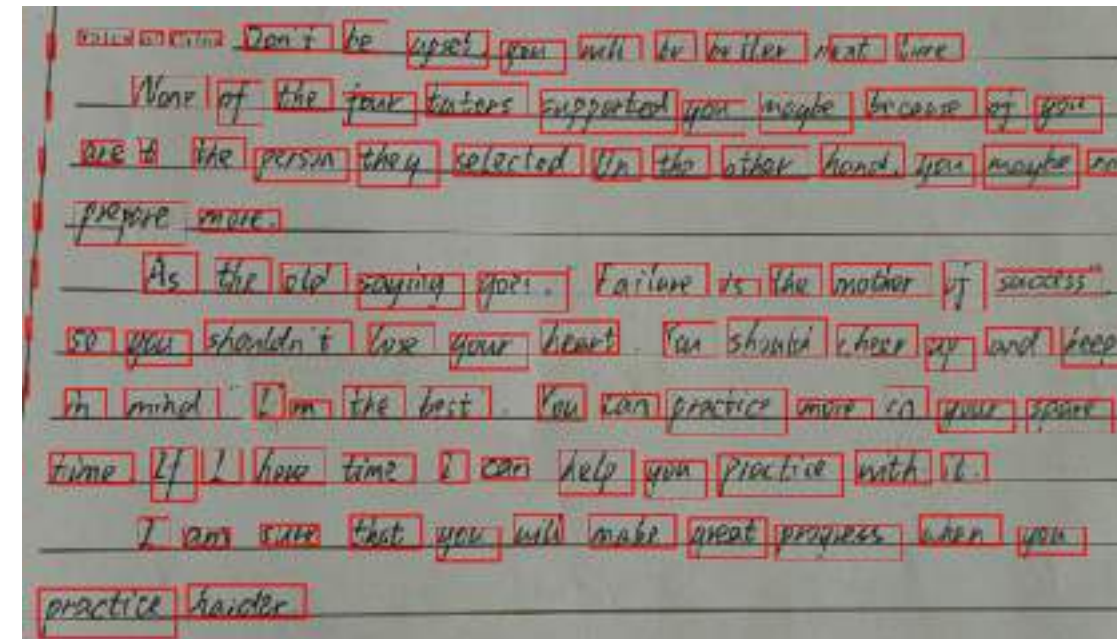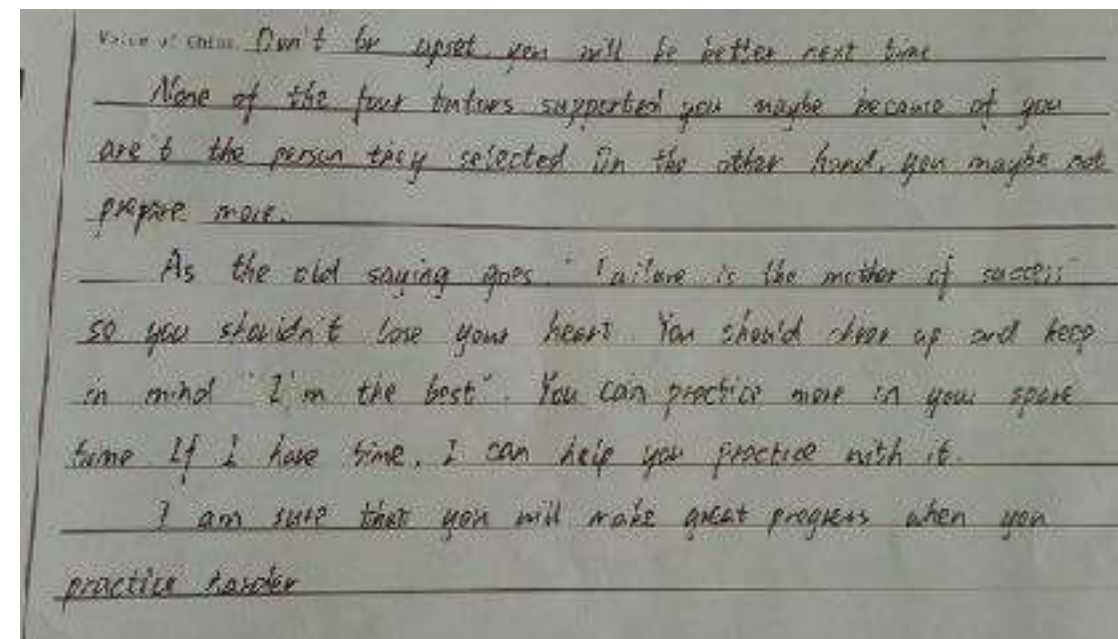"机器学习落地" ==> "机器学习（中的深度学习）落地"

场景

What is your solution?

Don't be upset, you will be better next time.
……

场景

解决方案

再看一次场景

解决方案的落地

回顾

解决方案

Don't be upset, you will be better next time.
……
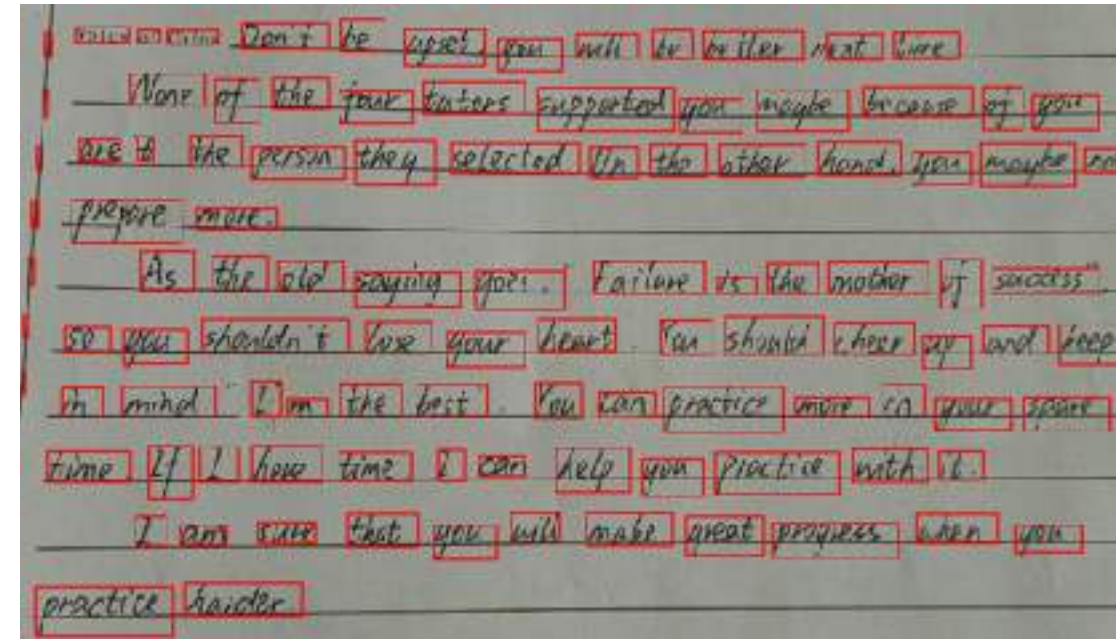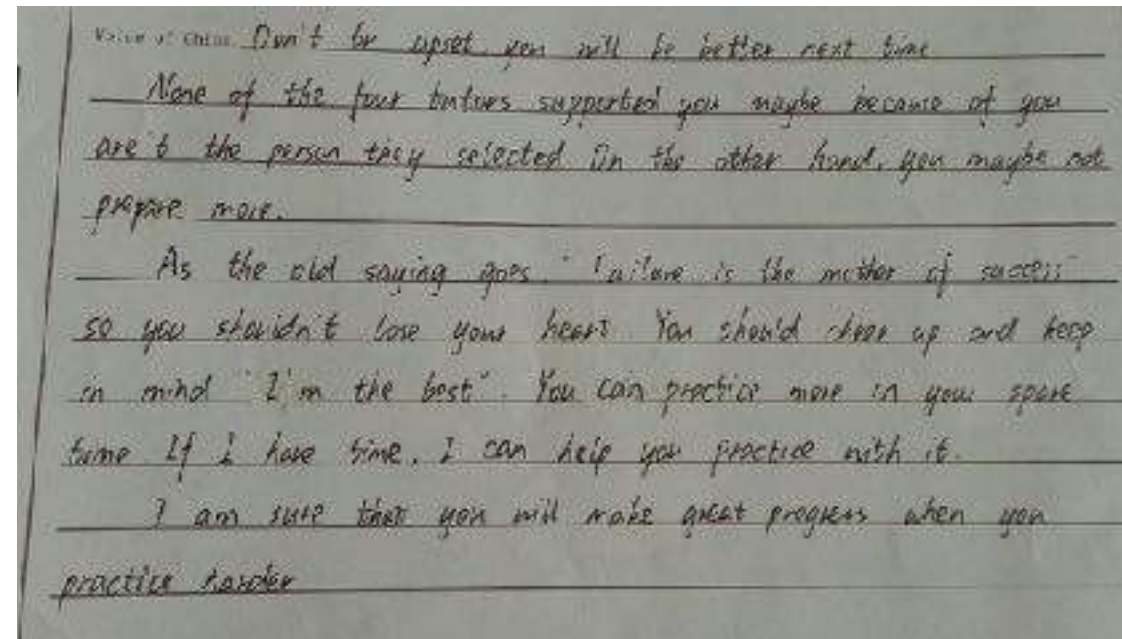
充分的单词框标注数据
+
*detection*模型

