

程昌旭

(+86) 159-2720-5178 · ccx0127@foxmail.com · AI 算法工程师

专业技能

- **专业领域:** 文字识别、文档理解、多模态建模、图像识别、语言模型、预训练技术
- **编程语言:** Python (熟练), C/C++

教育背景

华中科技大学, 信息与通信工程, 工学硕士 2018.9 - 2021.6

曾获得研究生知行优秀奖学金; ICDAR2019 ReCTS-中文字符识别赛道学生组第一名, 参与第五届“互联网+”大学生创新创业大赛获得主赛道金奖。

华中科技大学, 水利水电工程, 工学学士 2014.9 - 2018.6

获得过国家励志奖学金 (学院 5%)、华中科技大学“三好学生” (6%)、“优秀共青团员”等荣誉; 2017年美国大学生数学建模竞赛一等奖 (Meritorious Winner, 9%)、第7届全国大学生数学竞赛一等奖等。担任过华中科技大学院系资助助理、资助委员会宣传部副部长等学生干部。

工作经历

阿里巴巴达摩院 (杭州) 2021.7 至今

算法工程师, 负责图像文字识别与纠错技术、文档理解技术的研究与业务支持。

研究成果

文档理解

(共一) GeoLayoutLM: Geometric Pre-training for Visual Information Extraction. CVPR'23 Highlight (Top 2.5%)

- 针对视觉富文档信息抽取问题, 提出了一种几何知识预训练方法, 可以显著提升模型的结构化文本关系抽取能力, 同时也具有良好的实体识别能力。

文字识别

(一作) LISTER: Neighbor Decoding for Length-Insensitive Scene Text Recognition. ICCV'23

- 一种基于新的注意力机制的文字识别模型, 既拥有 CTC 解码器对文本长度不敏感的优点, 也具有传统注意力机制特征学习更好的优点, 且模型计算量小。

(一作) Decoupling Visual-Semantic Feature Learning for Robust Scene Text Recognition. arXiv'21

- 证明额外的语料可以缓解图像文本识别器在 OOV 词汇上识别效果不好的问题。

(一作) Maximum Entropy Regularization and Chinese Text Recognition. DAS'20 oral

(一作) Patch Aggregator for Scene Text Script Identification. ICDAR'19 oral

专利成果

(第二发明人) 一种结合全局和局部信息的场景文本图像的语种识别方法。

项目经历

智能文档处理项目 2023.06-至今

该项目的初衷的实现办公自动化。在大模型时代, 也为大模型提供了来自海量文档的训练数据。我主要负责文档结构化部分, 预测文档内文本块的精细类别, 并提取文档内部层级结构树。具体来说, 对每一个文档 (多页), 对每个版面框做语义分类, 并预测出其父结点。这个任务的主要难点在于, 如何去处理任意长度的文档, 包括成百上千页的情况, 保证速度和准确度都合适。我采用 GeoLayoutLM 为 backbone 进行, 对文档进行分块和融合处理, 高并行地建模全局依赖关系。相比于基于 BERT 的旧有模型, 在速度和精度上都取得了十分显著的提升。

自学习平台的文档 AI 能力建设 2022.11-2023.1

将文档理解研究成果 (GeoLayoutLM) 在实际业务使用。训练支持中英文的模型, 在 modelscope 开源, 并在内部自学习平台业务中提供给业务同学使用。

某统计局的文档 AI 项目 2022.10-2023.4

负责结构化规则引擎维护。

读光 OCR 系统的搜索纠错功能建设	2021.10-2022.3
基于知识库，对 OCR 结果中的若干字段进行纠错。主要包括搜索（粗排）、多字段联合校验、精排、纠错决策等过程。	
读光 OCR 公有云的卡证票据识别业务	2021.7-2023.4
负责升级和维护结构化规则引擎，参与建设的卡证票据场景超过 50 个。此外，针对生僻字识别困难的问题，提出了一种对抗学习模型，在营业执照识别场景上线。	