

方案说明书

(1) 比赛介绍+赛题重点难点剖析；

比赛介绍：本项目是参加飞桨常规赛：中文场景文字识别的项目，项目任务为识别包含中文文字的街景图片，准确识别图片中的文字。

赛题重点难点剖析：中文场景中的文字面临着包括光照变化、低分辨率、字体以及排布多样性、中文字符种类多等复杂情况。如何解决上述问题成为一项极具挑战性的任务。此外，从自然场景图片中进行文字识别，需要包括 2 个步骤：

- 1) 文字检测：解决的问题是哪里有文字，文字的范围有多少？
- 2) 文字识别：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。

(2) 思路介绍+方案亮点；

思路介绍：第一步是针对中文场景下的数据预处理（包括：把繁体字转成简体字，大写->小写，删除空格，删除符号等操作），结合相应的中文字典来提升文字识别的准确率。第二步是在飞桨框架下采用当前业界最经典的 CRNN 算法架构来建模与求解，以保证模型的性能。

方案亮点：结合中文场景下的字典资源来完成数据的预处理，可以更好的构建训练模型的语料；参考了 Tensorflow 版本的 CRNN 模型的代码，然后将其改造并迁移到飞桨平台，使其与国产化平台相适配；采用 CRNN 架构下的 rec_r34_vd_none_bilstm_ctc 模型来建模与求解。

(3) 具体方案分享（思路框架图、思路步骤详述、模型应用+调优过程、部分代码展示）



图 1 文字识别的流程图

上图展示了整个识别过程的流程，CRNN 模型的框架和相应超参数设置如下所示：

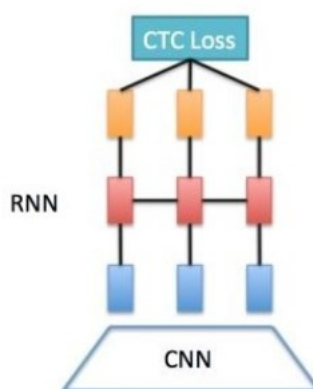


图 2：RCNN 模型的网络层次结构图

Global:

algorithm: CRNN

```

use_gpu: true
epoch_num: 30
log_smooth_window: 20
print_batch_step: 100
save_model_dir: output/rec_CRNN_aug_341
save_epoch_step: 1
eval_batch_step: 1800
train_batch_size_per_card: 256
test_batch_size_per_card: 128
image_shape: [3, 32, 256]
max_text_length: 64
character_type: ch
loss_type: ctc
reader_yaml: ./configs/rec/rec_icdar15_reader.yml
pretrain_weights: /home/aistudio/work/PaddleOCR/model/latest
checkpoints:
save_inference_dir: /home/aistudio/work/test
character_dict_path: /home/aistudio/work/dict.txt
Architecture:
  function: ppocr.modeling.architectures.rec_model,RecModel
Backbone:
  function: ppocr.modeling.backbones.rec_resnet_vd,ResNet
  layers: 34
Head:
  function: ppocr.modeling.heads.rec_ctc_head,CTCPredict
  encoder_type: rnn
SeqRNN:
  hidden_size: 256
Loss:
  function: ppocr.modeling.losses.rec_ctc_loss,CTCLoss
Optimizer:
  function: ppocr.optimizer,AdamDecay
  base_lr: 0.00001
  beta1: 0.9
  beta2: 0.999

```

(4) 模型应用结果分析

不同模型结果的对比分析：

模型名称	score	norm_distance	word_acc
官方基线模型 (PaddleOCR：中文场景文字识别)	82.87	0.93946	0.82872
CRNN rec_r34_vd_none_bilstm_ctc	85.94	0.95722	0.85941

调参优化过程分析：

- 1.将 epoch_num 改为 30
- 2.将 train_batch_size_per_card 改为 256
- 3.将 test_batch_size_per_card 改为 128
- 4.将 base_lr 改为 0.00001

经测试这样能提高 score

(5) 总结+改进完善方向

总结：通过参加本次比赛，大大的拓宽了自己的眼界，对模型有更加深刻的认识，可以根据不同的应用场景，去阅读国内外最新的文献，并将相关算法进行改造用以解决实际问题。在中文场景文字识别任务上，可以采用本模型来解决相关实际问题。

