

Research Report

利用视觉语言模型解决文字检测中的歧义问题

2022年 6月

汇报人：舒言

AE TextSpotter: Learning Visual and Linguistic Representation for Ambiguous Text Spotting


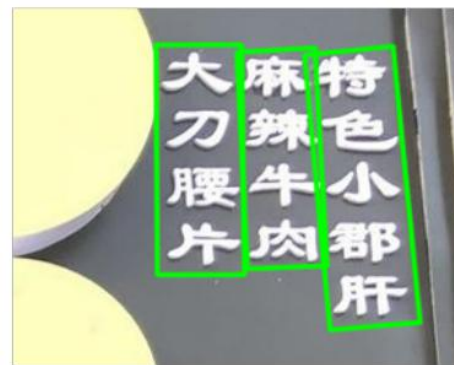
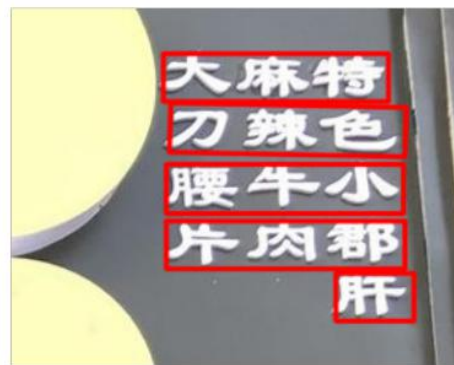
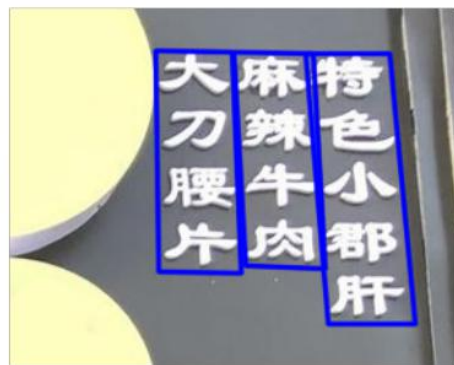
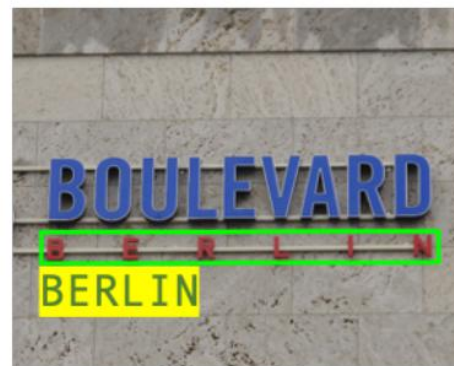
Wenhai Wang¹, Xuebo Liu², Xiaozhong Ji¹, Enze Xie³, Ding Liang²
Zhibo Yang⁴, Tong Lu^{1,✉}, Chunhua Shen⁵, and Ping Luo³

¹National Key Lab for Novel Software Technology, Nanjing University

²SenseTime Research ³The University of Hong Kong

⁴Alibaba-Group ⁵The University of Adelaide

wangwenhai362@smail.nju.edu.cn, {liuxuebo, liangding}@sensetime.com,
shawn_ji@163.com, xieenze@hku.hk, zhibo.yzb@alibaba-inc.com,
lutong@nju.edu.cn, chunhua.shen@adelaide.edu.au, pluo@cs.hku.hk

