



Trap x00

Tag x00 展示

Trap x00

陈俊达 张凌哲 梁楠 张美玲

2018年6月21日





着力点

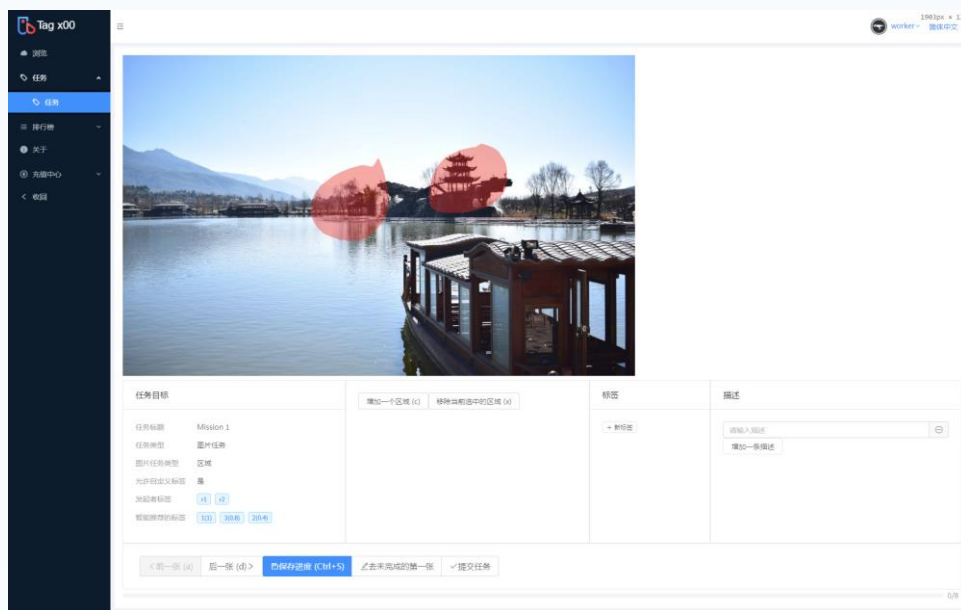
- 标注工作本身
- 让工人爽
- 兼具广度和深度

特色

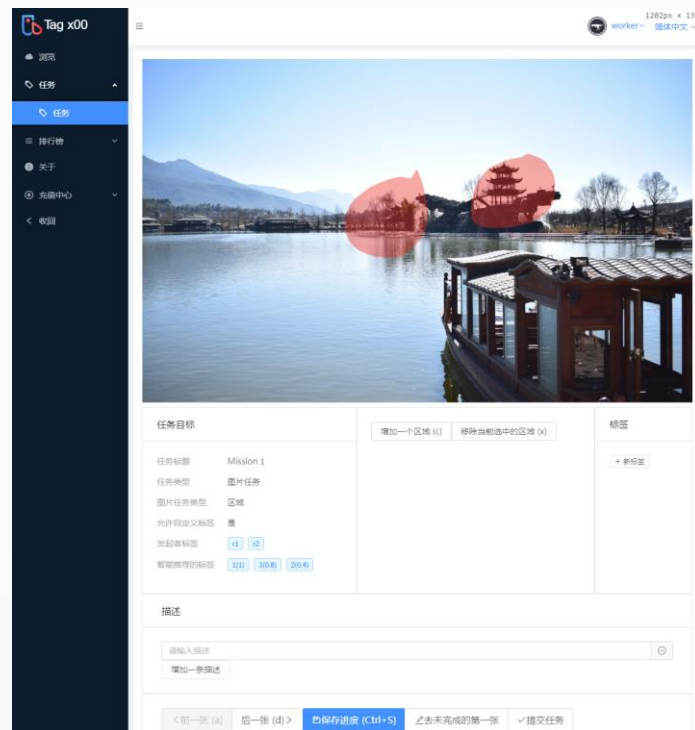
- 图片、文本、音频、视频、3D标注
- 用户体验
 - 精心设计的UI，完全的响应式设计（支持各类设备）
 - 快捷键
 - 国际化
- 智能模块：
 - 在现有**API**基础上的，基于图片物体识别的标签推荐
 - 文本分词
- 丰富的、可交互的数据可视化