充分的字符标注数据
+
识别模型*(CNN, RNN, CNN+RNN, …)*

解决方案

为了能快速落地…

传统方法：
1. 规则模型（如，对于图像来说基于连通域的规则）
2. 人工提取feature的判别模型
3. ……

传统方法
快速出baseline结果

→

积累标注的Dataset

→

DL模型
训练

更准确的结果、
替代传统模型

补刀

解决方案

最后的模型很可能是

(传统方法 ⊕ DL模型 ⊕ (传统方法补刀) ⊕ DL模型 ⊕ ⋯

场景

解决方案

再看一次场景

解决方案的落地

回顾

再看一次场景 难点！
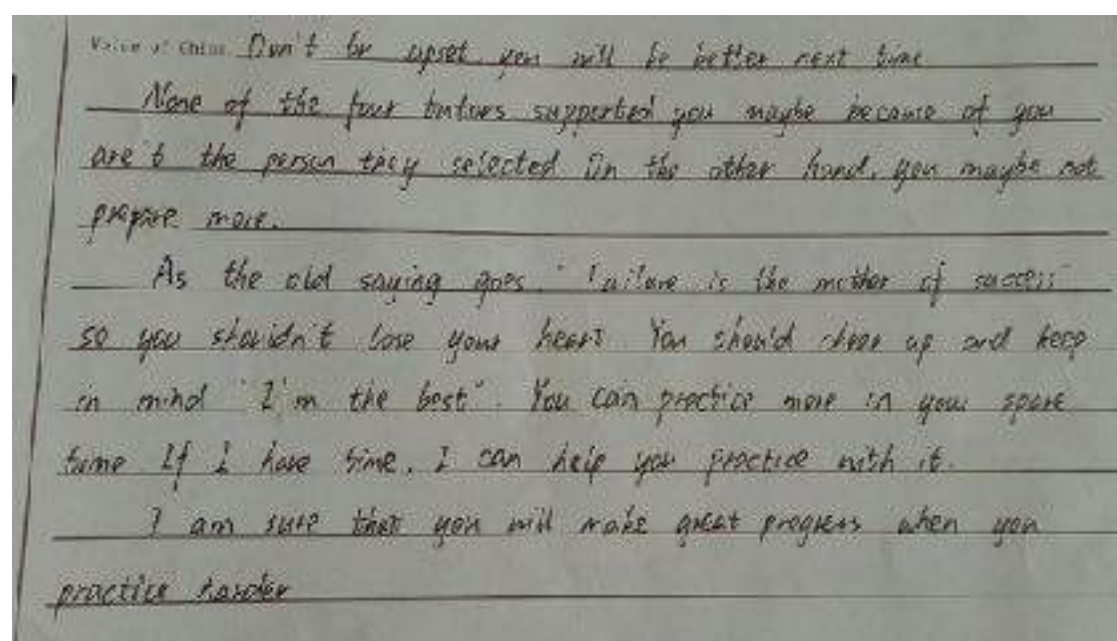
+光照影响
+模糊
+倾斜

→ 常见的真实
图片状况

+印刷体的干扰 （印刷体应不参与评分）
+批改笔记的干扰
+删除字符

→ 英语批改任务
相关的干扰项

（+横格线的干扰）
（+串行对顺序的干扰）

难点！

场景

解决方案

再看一次场景

解决方案的落地

回顾

传统方法baseline
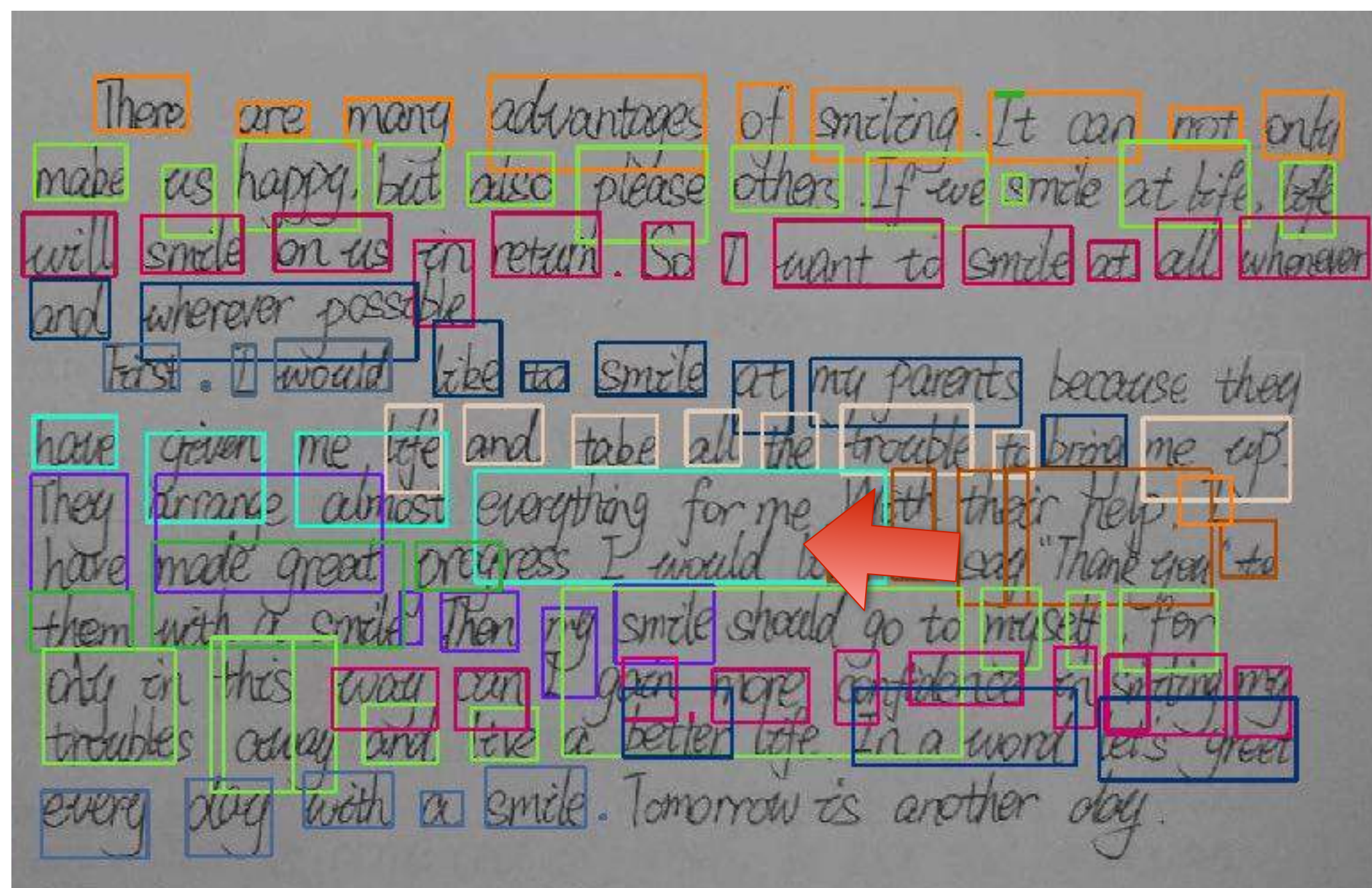