 Ambiguity Ground Truth Incorrect Result Correct Result

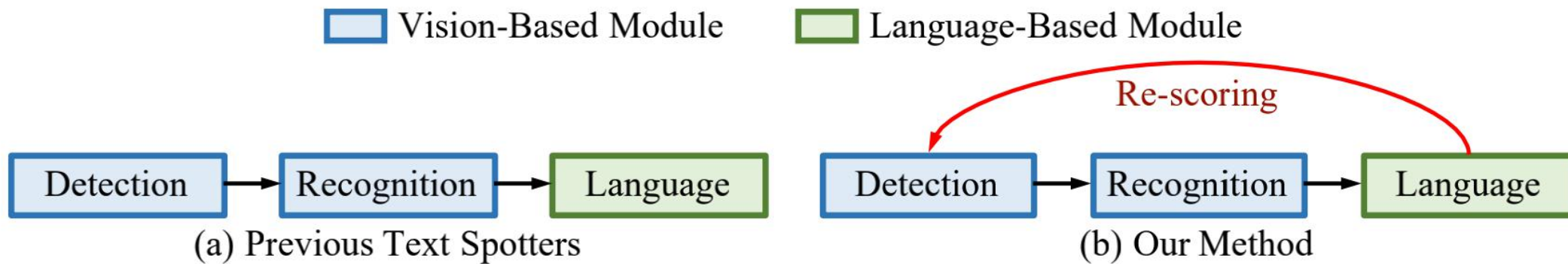
(a) Input Image

(b) Ground-Truth

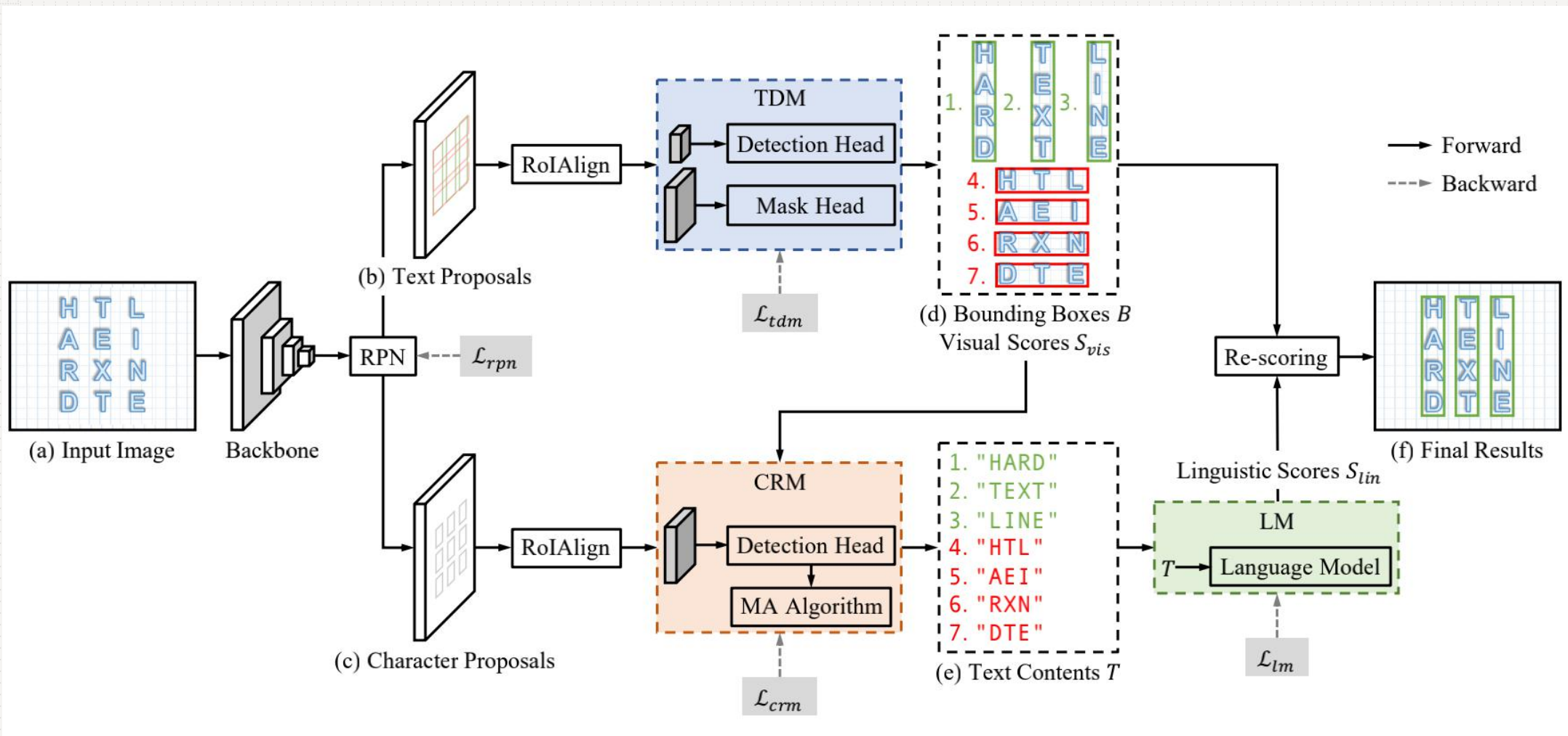
(c) Previous Method

(d) Our Method