响应式

- 完全的响应式，所有功能完美支持各种屏幕大小，支持鼠标和触摸

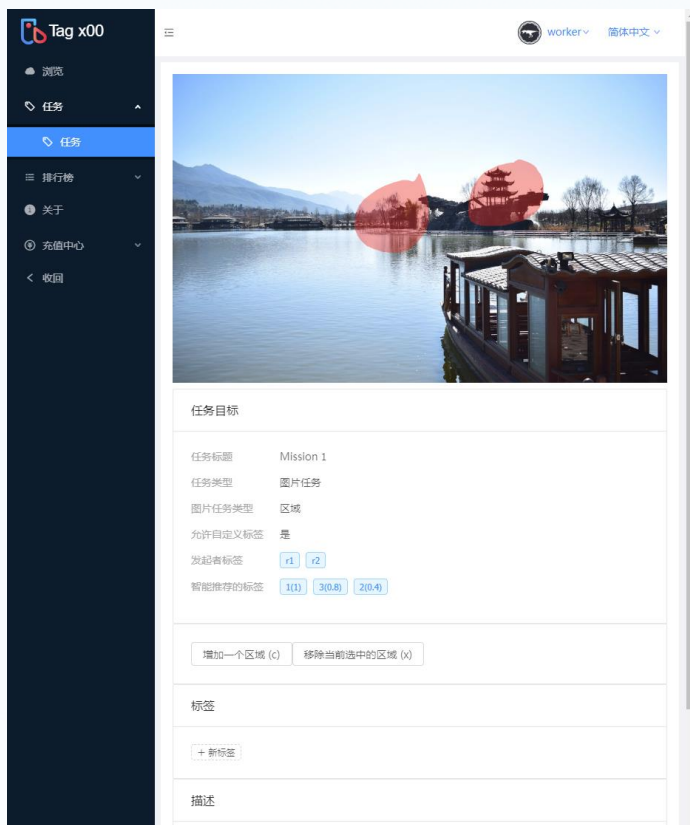


宽度1903px，鼠标，画下两个区域
图片大小960*600
缩放比例1.04（1000宽度上限阈值）

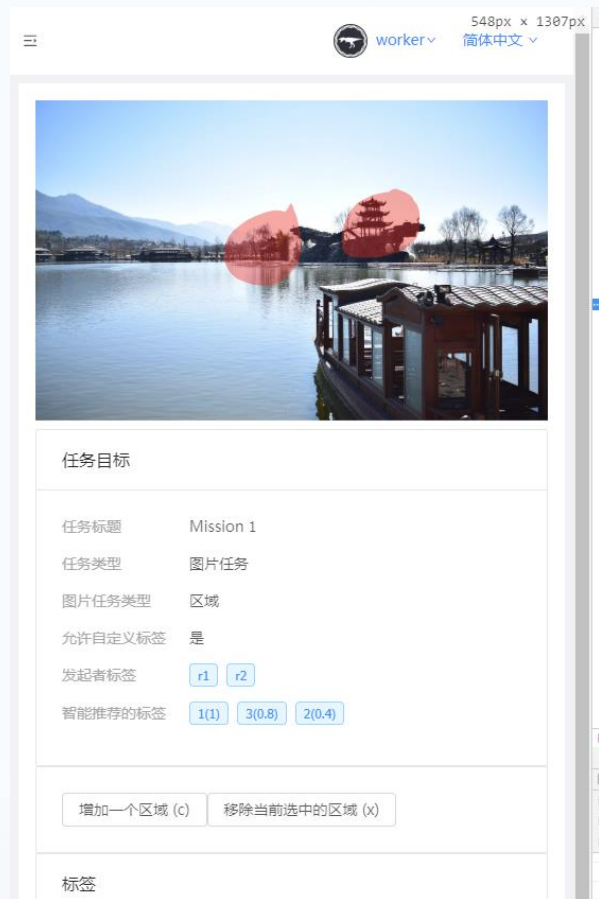


宽度1202px，鼠标，图片缩放比例0.97

响应式



宽度934px，鼠标
图片缩放比例0.69

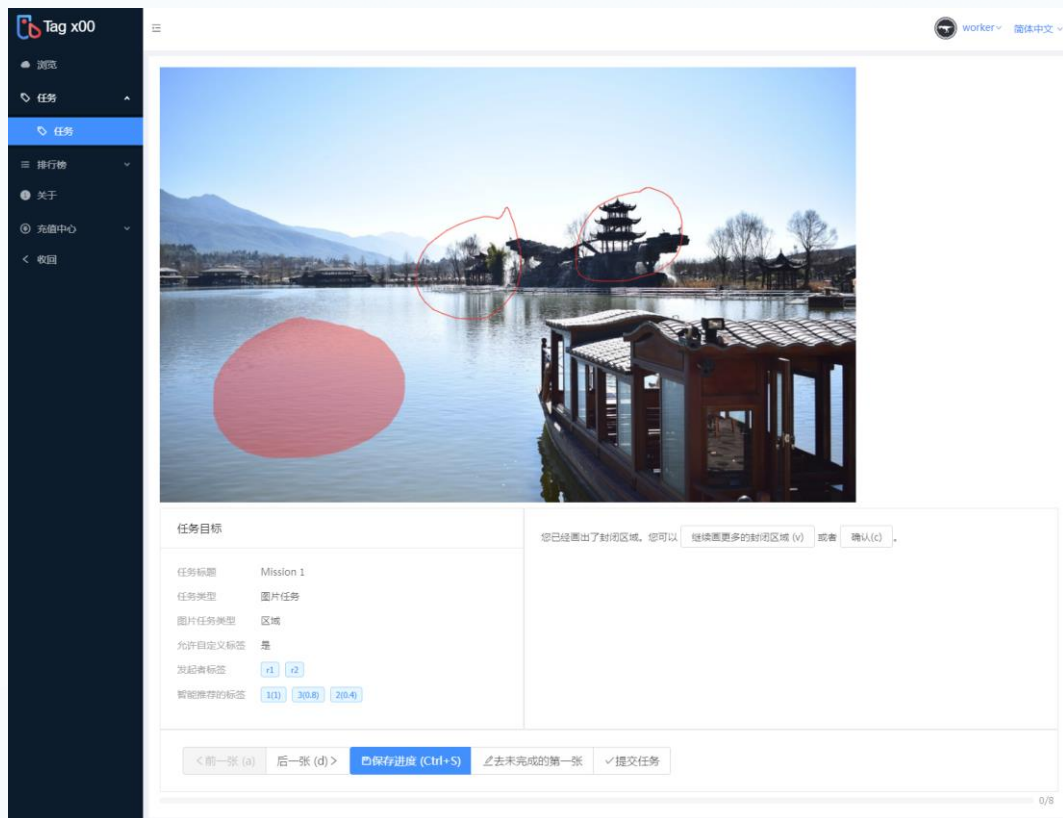


宽度548px，鼠标
图片缩放比例0.5

响应式



宽度667px，触摸
缩放比例0.644
正确检测到触摸位置



宽度1539px，触摸
缩放比例1.04
区域位置正确

快捷键



任务目标

您已经画出了封闭区域。您可以 或者 。

任务标题	Mission 1
任务类型	图片任务
图片任务类型	区域
允许自定义标签	是
发起者标签	<input type="button" value="r1"/> <input type="button" value="r2"/>
智能推荐的标签	<input type="button" value="1(1)"/> <input type="button" value="3(0.8)"/> <input type="button" value="2(0.4)"/>

< 前一张 (a) 后一张 (d) >

保存进度，前后，确认区域
(Ctrl+s, a, d, c)

任务目标	标签	描述
任务标题	Mission 1	
任务类型	音频任务	
音频任务类型	整体	
允许自定义标签	否	
预设标签	<input type="button" value="1"/> <input type="button" value="2"/>	