图片

二值化 （去横格线）

"行"切分 （外接矩形框）

行内单词切分

传统方法baseline

光照影响
模糊
倾斜
印刷体的干扰
批改笔迹的干扰
删除字符
（横格线的干扰）
（串行）

图片

⬇

二值化（去横格线）

⬇

"行"切分（外接矩形框）

⬇

行内单词切分

*粘连判断

⬇

*粘连切分

⬇

传统方法baseline

☑光照影响
☑模糊
☑倾斜
☑印刷体的干扰
☑批改笔迹的干扰
☑删除字符
☑（横格线的干扰）
☑（串行）

图片
↓
二值化（去横格线）
↓
"行"切分（外接矩形框）
↓
行内单词切分

↓
*粘连判断
↓
*粘连切分
↓
切分后box分类

↓
手写单词块
↓
英文字符
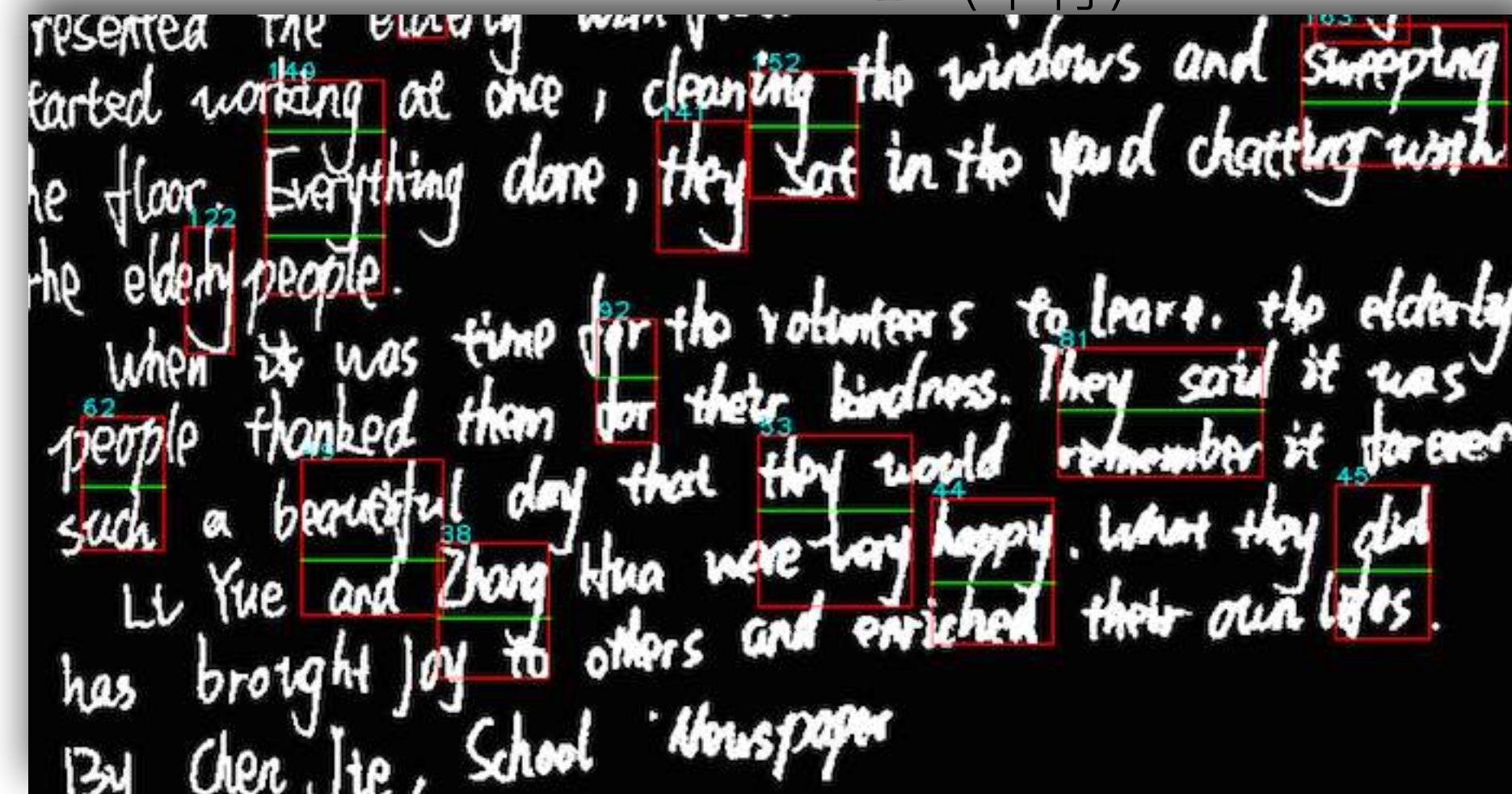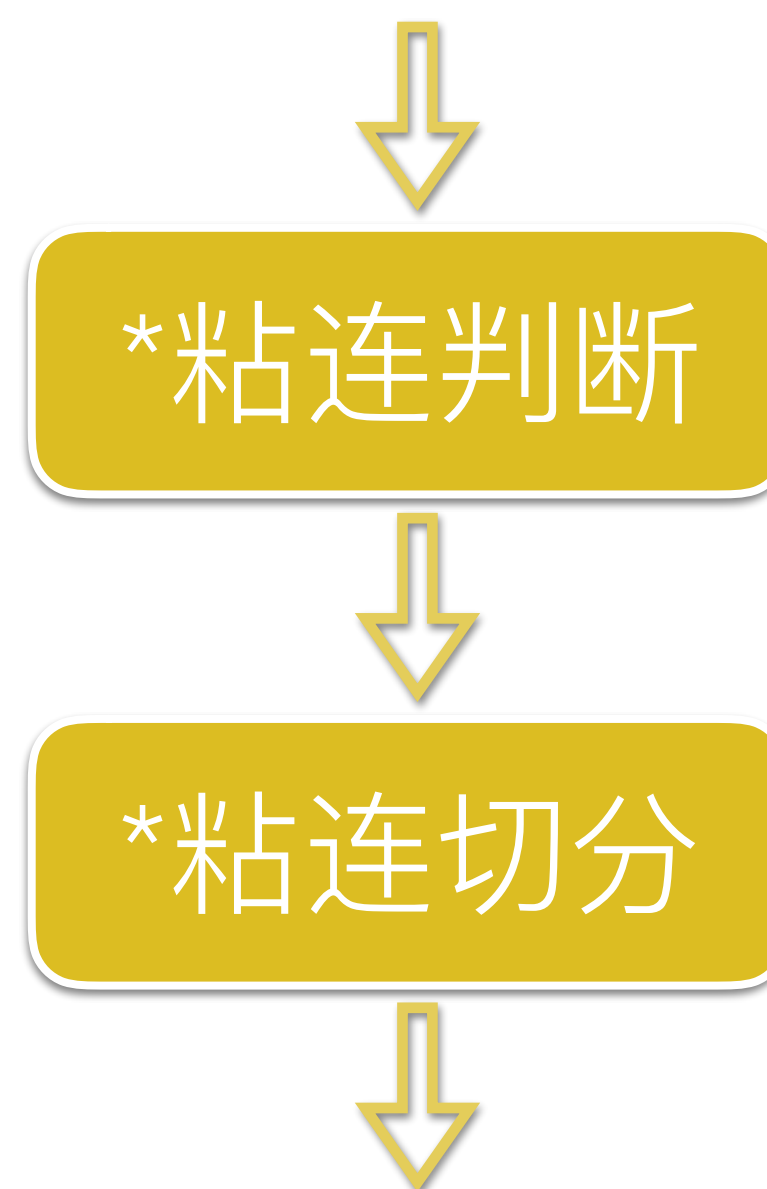
Class 1: 印刷体区域
Class 2: 删除区域
Class 3: 批改区域
Class 4: 常规英文手写区域
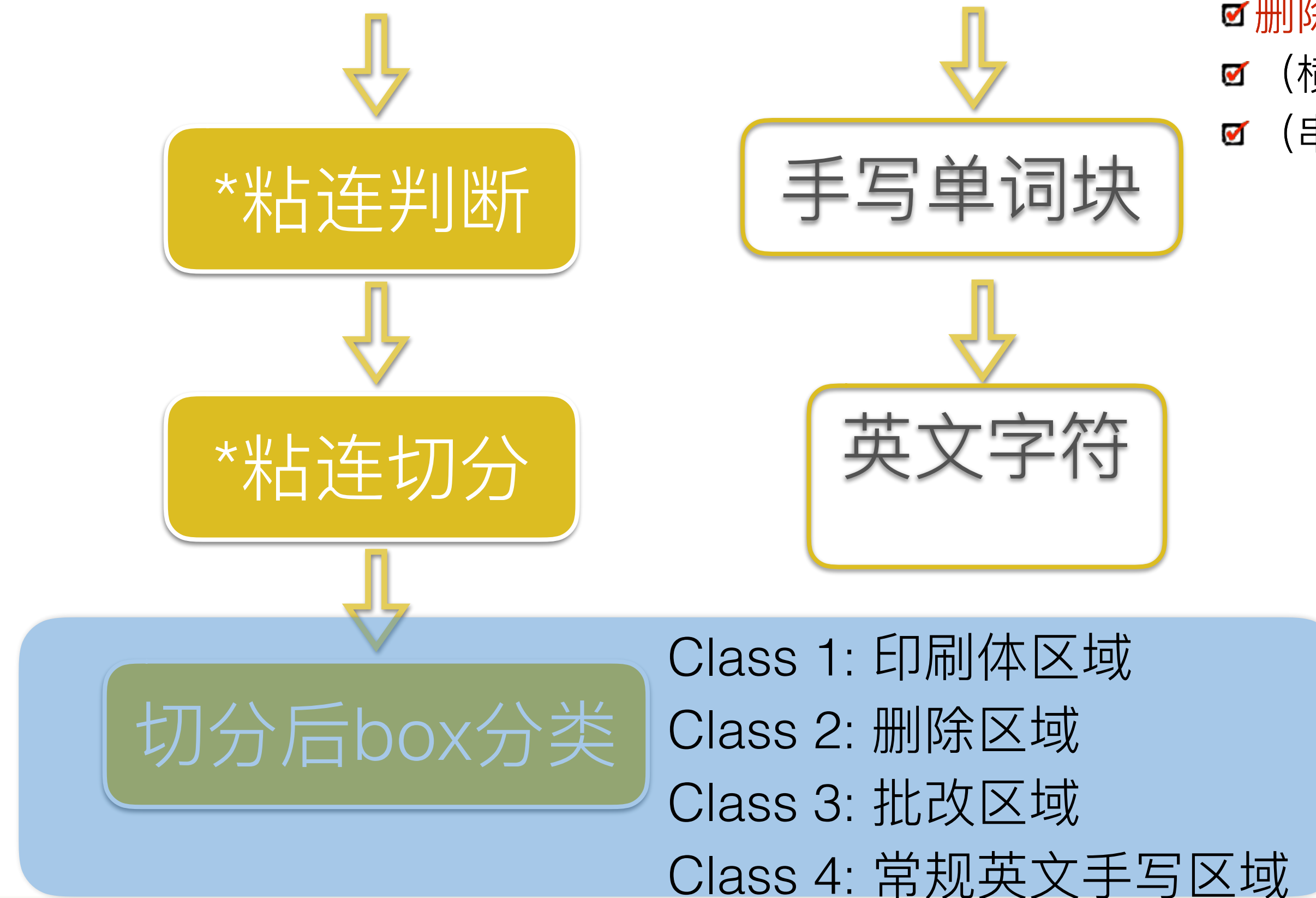
从何处推进DL模型的落地？

