

CaMiTextSpotter

Abstract

- 概述该研究的发展现状
- 该研究的发展瓶颈问题
  - 字符集过大带来的长尾效应（引出对字符集做拆分、并引入语种分类任务的必要性）
  - 多字符集的混合样本（由于字符集拆分，会导致无法解决的难题）
  - 针对端到端引入语种分类，模型层面多任务的耦合性，需要进一步优化
- 所提方法的名称以及形式化（多标签、字符级别的语种识别）
- 实验结果可比较

Introduction

- ...
- ...
- ...
- Contribution
  - 提出第一个基于transformer架构的多语种端到端OCR框架。将单标签分类转化成序列、多标签预测，并且提出两个约束项确保模型可以收敛。
  - 复用attention机制，使得细粒度的分类信息能够可信地作用于后处理。
  - 实验表明所提方法能够达到 SOTA 可比较的性能

Related Work

- Text detection and recognition methods
- Script identification
- Text spotting

Methodology

- Overview
  - 方法基于SwinTextSpotter，包含xxx组成构件
  - 端到端，多语种识别存在的问题
    - 如何在固定字符集的情况下，解决混合样本？-> 分类形式化转换成序列级、多标签识别，并且辅以后处理多头集成投票策略
    - 如何确保有效地端到端优化？-> 复用attention机制统一语种识别和OCR两个任务、RC（SwinTextSpotter提出）结构中引入 分类的 logits、序列attention信息。
  - 如何定义字符集？-> t-sne分析+多标签分析（归入消融实验里面作为一个小标题）
- Character-aware Multi-script Recognition（CMR）
  - 模块结构的概述
  - 标签生成
  - 为了解决混合样本，将分类形式化转为序列+多标签
- The usage of character-level, multi-script information
  - Character-aware recognition conversion（CA-RC）改进baseline RC结构，引入语种识别的细粒度信息
  - 复用attention确保后处理策略实现多头预测结果的可信集成，提出一种后处理字符替换策略 Script voting strategy
- Optimization
  - CMR模块的序列学习损失、以及两个约束项
  - SwinTextSpotter架构包含的检测、分割、识别损失

Experiments

- Datasets
- Implementation details
- Ablation study
  - Multi-label ambiguity
    - 通过t-sne分析通用OCR head字符之间的距离关系
    - 分析MLT的标签和多标签映射结果
  - Character-aware Multi-script Recognition
    - 两个约束项的消融
    - 和通用OCR的性能对比
    - 语种识别替换成CMR模块的性能对比
  - Compared with SOTA
    - MLT17、19数据上 Task1、2、3
    - 样张的可视化
  - 总结
    - 构建字符集的合理性
    - 字符树
    - 利用大规模合成数据，仅考虑分类+OCR的场景