+ 删除标签 增加一条描述

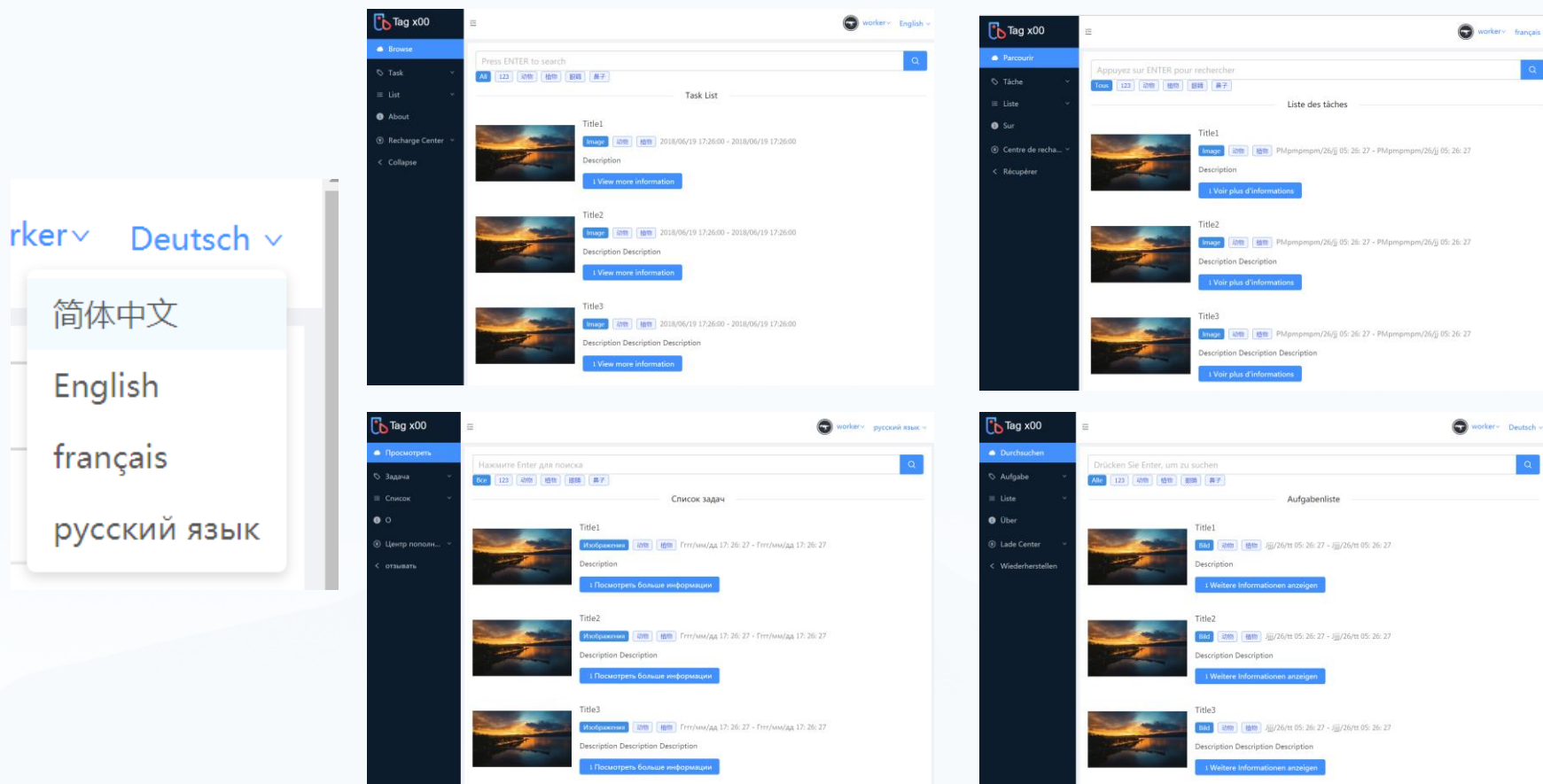
< 前一张 (a) 后一张 (d) >

0/4

空格播放或者暂停

国际化

- 简体中文，英文，法语，德语，俄语，热切换，按需加载语言文件



图片标注

- 区域标注，支持非连续的多个区域成为一个标注单位



任务目标	
任务标题	Mission 1
任务类型	图片任务
图片任务类型	区域
允许自定义标签	是
发起者标签	<input type="text" value="r1"/> <input type="text" value="r2"/>
智能推荐的标签	<input type="text" value="1(1)"/> <input type="text" value="3(0.8)"/> <input type="text" value="2(0.4)"/>
<div>增加一个区域 (c) 移除当前选中的区域 (x)</div>	
标签	
<div>+ 新标签</div>	
编辑	

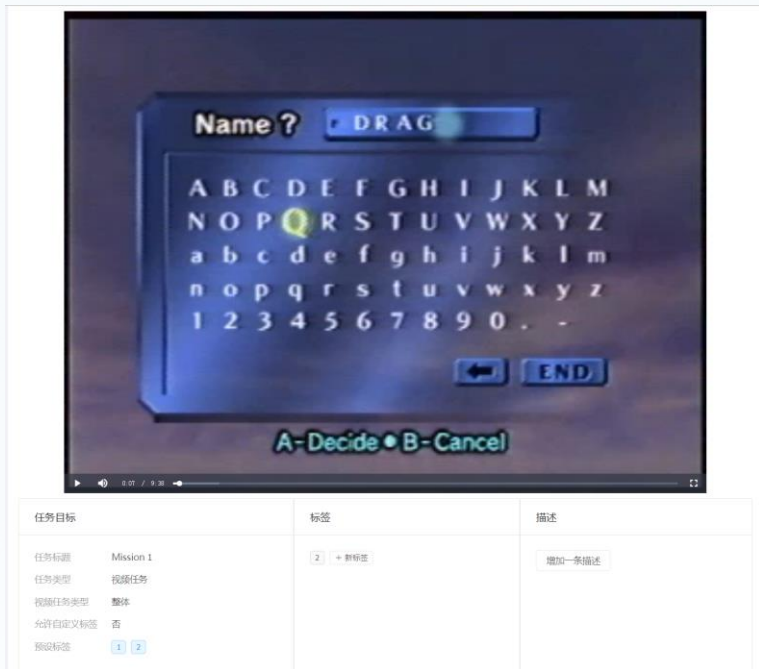
- 提取主题词（可自定义）， 文本分类（不可自定义）
- 机器分词辅助

[illegible]