传统模型不够简洁优美？实用性上考虑！

传统方法baseline

☑ 光照影响
☑ 模糊
☑ 倾斜
☑ 印刷体的干扰
☑ 批改笔迹的干扰
☑ 删除字符
☑ （横格线的干扰）
☑ （串行）

图片
↓
二值化（去横格线）
↓
"行"切分（外接矩形框）
↓
行内单词切分

↓
*粘连判断
↓
*粘连切分
↓
切分后box分类

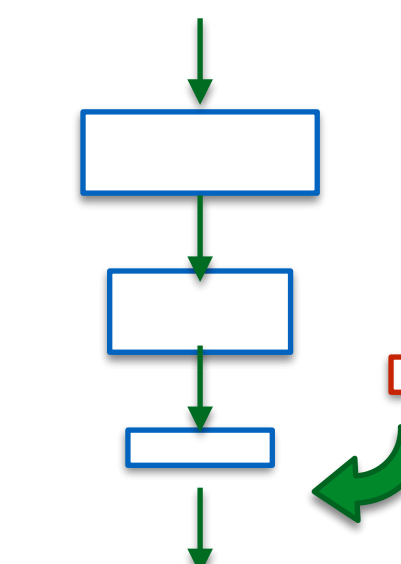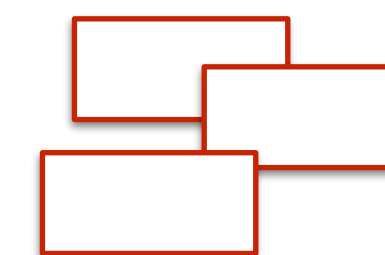手写单词块
↓
英文字符

Class 1: 印刷体区域
Class 2: 删除区域
Class 3: 批改区域
Class 4: 常规英文手写区域

DL模型替代

切分后box分类

Class 1: 印刷体区域
Class 2: 删除区域
Class 3: 批改区域
Class 4: 常规英文手写区域

手写单词
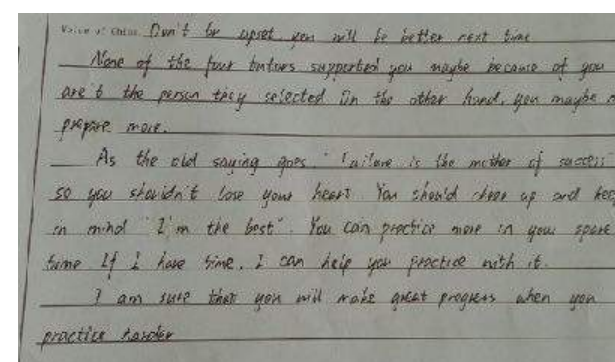
多类别分类
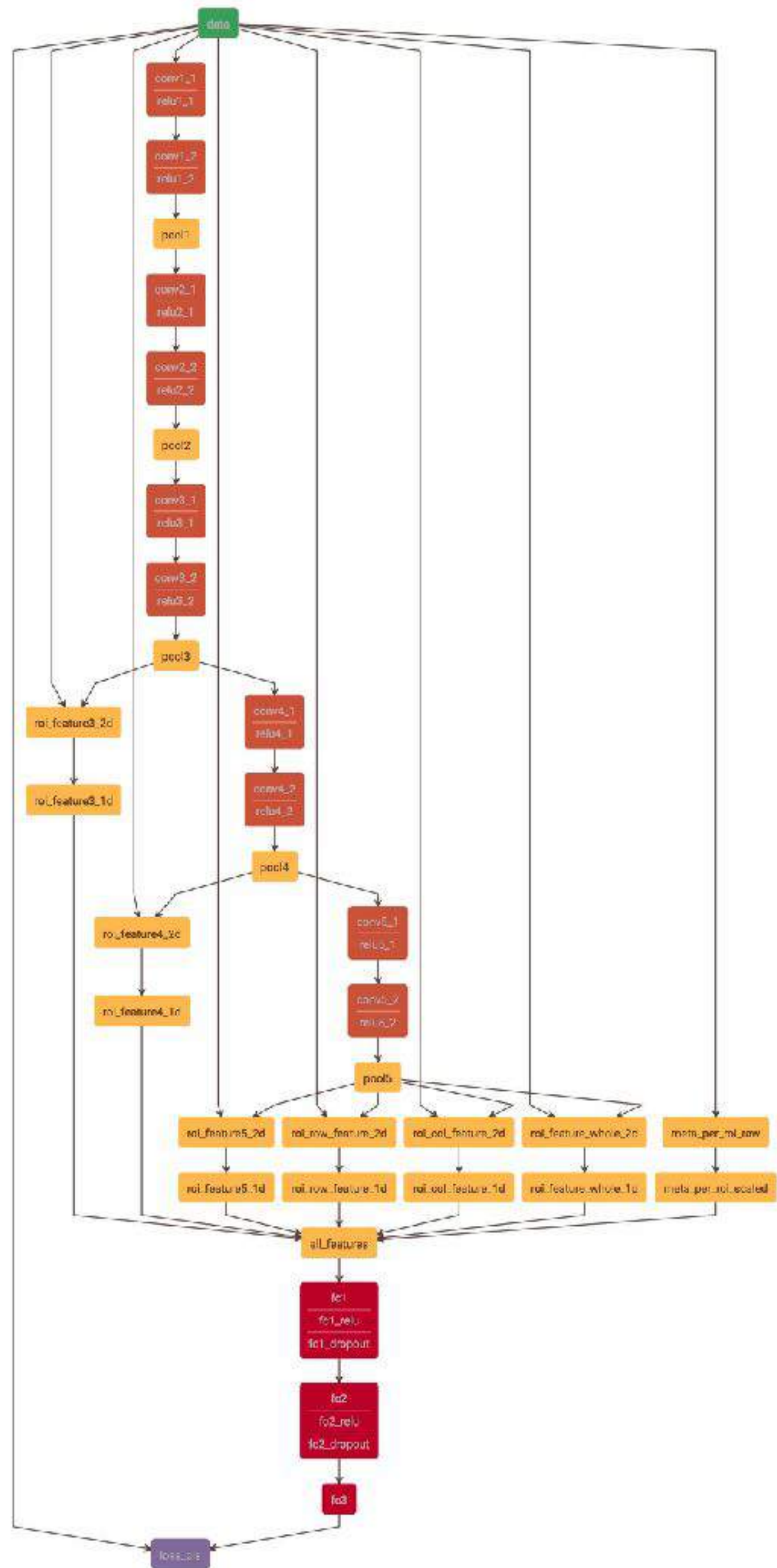
DL模型替代