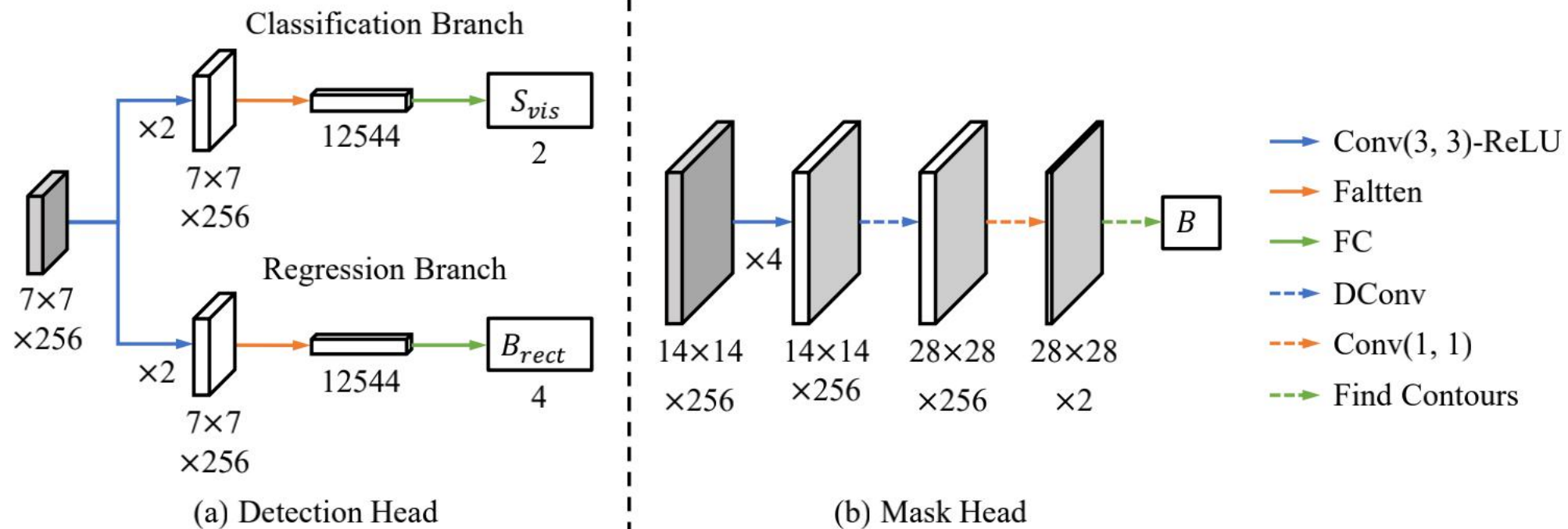
对于检测-识别-矫正框架的改善

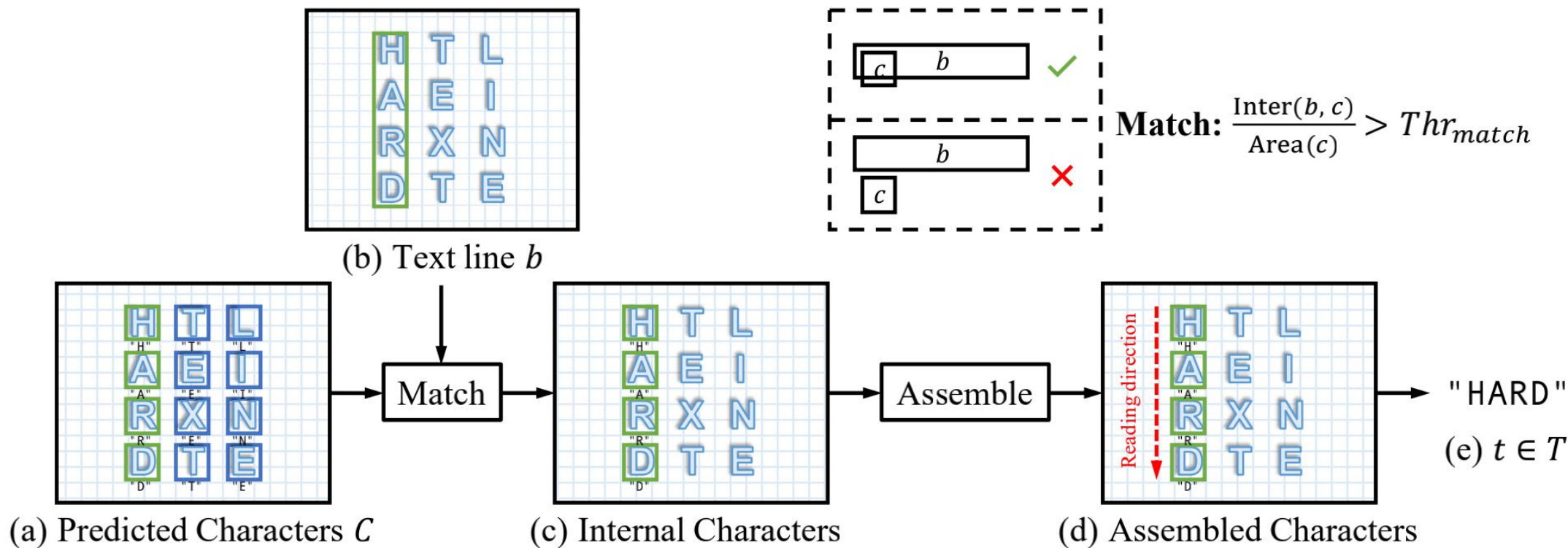


1. 首次在文字检测任务中引入了语言的表示，用以解决检测的歧义性问题。
2. 提出了新的字符识别框架用于文字识别，速度较快
3. 实验结果取得IC19-RECTS的SOTA.



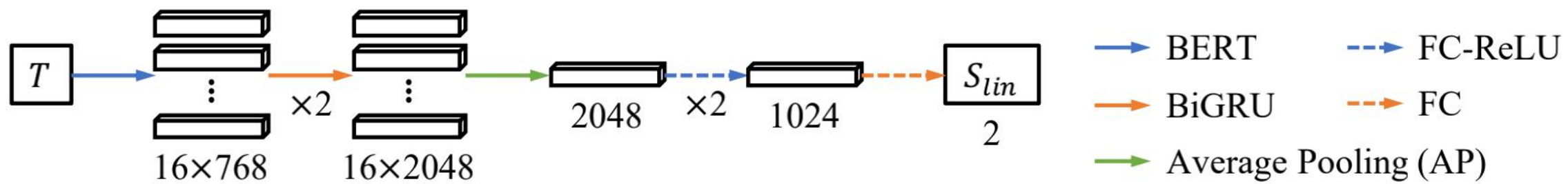
$$S = \lambda S_{vis} + (1 - \lambda) S_{tex}.$$





Character detection + Character classification

2 LM



$$\mathcal{L}_{vis} = \mathcal{L}_{rpn} + \mathcal{L}_{tdm} + \mathcal{L}_{crm}$$