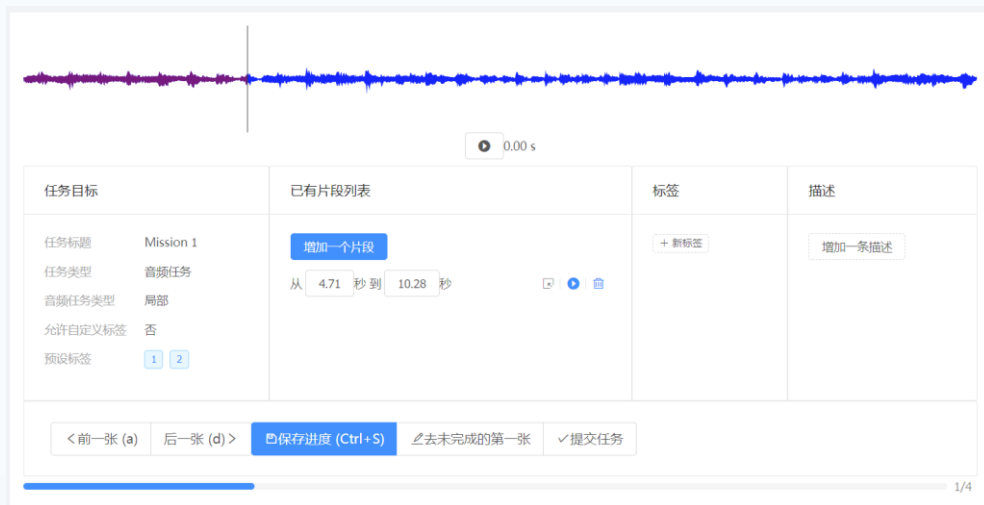
提取关键词
机器分词
点击即成为标签

音频、视频标注

- 整体，片段



视频整体

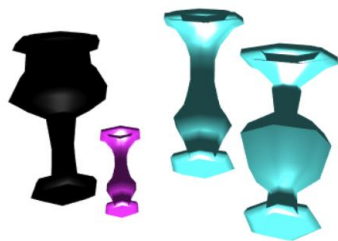


音频片段

点击秒数按钮设定时间

点击播放按钮播放此片段

3D标注



Mtl, obj格式3D模型
可自由移动相机查看3D模型

任务目标	标签	描述
<div>任务标题Mission 1</div> <div>任务类型3D任务</div> <div>3D任务类型整体</div> <div>允许自定义标签否</div> <div>预设标签12</div>	<div>+ 新标签</div>	<div>增加一条描述</div>

< 前一张 (a)

后一张 (d) >

保存进度 (Ctrl+S)

去未完成的第一张

提交任务

标签推荐

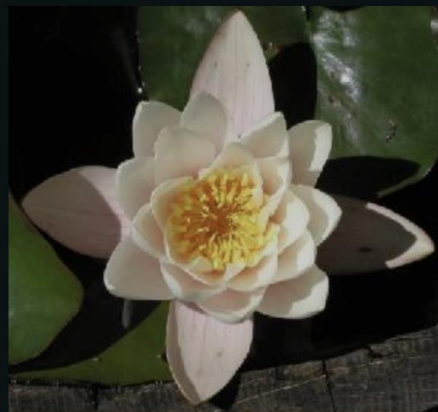
• 现有API的问题（阿里云）



图像标签：

发卡	32%
煎锅	16%
开瓶器	16%
小刀	13%
电熨斗	13%

识别错误



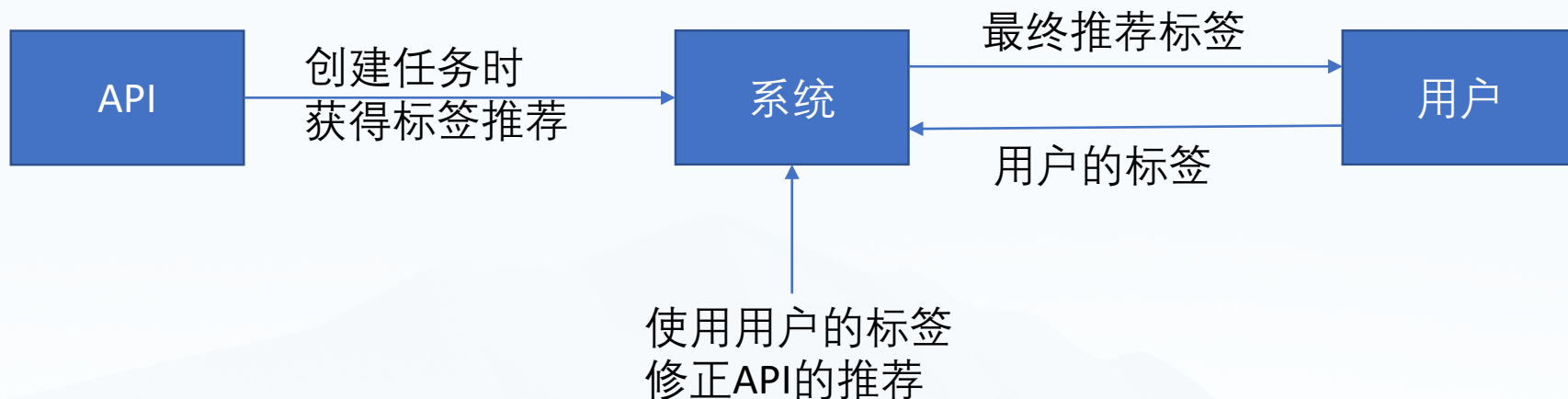
图像标签：

花	43%
餐巾	23%
植物	17%
其他	13%
雏菊	12%

识别不准确

标签推荐

- 基本思路：
 - 在现有API的基础上，通过工人的数据修正API给出的数据
 - 迁移学习



效果

- 使用我们自己做出来的110份标注结果、40张图片进行学习
- 使用10张图片（出现在训练集中），使用如下公式度量算法效果（结果称为采纳率）

$$\phi = \frac{\sum_{j=1}^m \sum_{i=1}^n \hat{y}_{ij} * y_{ij}}{m}$$

以上 X_{ij} 为输入提示标签, Y_i 为输入标注答案, m 为批度大小, \hat{y}_i 为预测值, y_i 为标签值, y_i 为单组标签的总预测值 $\sum_{i=1}^n \hat{y}_{ij}$

算法

- 使用单隐层神经网络，使用交叉熵来作为神经网络的代价函数，批量梯度下降来学习参数

- 首先为了提高系统的鲁棒性，我们在对数据进行计算前加入高斯白噪声 $X = X + e$
- 接下来使用 $O = g(W_{hidden}X + B_{hidden})$ 计算隐层输出,这里使用隐层神经元128个，激活函数 g 使用relu函数，并为了防止过拟合，训练时在此使用dropout=0.7
- 然后直接连接输出层 $\hat{y} = \theta(W_{output}O + B_{output})$ 得到输出值， θ 函数对结果进行归一化，使用softmax进行归一化的处理，并再计算一次 $y_i = \frac{y_i}{\sum y}$
- 然后根据已有的数据，计算 \hat{y} 的拒绝域。
首先根据不被接受的提示的 \hat{y} 进行降序排列,即对于一个结果数组P，如果 $y_{ij} = 0$ 则将 \hat{y} 加入P中，然后对P进行降序排列，取第 $\alpha * N$ 项，其中 α 为显著值， N 为P数组长度。

