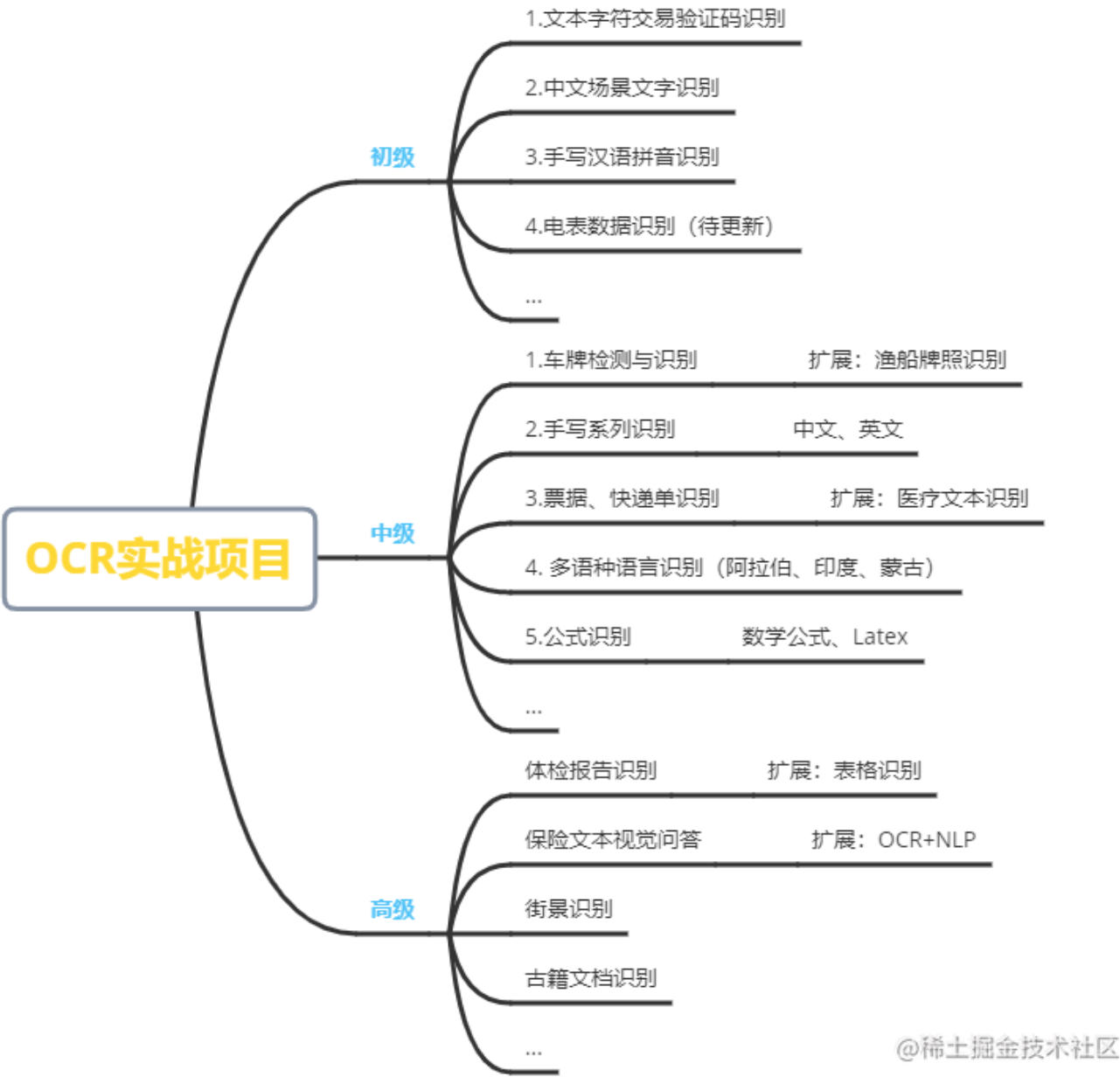


#项目



## 基于 CAGNN 的端到端识别

本项目以 DCID2022 赛题为背景，以已标记字符信息的实例字符验证码图像数据为训练样本，参赛选手需基于提供的样本构建模型，对测试集中的字符验证码图像进行识别，提取有效的字符信息。训练数据集不局限于提供的数据，可以加入公开的数据集。

### 1 数据与评测

#### 1.1 数据简介

提供标注信息训练数据集 15000 张，每张训练数据由一个 4 位文本字符验证码图像组成，字符随机产生，标注集对当前图像中的文本字符进行内容标注；测试数据集含 25000 张验证码图像。



1.2 数据说明

本项目提供训练数据集文件 train\_imgs.zip，其中文件名称对应图片文本字符标签；测试数据集文件 test\_imgs.zip，包含待识别的图像文件。文件名称 说明 train\_imgs.zip 训练集图片，15000 张验证码图片 test\_imgs.zip 测试集图片，25000 张待识别验证码图片 submit\_example.csv 最终数据格式提交样例 done

基于 PGNet 的端到端识别

框架介绍与环境安装

PaddleOCR 框架