图片

↓

二值化 （去横格线）

↓

"行"切分 （外接矩形框）

↓

行内单词切分

↓

*粘连判断

↓

*粘连切分

↓

切分后box分类的DL模型

↓

手写单词块

↓

英文字符

☑光照影响
☑模糊
☑倾斜
☑印刷体的干扰
☑批改笔迹的干扰
☑删除字符
☑ （横格线的干扰）
☑ （串行）

AiCon
全球人工智能与机器学习技术大会

主办方： Geekbang 极客邦科技 InfoQ

DL的推进

标注数据集

对象：每个现有框

分类
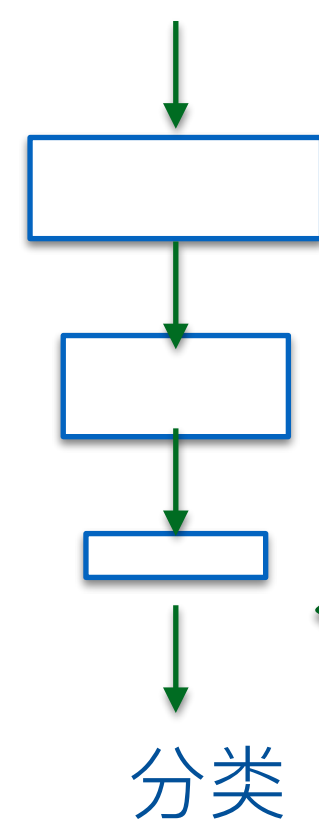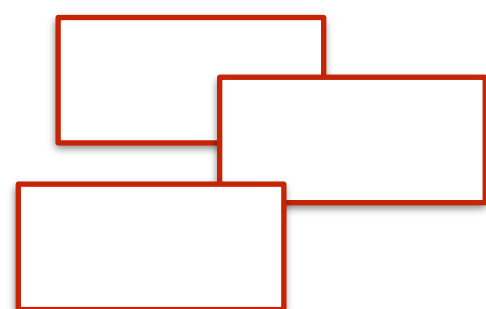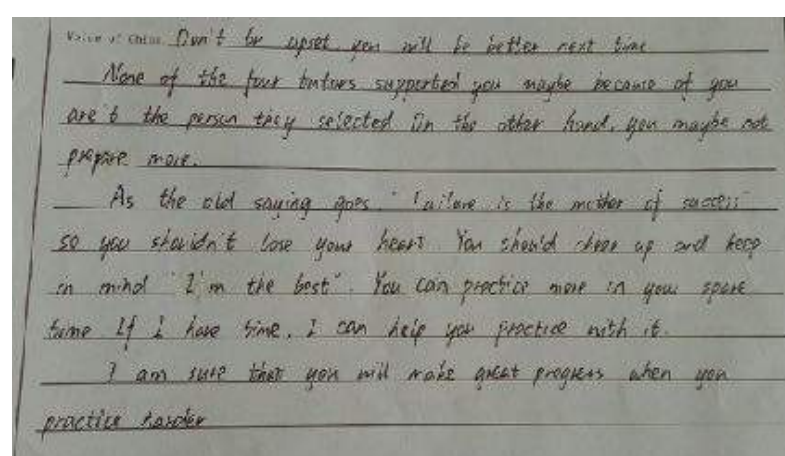
feature map上每一点（原图8x8区域）的"是/不是英文手写"的分类