过滤

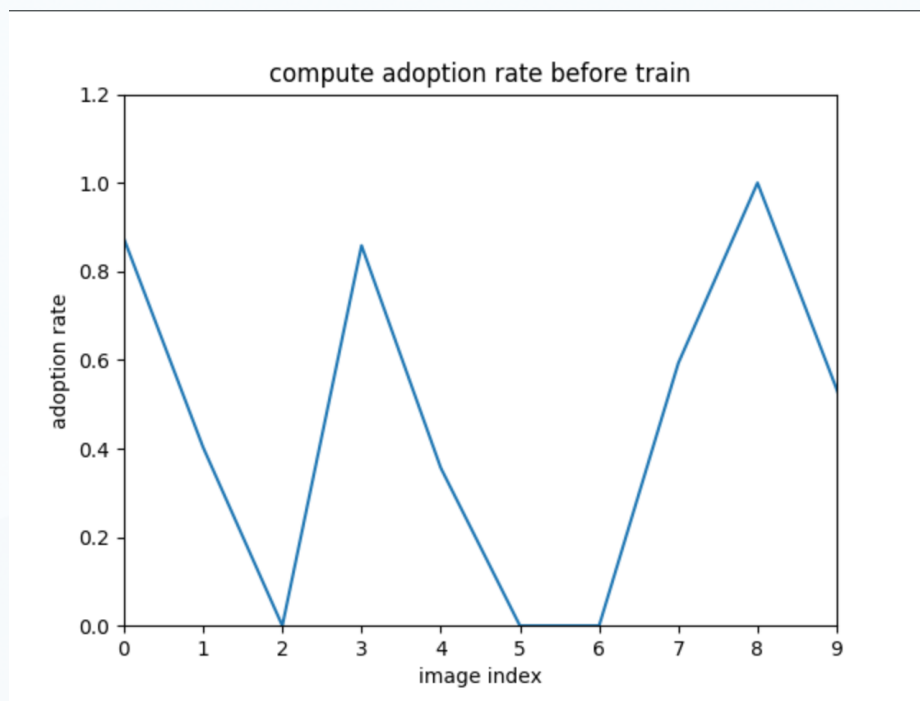
- 通过对数据的研究，我们发现，有一些数据其估计值非常低，所以我们认为，如果全组数据的估计值都极低，我们应该认为API无法识别此张图片

然后根据已有的数据，计算 \hat{y} 的拒绝域。

首先根据不被接受的提示的 \hat{y} 进行降序排列,即对于一个结果数组P，如果 $y_{ij} = 0$ 则将 \hat{y} 加入P中，然后对P进行降序排列，取第 $\alpha * N$ 项，其中 α 为显著值， N 为P数组长度。

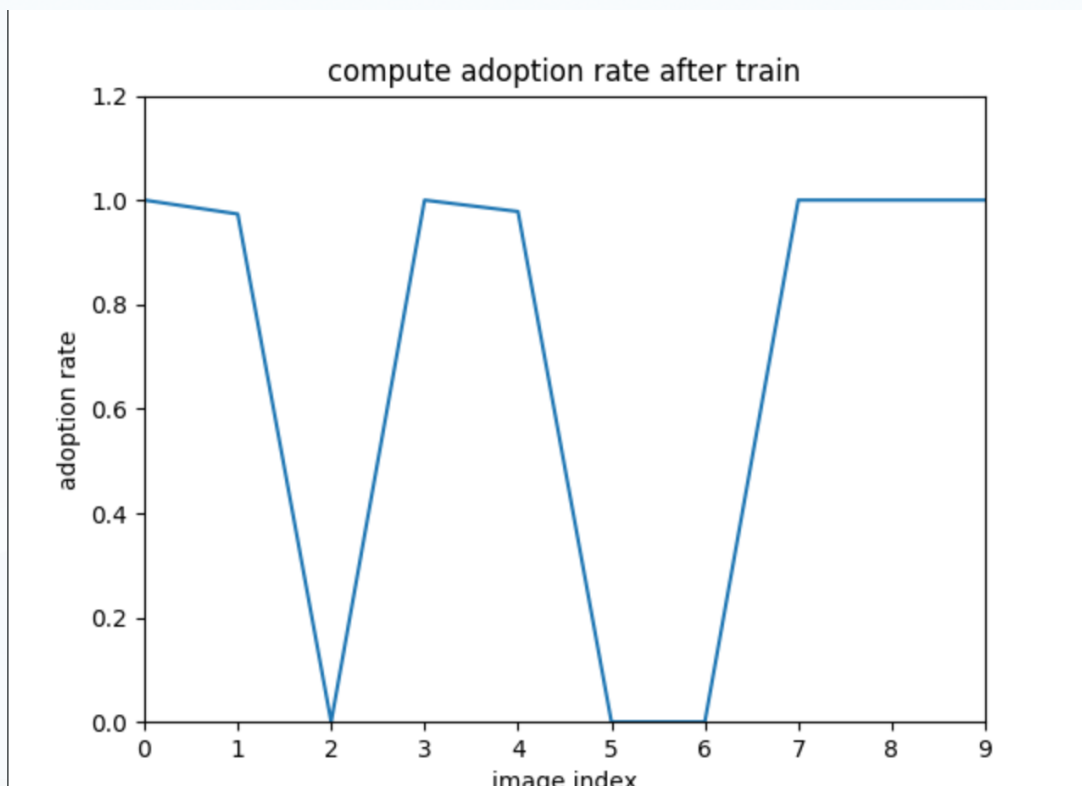
过滤

- 仅采用API，每张图片得到的采纳率如下图所示
- 均值为0.462



效果

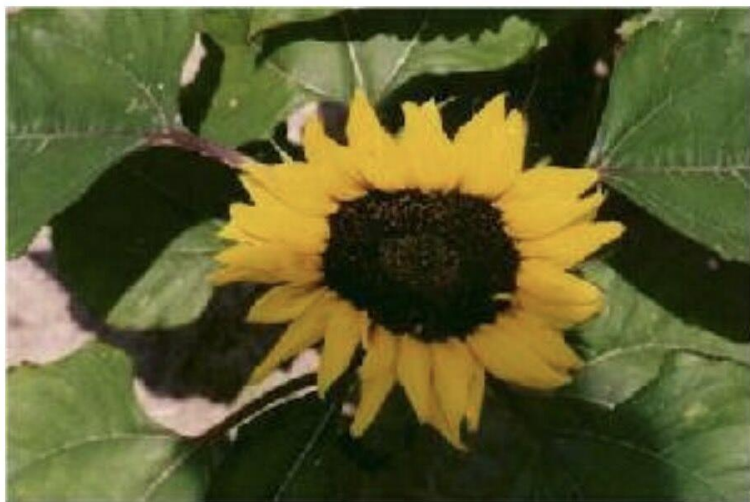
- 训练后，每张图的采纳率为如下图所示
- 均值为0.677



实例

- 做13张图，其中**不包含**以下这幅图。

系统对这幅图的推荐标签以及概率



第一次（纯API数据）

雏菊(0.35000000000000003) 植物(0.16) 花盆(0.13) 叶甲虫(0.13) 小青南瓜(0.11)

