

关于乐谱检测-识别项目的阶段报告

算法组：郭辉铭

2023-2-15

一：需求

1. 前端部署针对用户拍摄以及导入图片的曲谱标题识别
2. 针对后端曲谱以及用户拍摄的所有曲谱的前端部署曲谱术语识别
3. 针对后端曲谱以及用户拍摄的所有曲谱的前端部署曲谱作曲家，编曲家识别

二：需求分析

1) 标题识别

1. 首先需要对曲谱进行检测。将曲谱中的中文以及外文文字信息进行准确框选，且降低假阳率
2. 曲谱标题识别需要针对谱集的首页标题进行识别。包括：中文，英语，法语，德语，俄语。
3. 识别中包含多种复杂情况（标题位置影响，标题间隔，标题分级，以及广告信息和作词作曲家，数字谱。吉他谱的影响）以及标题分级情况。
4. 部署使用c++，ios和安卓依赖于不同的三方库。以及内存管理及其优化

2) 术语检测

1. 检测-识别-模型量化-蒸馏-裁剪-准确率测试-提交上级认可-c++部署（标题识别已部署，该部分通用，只需要做少许修改）。流程同标题识别

2. 标题的检测没有太多的符号干扰，因此标题的检测准确率在99%以上（测试过近1000张包含电子谱以及不同型号手机拍摄的曲谱图片）。

而术语则被音符，谱线以及术语符号和噪声干扰，因此在检测，尤其是识别上，要求更高的准确率。

3) 作曲家识别, 同1,2

三) : 现阶段进展 (2023.2.15)

以下测试是基于大量的不同乐器的曲谱, 不同手机型号拍摄的曲谱, 不同格式的曲谱进行的。数据可靠

1) 曲谱检测

A: 模型大小7M。检测准确率93%

B: 模型大小102M。检测准确率96%

C: 102M蒸馏量化后大小5.6M。检测准确率91%

目前: 前端使用A模型。后端数据库在线下使用B模型

2) 文字识别

A: 中文官方模型大小10.8M。识别的完全准确率96.48%; 编辑距离匹配率: 98.78%。(官方模型数据充足, 且中文的字体干扰较小。重复性较低)

B: 自训练外文模型大小2.1M。识别准确率84.89%; 完全匹配率97.30%。(训练样本较少, 字母较少, 且相似度较高的字母占比不低, 字母比中文更容易受噪声点和无关符号的干扰)

目前以上两个模型均应用于前端

四) : 维护(根据需求而来, 或者利用空余时间加班尝试)

1.识别的准确率有待提升。尤其是一些特殊术语的识别: mf,sfp,fp,tr等。字体太花, 以及清晰度的干扰导致识别准确率很低

2.对于外文识别模型的数据量需要加大

3.主干网络的替换和设计。当前外文模型只有2.1m。