对象：原图每8x8像素块

```
# Mapping 7
n.map_conv7_1, n.map_relu7_1 = conv_relu(n.pool7, num_output=512)
n.up_conv7, n.up_relu7 = deconv_relu(n.map_relu7_1, num_output=512,
        kernel_size=16, stride=16)

# Mapping 6
n.map_conv6_1, n.map_relu6_1 = conv_relu(n.pool6, num_output=512)
n.up_conv6, n.up_relu6 = deconv_relu(n.map_relu6_1, num_output=512,
        kernel_size=8, stride=8)

# Mapping 5
n.map_conv5_1, n.map_relu5_1 = conv_relu(n.pool5, num_output=512)
n.up_conv5, n.up_relu5 = deconv_relu(n.map_relu5_1, num_output=512,
        kernel_size=4, stride=4)

# Mapping 4
n.map_conv4_1, n.map_relu4_1 = conv_relu(n.pool4, num_output=256)
n.up_conv4, n.up_relu4 = deconv_relu(n.map_relu4_1, num_output=256)

# Mapping 3
n.map_conv3_1, n.map_relu3_1 = conv_relu(n.pool3, num_output=128)
n.map_conv3_2, n.map_relu3_2 = conv_relu(n.map_relu3_1, num_output=128)
n.concated_all = L.Concat(n.map_relu3_2, n.up_relu4, n.up_relu5, n.up_relu6, n.up_relu7)

# fc 1
n.fc_conv1_1, n.fc_relu1_1 = conv_relu(n.concated_all, num_output=1024)
n.fc_conv1_2, n.fc_relu1_2 = conv_relu(n.fc_relu1_1, num_output=1024)
n.fc_conv1_3, n.fc_relu1_3 = conv_relu(n.fc_relu1_2, num_output=4)
```
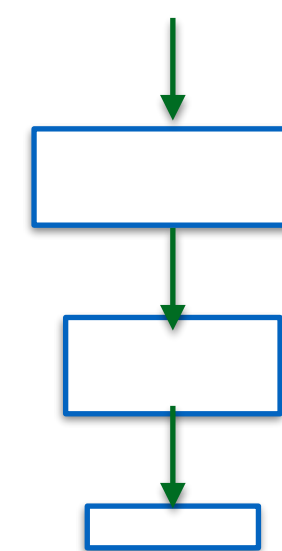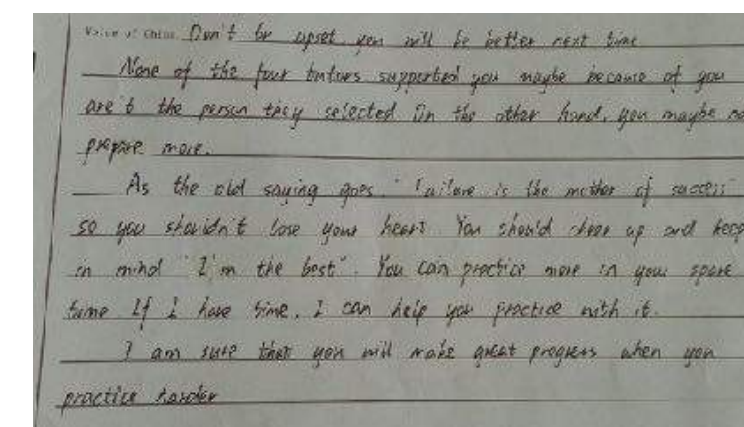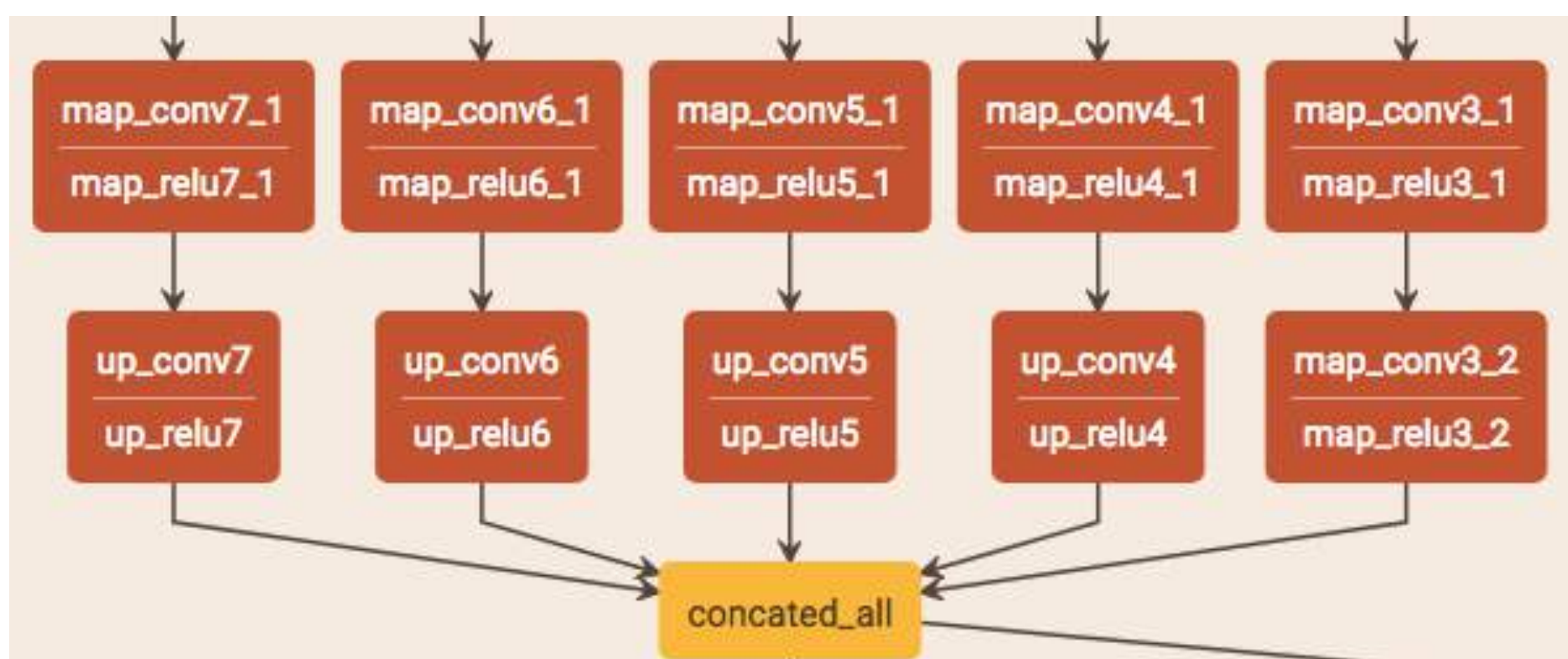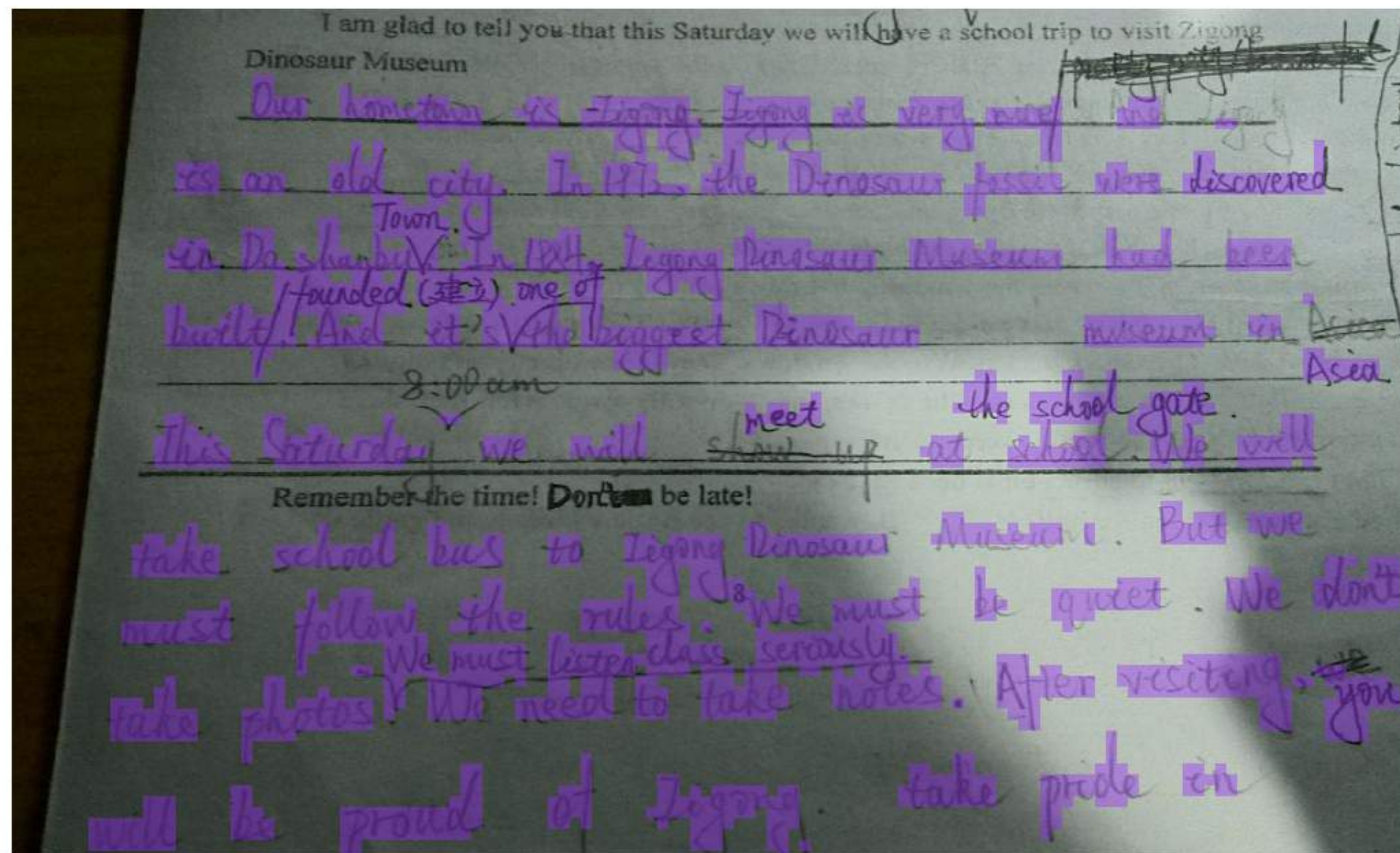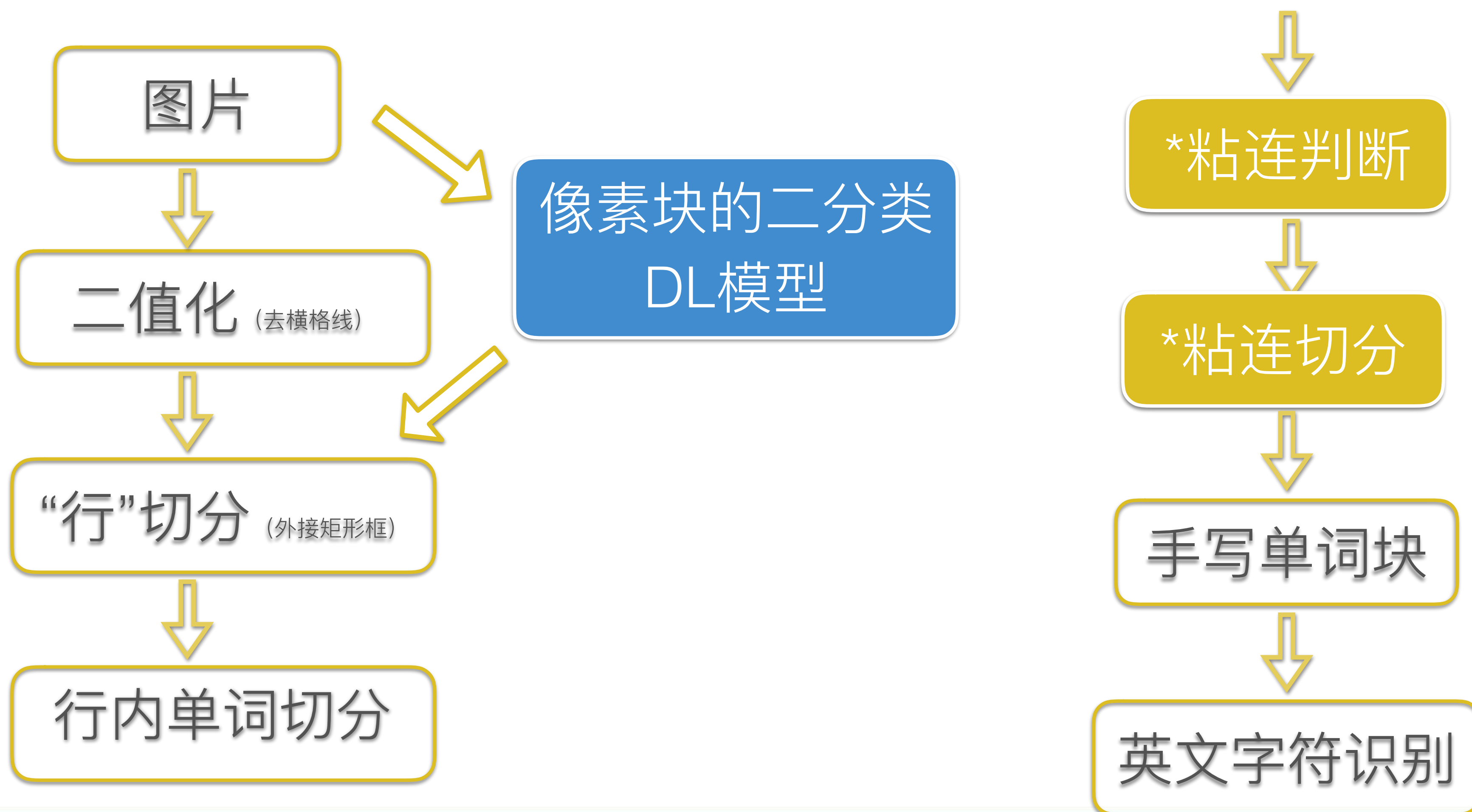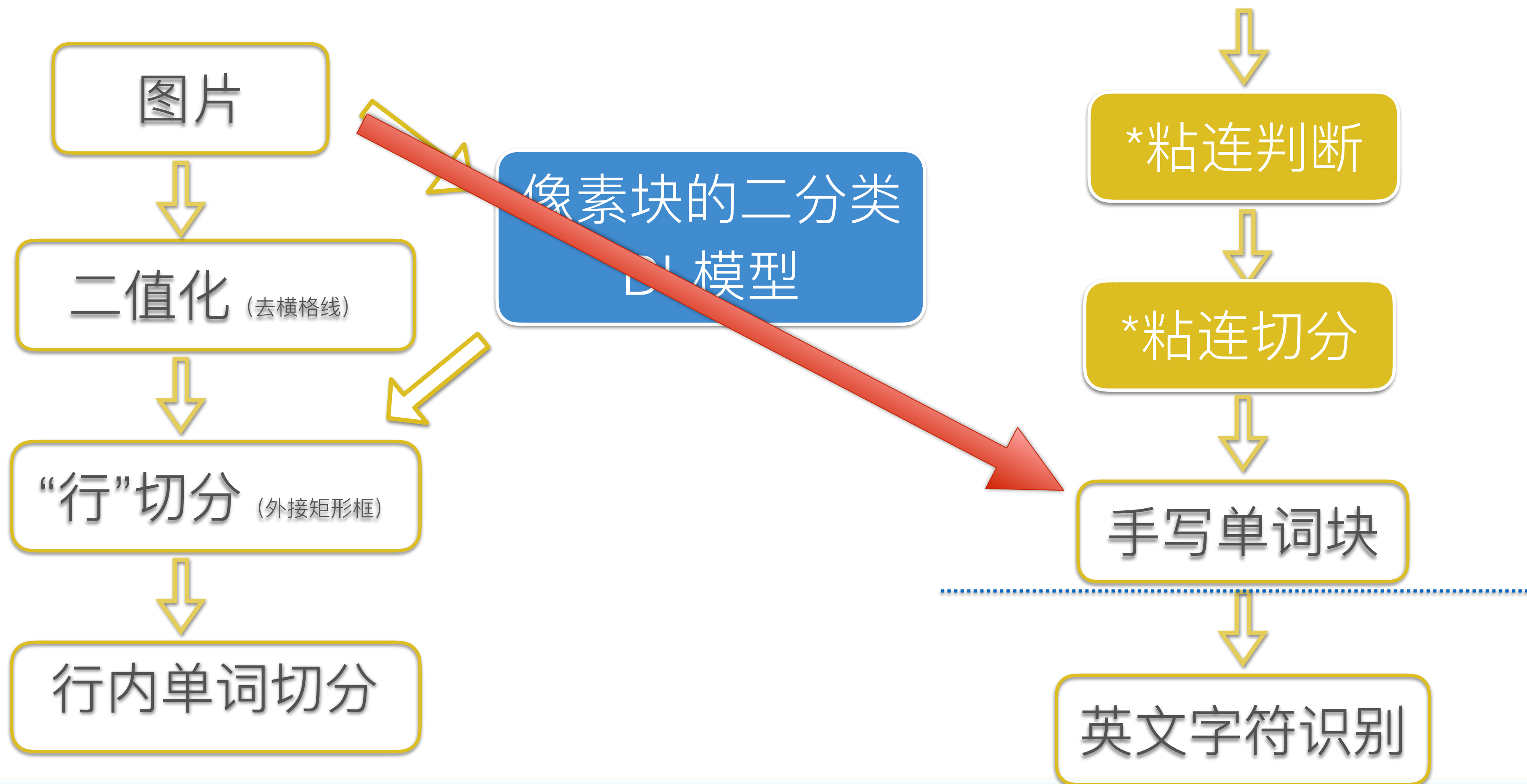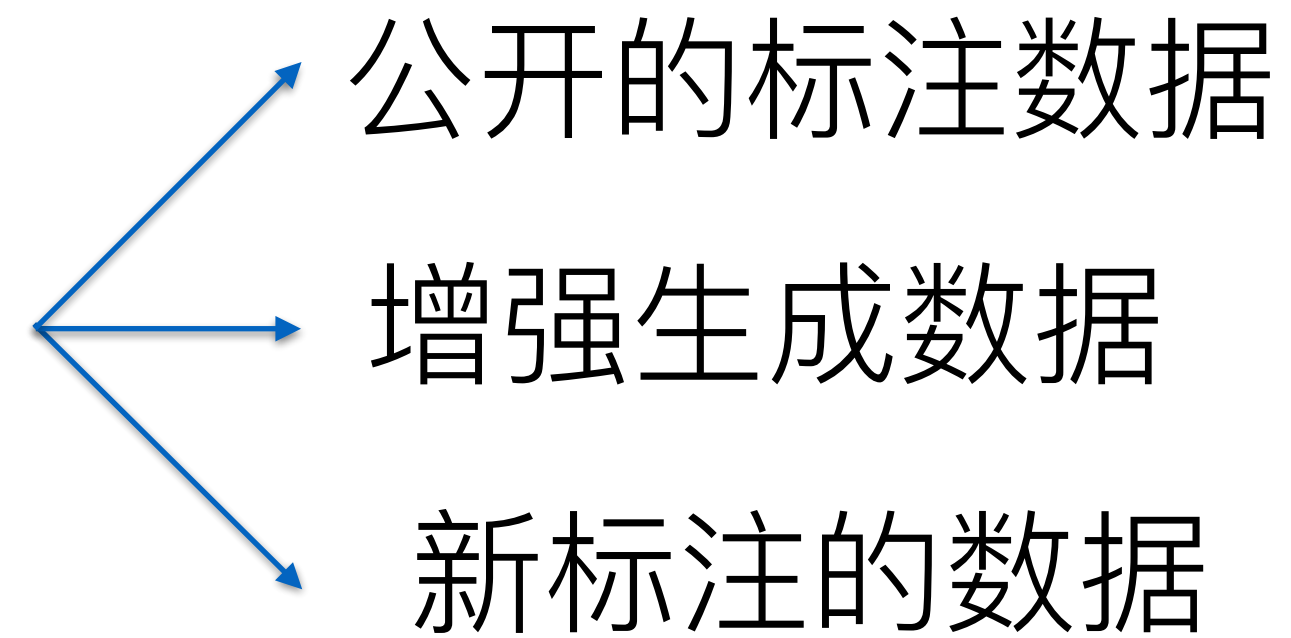
DL的推进