做13张图之后，第二次

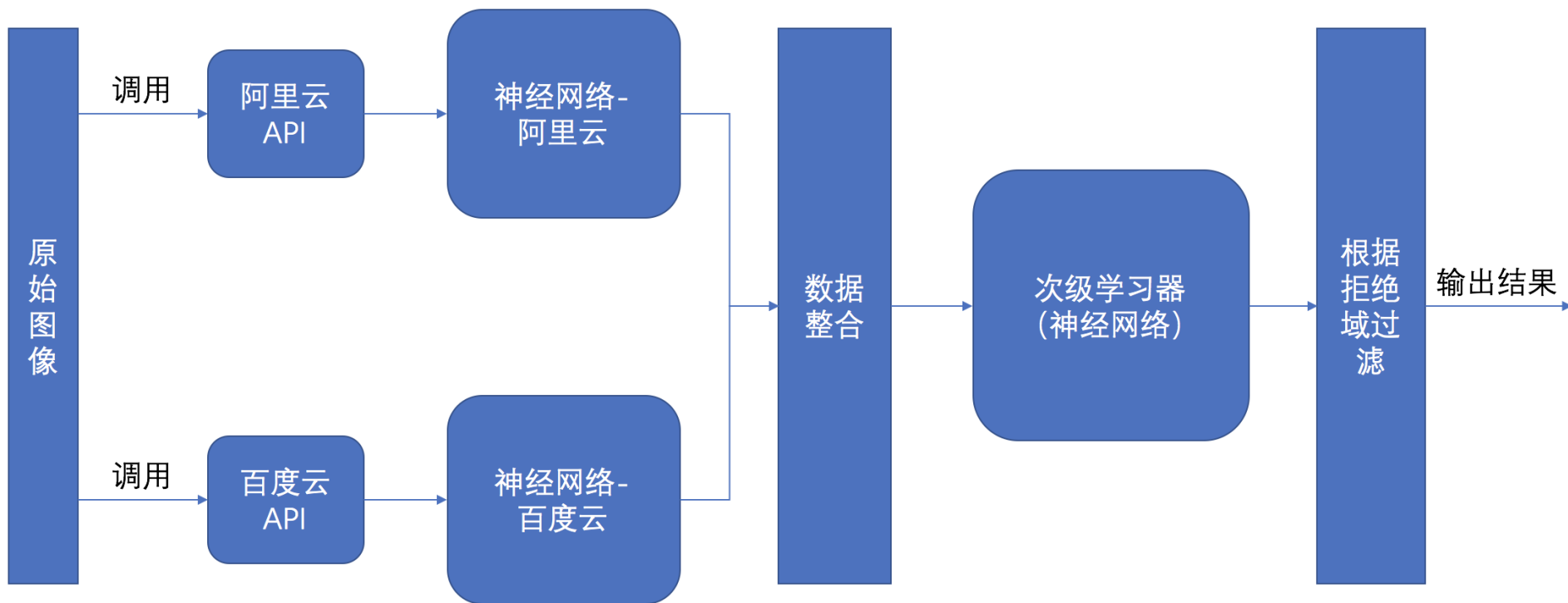
雏菊(1) 小青南瓜(0.99) 植物(0) 花盆(0) 叶甲虫(0)

解释和后续

- 研究数据发现index为2、5、6的数据都为0，都是因为其最终的提示没有被用户采用
- 初步想法是是因为这样的完全无用提示分布较为稀疏，数据量太少，神经网络层数太少，这样的数据无法被很好地识别出来
- 对于这样的情况，我们认为，执着于挑选出全无效提示对于整个系统没有太大的用处，所以我们接下来使用集成学习进行对提示进行改进.

改进

- 使用阿里云和百度云两种API，在这里，我们基于之前的模型使用Stacking算法融合两种api的结果



实例

- 做8张图，其中**包含**以下这幅图。



系统对这幅图的推荐标签以及概率

第一次（纯阿里云API数据）

小刀(0.68)

剑（刀）鞘(0.18)

钩子(0.17)

饰品(0.15)

长柄勺(0.12)

当时手动打入剪刀

理发剪(1)

剪刀(0.74)

小刀(0.66)

剑（刀）鞘(0)

钩子(0)

眉毛剪(0)

实例

- 仅仅在以上的基础上，做新的任务。

系统对这幅图的推荐标签以及概率



纯阿里云API数据

聚光灯(0.49)

磁罗盘(0.16)

太阳镜(0.13)

放大镜(0.13)

头盔(0.13)

加入刚刚已经训练好的模型

头戴式耳机(1)

头戴耳机(0.54)

聚光灯(0.4)

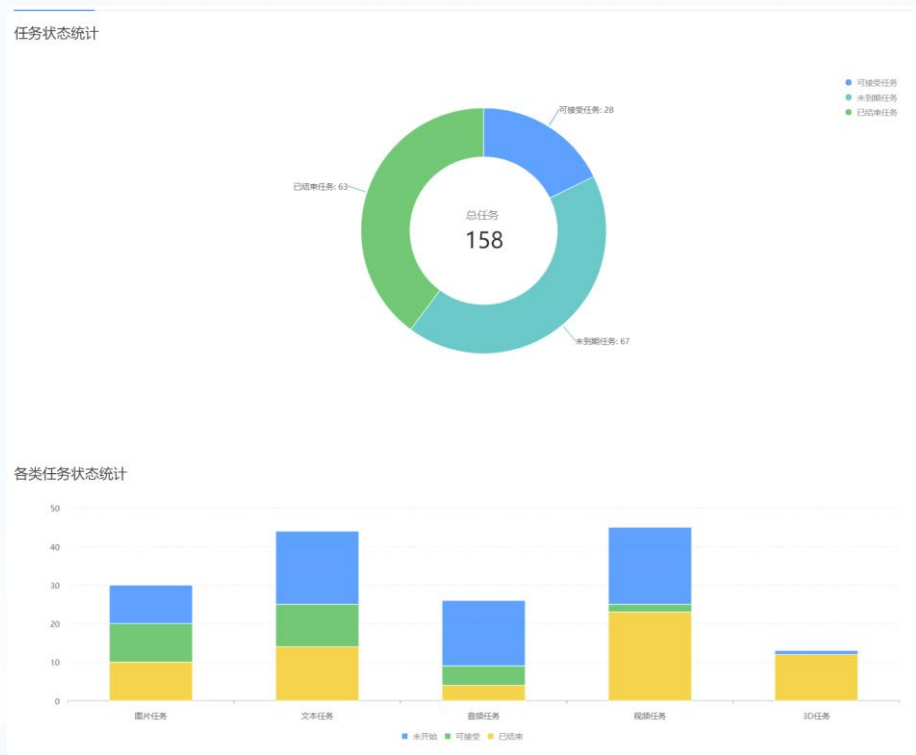
磁罗盘(0)