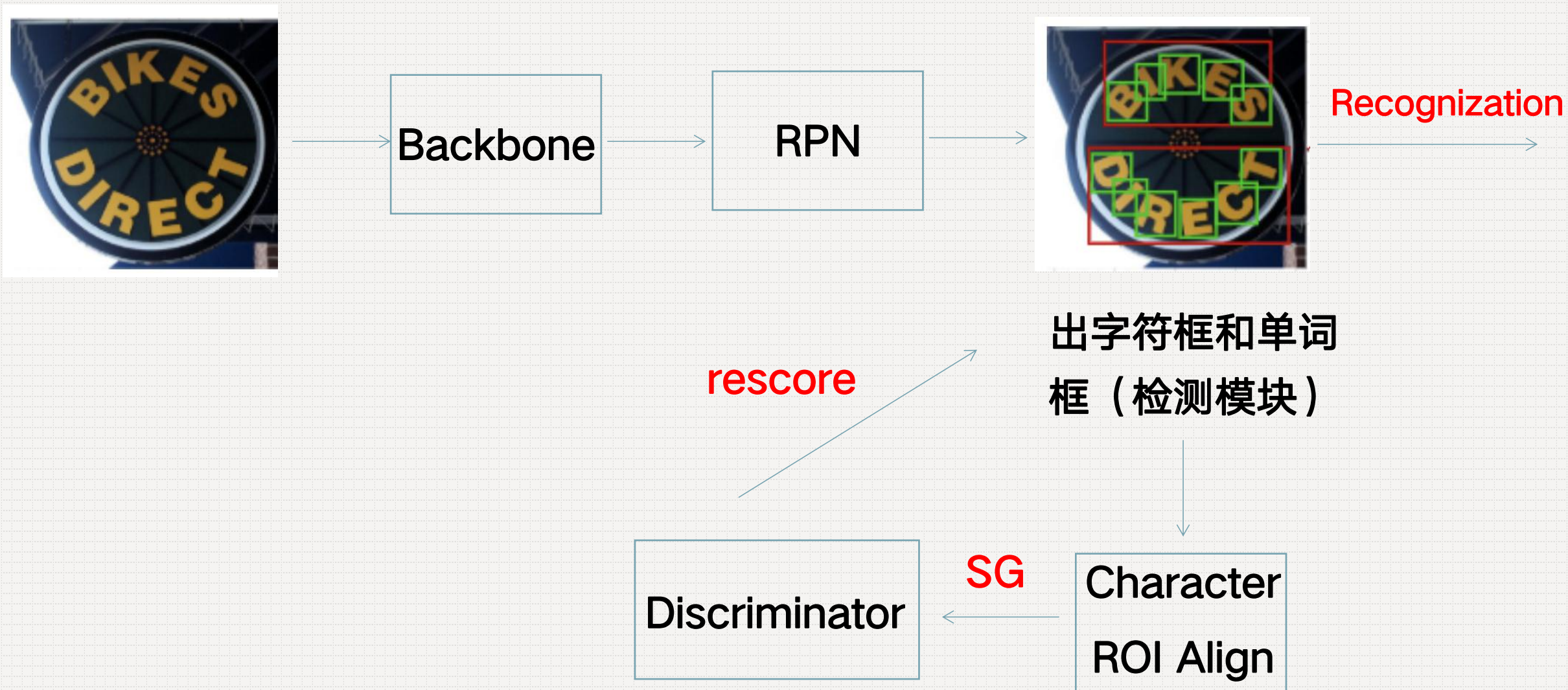
$$\mathcal{L}_{tdm} = \mathcal{L}_{tdm}^{cls} + \mathcal{L}_{tdm}^{box} + \mathcal{L}_{tdm}^{mask}$$

$$\mathcal{L}_{crm} = \mathcal{L}_{crm}^{cls} + \mathcal{L}_{crm}^{box}.$$

语言模块用交叉熵损失函数，两个模块分开独立训练。

对于每个检测出的候选框，都需要去进行识别出来看看是否合乎语义。

能否做到直接根据检测出的候选框，结合视觉和语义信息选取那些正确的候选框进行识别。（将语义判别器做进整个模型中）





(a)

RoIAlign



(b)



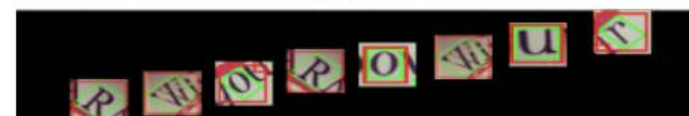
(c)