改进完善方向：

第一，CNN 部分目前用的 RESNET，后续可以考虑改成 VGG 网络；

第二，可以进一步加大高质量的标注数据集来训练模型，以增强模型的泛化性能；

第三，后续可以进一步优化损失函数和训练策略，以便提升模型的收敛速度。

(6) 飞桨使用体验+给其他选手学习飞桨的建议

飞桨使用体验：

飞桨作为国内第一个自主研发的深度学习框架，虽然暂时还不能完全比肩 Tensorflow 或者 PyTorch 这样的成熟框架，但是经过努力，现在的飞桨 2.0 版本已经可以大幅满足国产化适配的要求，很多场景都有公开的算法支持和 API 说明文档，还有免费的 GPU 资源可以使用，所以整体使用体验还是非常赞的，后面会逐步响应国家的号召，慢慢从其它框架转向飞桨框架，支持国产化发展。

给其他选手学习飞桨的建议：

鼓励国内的研发人员和科研人员积极响应国家号召，投身到自主可控技术的研发当中来，积极贡献自己的力量，为国产化发展打好基础，有时间的用户，可以多参加飞桨平台组织的比赛。此外，建议百度公司可以组织飞桨平台的核心研发人员撰写并出版《飞桨框架的相关权威指南》来指导用户的使用。有了这类教材之后，以后国内的高校也可以开设相关的课程，从理论方面开始引导与培养，这样可以更好的普及与推广飞桨平台。

(7) 参考资料（论文、课程、其他资料）

1. EAST:

```
@inproceedings{zhou2017east,
  title={EAST: an efficient and accurate scene text detector},
  author={Zhou, Xinyu and Yao, Cong and Wen, He and Wang, Yuzhi and Zhou, Shuchang and He, Weiran and Liang, Jiajun},
  booktitle={Proceedings of the IEEE conference on Computer Vision and Pattern Recognition},
  pages={5551--5560},
  year={2017}
}
```

2. DB:

```
@article{liao2019real,  
  title={Real-time Scene Text Detection with Differentiable Binarization},  
  author={Liao, Minghui and Wan, Zhaoyi and Yao, Cong and Chen, Kai and Bai,  
Xiang},  
  journal={arXiv preprint arXiv:1911.08947},  
  year={2019}  
}
```

3. DTRB:

```
@inproceedings{baek2019wrong,  
  title={What is wrong with scene text recognition model comparisons? dataset and  
model analysis},  
  author={Baek, Jeonghun and Kim, Geewook and Lee, Junyeop and Park, Sungrae  
and Han, Dongyoon and Yun, Sangdoo and Oh, Seong Joon and Lee, Hwalsuk},  
  booktitle={Proceedings of the IEEE International Conference on Computer Vision},  
  pages={4715--4723},  
  year={2019}  
}
```

4. SAST:

```
@inproceedings{wang2019single,  
  title={A Single-Shot Arbitrarily-Shaped Text Detector based on Context Attended  
Multi-Task Learning},  
  author={Wang, Pengfei and Zhang, Chengquan and Qi, Fei and Huang, Zuming  
and En, Mengyi and Han, Junyu and Liu, Jingtuo and Ding, Errui and Shi, Guangming},  
  booktitle={Proceedings of the 27th ACM International Conference on Multimedia},  
  pages={1277--1285},  
  year={2019}  
}
```

5. SRN:

```
@article{yu2020towards,  
  title={Towards Accurate Scene Text Recognition with Semantic Reasoning  
Networks},  
  author={Yu, Deli and Li, Xuan and Zhang, Chengquan and Han, Junyu and Liu,  
Jingtuo and Ding, Errui},  
  journal={arXiv preprint arXiv:2003.12294},  
  year={2020}  
}
```

6. end2end-psl:

```
@inproceedings{sun2019chinese,  
  title={Chinese Street View Text: Large-scale Chinese Text Reading with Partially
```

```
Supervised Learning},  
  author={Sun, Yipeng and Liu, Jiaming and Liu, Wei and Han, Junyu and Ding, Errui  
and Liu, Jingtuo},  
  booktitle={Proceedings of the IEEE International Conference on Computer Vision},  
  pages={9086--9095},  
  year={2019}  
}
```