太阳镜(0)

耳机(0)

数据可视化

任务统计，管理员可以管理系统任务的状态



任务统计

可接受任务

任务ID	任务类型	任务状态	操作
mission	图片任务	可接受	任务详细
mission	图片任务	可接受	任务详细
mission	图片任务	可接受	任务详细
mission	图片任务	可接受	任务详细
mission	图片任务	可接受	任务详细
mission	图片任务	可接受	任务详细
mission	图片任务	可接受	任务详细
mission	图片任务	可接受	任务详细
mission	图片任务	可接受	任务详细
mission	图片任务	可接受	任务详细

<

1

2

3

>

Cancel

OK

点击色块显示色块对应的任务信息

数据可视化

用户统计，帮助管理员管理系统用户

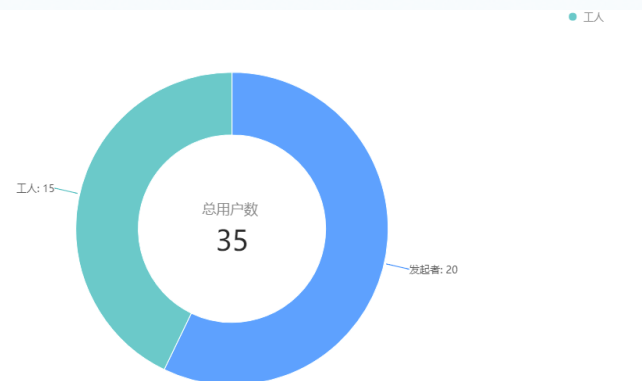
发起者

用户名	电邮	注册时间	创建任务数	任务实例数
requester undefined	smallda@outlook.com	2018-6-2	20	40
requester undefined	smallda@outlook.com	2018-6-2	20	40
requester undefined	smallda@outlook.com	2018-6-2	20	40
requester undefined	smallda@outlook.com	2018-6-2	20	40
requester undefined	smallda@outlook.com	2018-6-2	20	40
requester undefined	smallda@outlook.com	2018-6-2	20	40
requester undefined	smallda@outlook.com	2018-6-2	20	40
requester undefined	smallda@outlook.com	2018-6-2	20	40
requester undefined	smallda@outlook.com	2018-6-2	20	40
requester undefined	smallda@outlook.com	2018-6-2	20	40

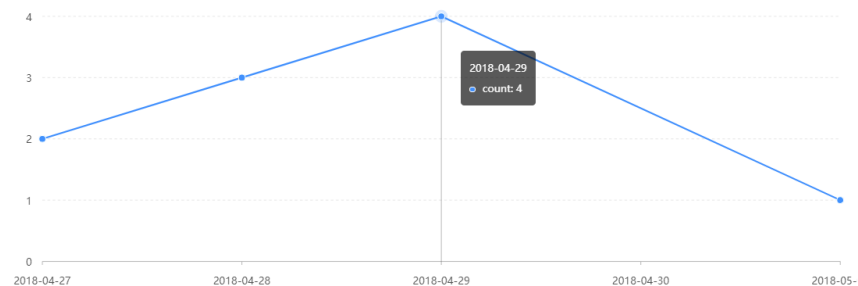
工人

用户名	电邮	注册时间	积分	经验	等级
worker undefined	smallda@outlook.com	2018-5-8	10	20	10
worker undefined	smallda@outlook.com	2018-5-8	10	20	10
worker undefined	smallda@outlook.com	2018-5-8	10	20	10
worker undefined	smallda@outlook.com	2018-5-8	10	20	10
worker undefined	smallda@outlook.com	2018-5-8	10	20	10
worker undefined	smallda@outlook.com	2018-5-8	10	20	10