(d)



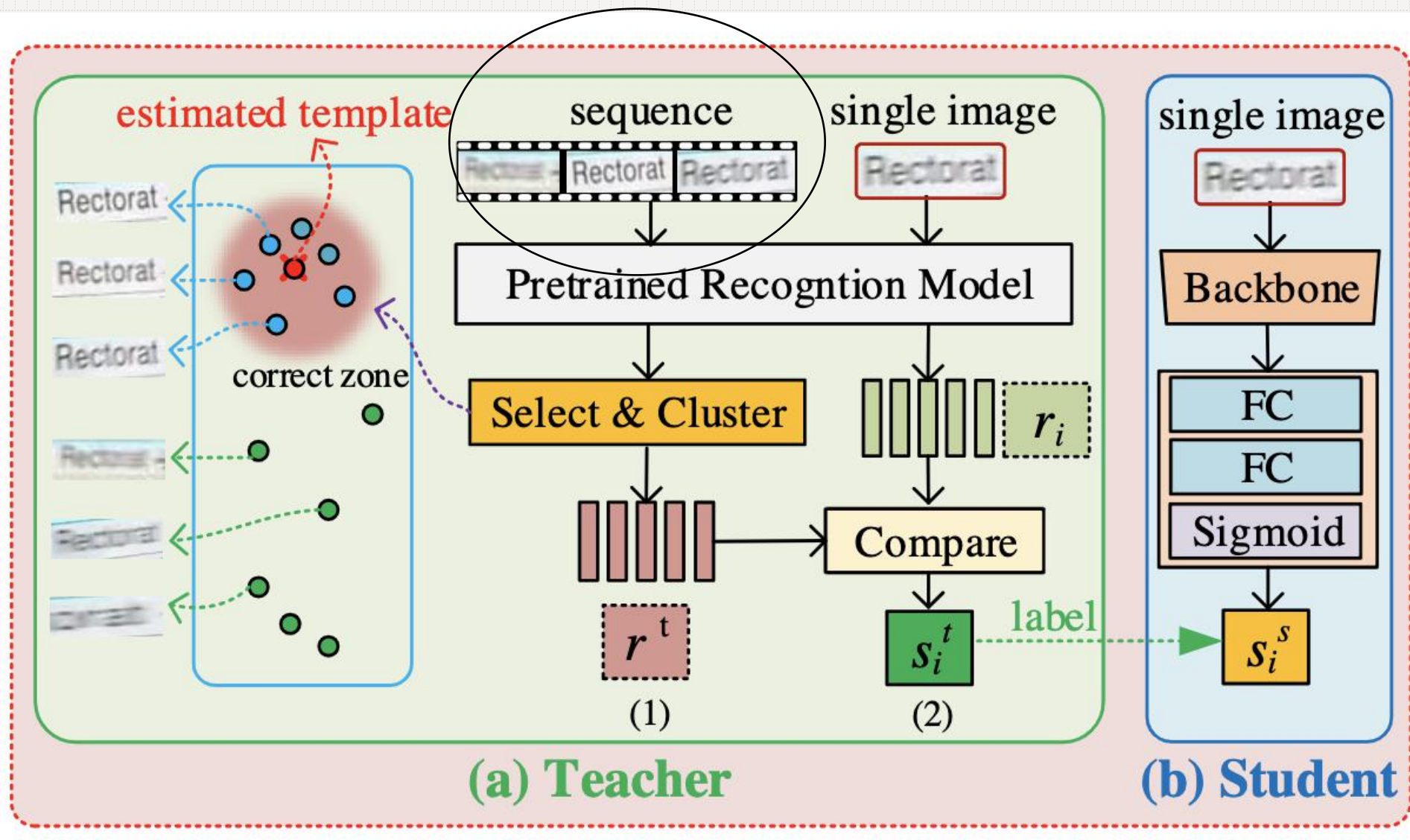
(e)



(f)

2

Discriminator





感谢聆听!

THANK YOU FOR WATCHING!