图片 →

像素块的二分类
DL模型

二值化（去横格线）

"行"切分（外接矩形框）

行内单词切分

*粘连判断

*粘连切分

手写单词块

英文字符识别

进一步推进？

图片

↓

二值化（去横格线）

↓

"行"切分（外接矩形框）

↓

行内单词切分

像素块的二分类 DL 模型

*粘连判断

↓

*粘连切分

↓

手写单词块

- - - - - - - - - - - - - - -

↓

英文字符识别

DL识别模型

手写单词块

⬇

英文字符识别

公开的标注数据

增强生成数据

新标注的数据

中国学生的英文手写vs. 外国人手写
【要符合你面对的真实数据的特点和分布】

- CNN
- RNN
- CNN+RNN
- CNN+RNN+CTC
- ……

场景

解决方案

再看一次场景

解决方案的落地

回顾

回顾

传统方法
快速出baseline结 → 积累Dataset → DL模型
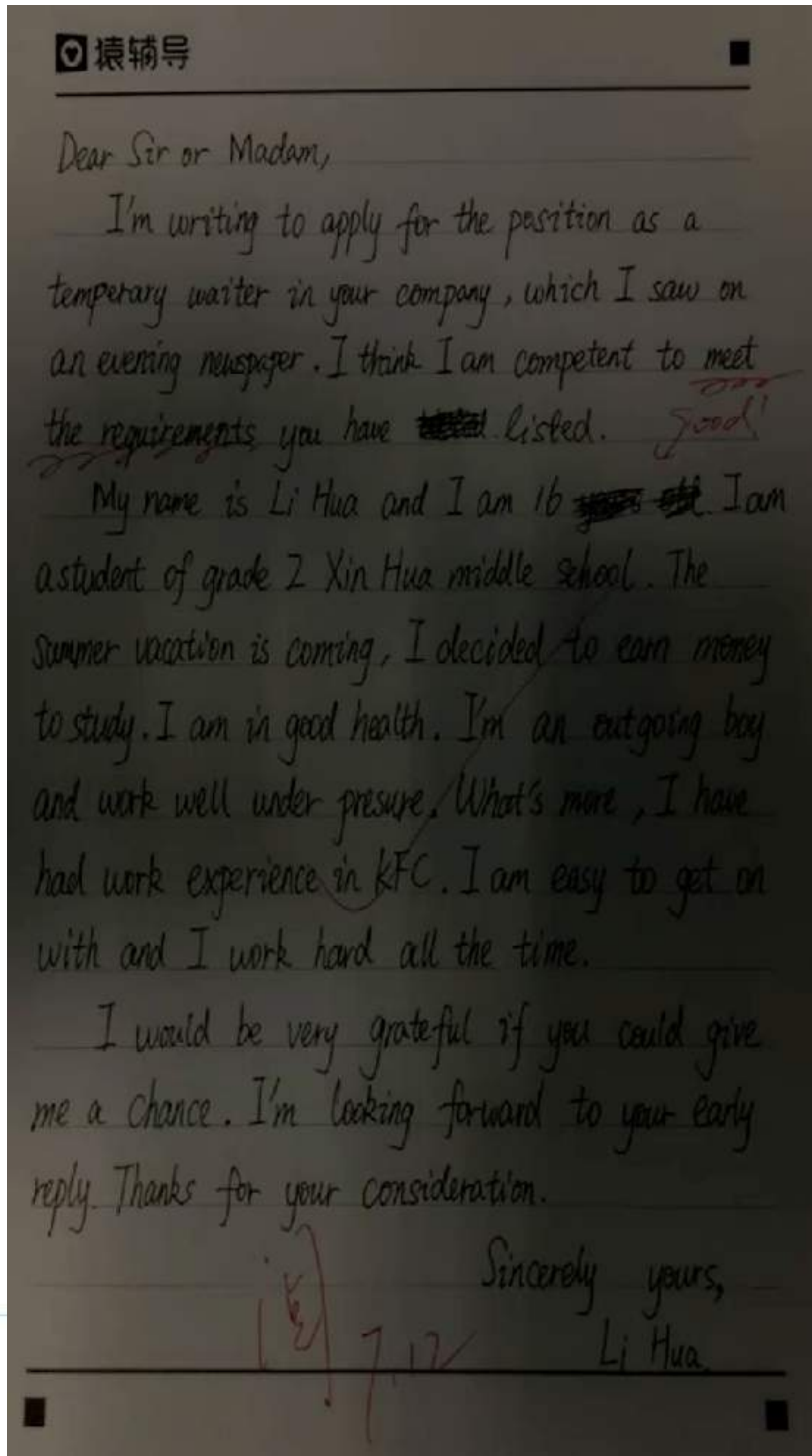训练

更准确的结果、
替代传统模型

补刀
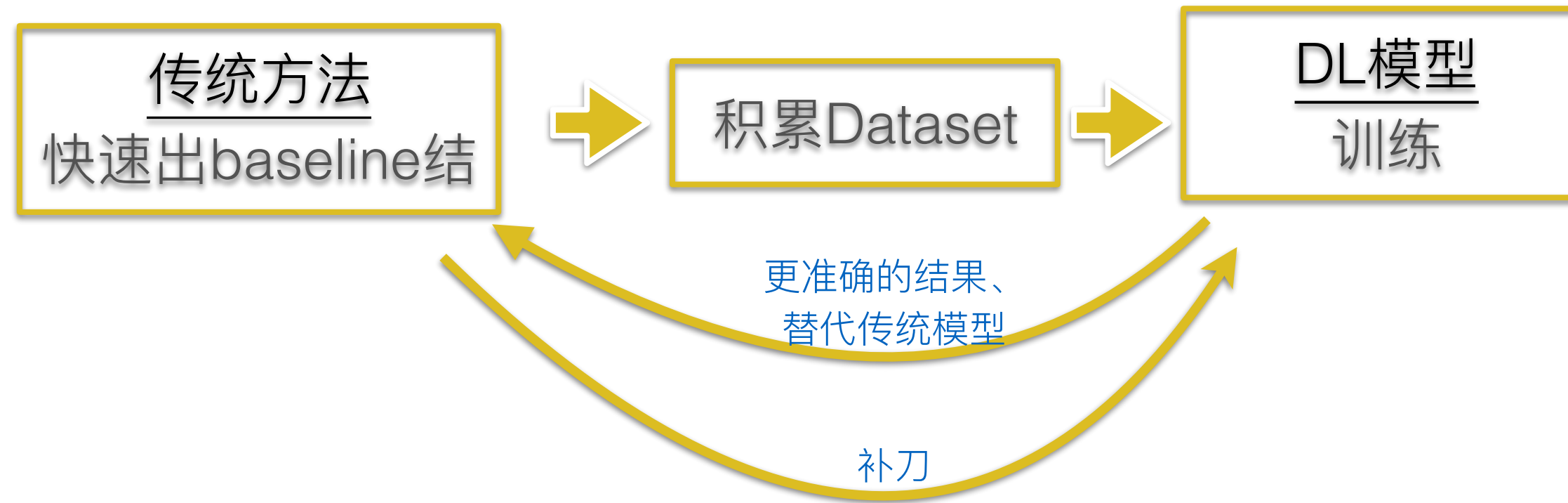
从何处推进DL模型的落地？传统模型不够简洁优美？实用性上考虑！✔

（传统方法）⊕ DL模型 ⊕（传统方法补刀）⊕ DL模型 ⊕ ···

落地

欢迎大家的
提问和交流！

落地

Thanks!

猿辅导