用户总表



注册人数/日期曲线



用户统计和注册人数曲线图

数据可视化

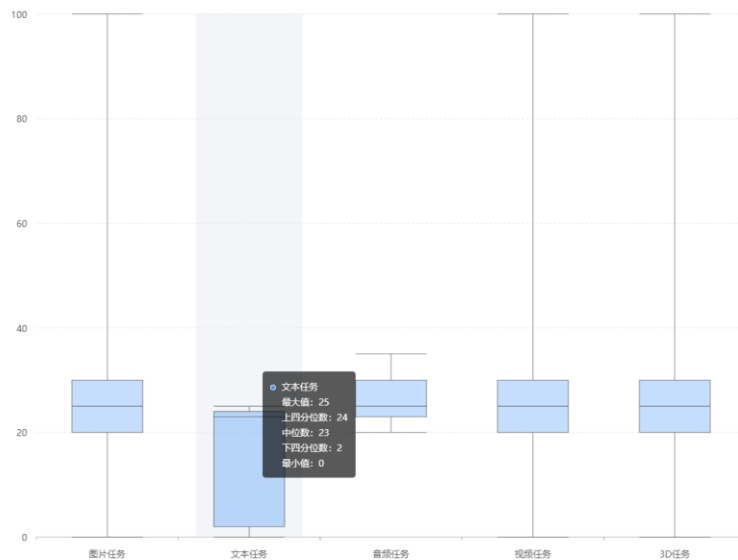
积分统计，帮助管理员管理公司资产

积分分布



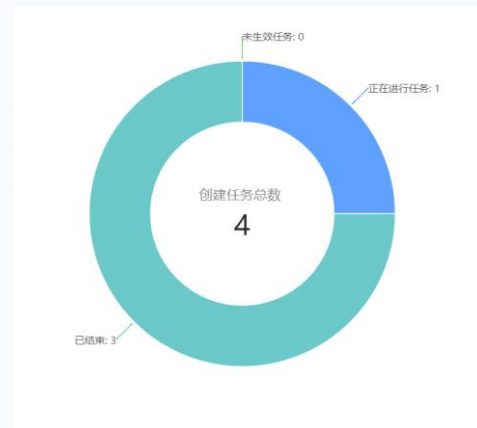
积分分布情况

各类型任务积分分配图

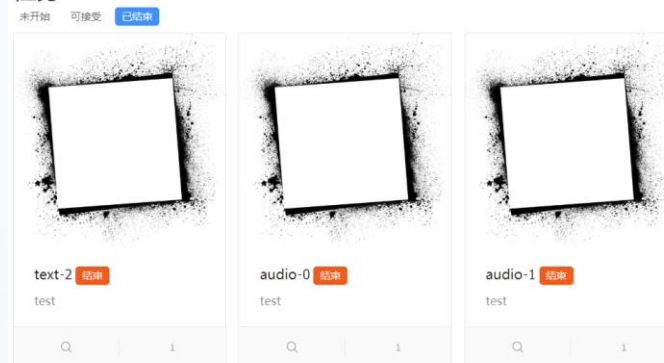


各类任务分发的积分的箱型图

数据可视化

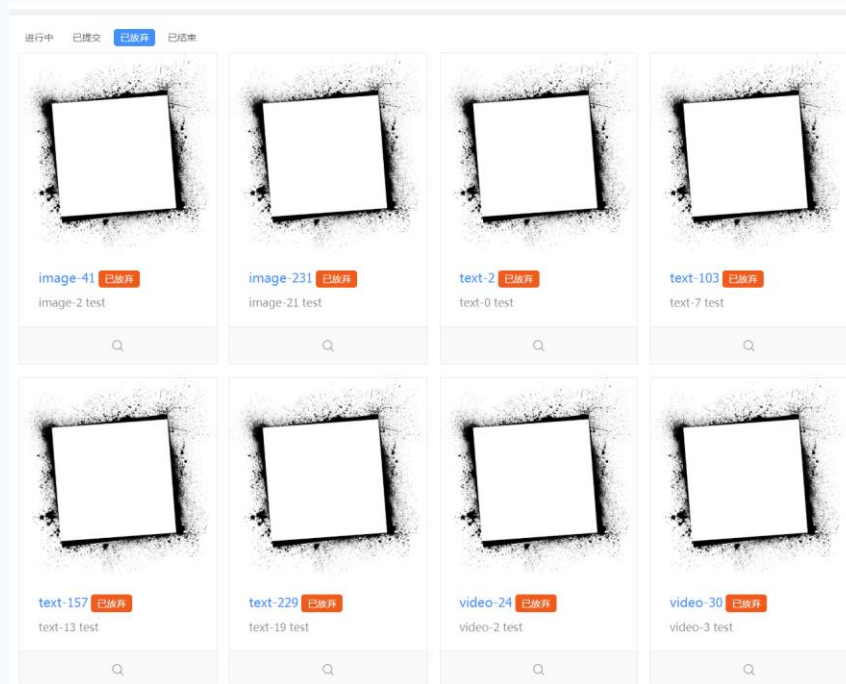
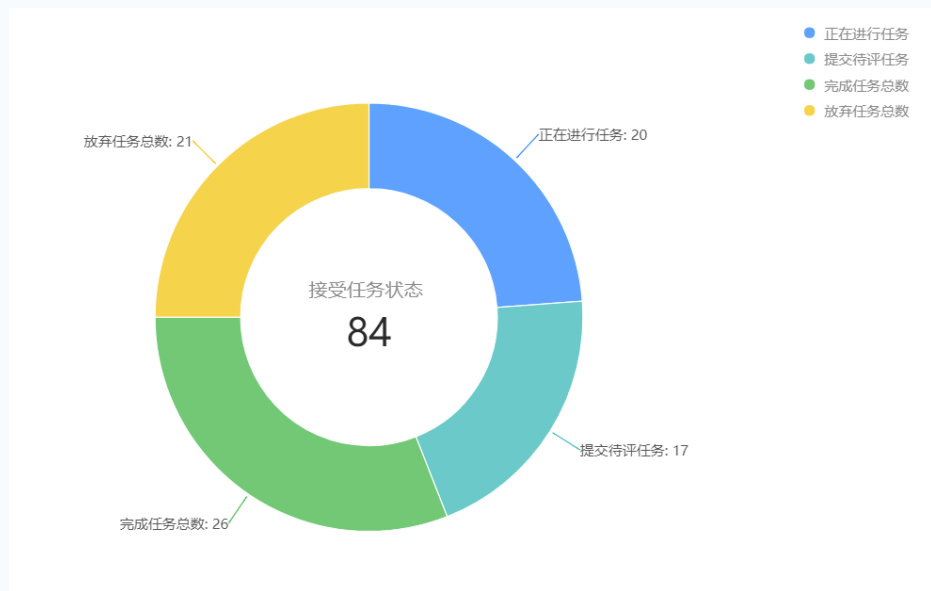


任务



发起者查看自己的所有实例/特定任务的实例的状态分布
点击进入任务列表

数据可视化

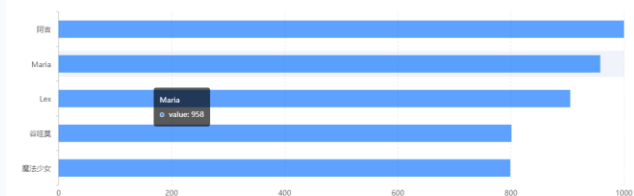


工人查看自己接受的任务的状态分布
点击进入任务列表

数据可视化

经验排行榜

巅峰榜单

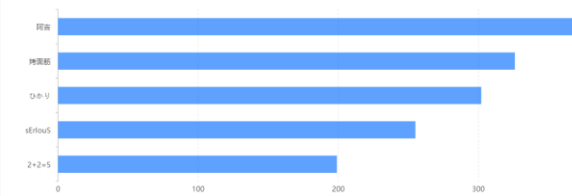


总排行榜

用户名	积分	等级	排名
阿吉	1000	10	1
Maria	958	9	2
Lex	905	9	3
谷粒菜	801	9	4
魔法少女	799	8	5
冯存族秀	888	8	6
替木工	886	8	7
一只猫说	880	8	8
芝士蛋糕	878	8	9

工人富人榜

巅峰榜单

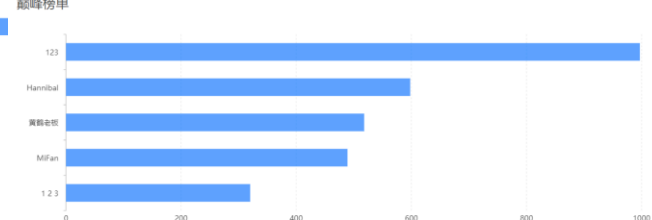


总排行榜

用户名	积分	排名
阿吉	397	1
sErlou5	255	2
神楽坂	326	3
D-k-i	302	4
2+2=5	199	5
刀马旦	296	6
老歪脖子树	193	7
一醉观人城	280	8

发起者富人榜

巅峰榜单



总排行榜

用户名	积分	排名
123	997	1
黄粉老板	518	2
MIFan	489	3
1 2 3	320	4
Hannibal	598	5
イチゴ	200	6
沙雕艺术家	798	7
黑小六六六	197	8

带Top5的排行榜



谢谢