

鉴别伪造图片

人工智能思维与伦理课程

2021.09

鉴别伪造图片



真实图片



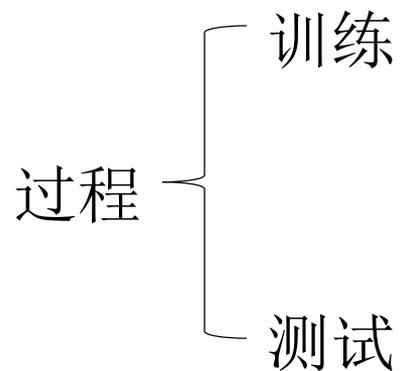
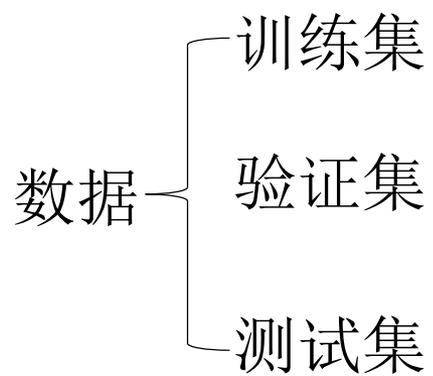
伪造图片

流程

- 从百度网盘下载Casia数据集。real文件夹包含800张真实图片，fake文件夹包含462张伪造图片。网盘链接：<https://pan.baidu.com/s/11b9YPp24t5xhLmHFgAUDrw>
提取码：46s9
- 采样若干张真实图片和伪造图片（比如400张真实图片，400张伪造图片）作为训练集(其中一部分会被划归为验证集)，其他图片作为测试集。
- 分类任务demo：

<https://bbs.huaweicloud.com/videos/58cae67982ba42efa806df5d6b9569d8>

流程



1. 进入华为云ModelArts->自动学习界面，选择图像分类
<https://www.huaweicloud.com/product/modelarts.html>

ModelArts 自动学习

ModelArts自动学习能力，可根据用户标注数据全自动进行模型设计、参数调优、模型训练、模型压缩和模型部署全流程。无需任何代码编写和模型开发经验，即可利用ModelArts构建AI模型应用在实际业务中。

[立即使用](#) [视频教程](#)

自动学习

零经验，玩转ModelArts自动学习

零编码，零AI基础，三步构建AI模型

ModelArts自动学习可以大幅降低AI使用门槛与成本，较之传统AI模型训练部署，使用自动学习构建将降低成本90%以上。

4+特定应用场景

目前，ModelArts支持图片分类、物体检测、预测分析、声音分类4大特定应用场景，可以应用于电商图片检测、流水线物体检测等场景。

ModelArts

总览
自动学习
数据标注 Beta
开发环境
训练作业
模型管理
部署上线
AI市场 Beta
专属资源池
全局配置

自动学习

- 图像分类**
识别一张图片中是否包含某种物体。[视频教程](#)
[创建项目](#)
- 物体检测**
识别出图片中每个物体的位置及类别。[视频教程](#)
[创建项目](#)
- 预测分析**
对结构化数据做出分类或数值预测。[视频教程](#)
[创建项目](#)
- 声音分类 New!**
识别一段音频中是否包含某种声音。
[创建项目](#)

所有分类

项目名称	项目类型	训练状态	训练数据集	创建时间	描述
flower-test	图像分类	运行成功	/flower-classification/...	2019/09/23 13:00:51 GMT+08:...	--

2. 上传数据后，标注数据，点击开始训练。

华为云 控制台 搜索 费用中心 资源 工单 企业 开发工具 备案 支持与服务 中文 (简体) hid_tj5eyoe6hod021

exeML-1123 < 返回自动学习 1 数据标注 2 模型训练 3 部署上线 数据版本 评价 使用指南 建议反馈

已标注 340 未标注 0 开始训练

删除图片 同步数据源 选择当前页

添加标签 已选 0 张图片
标签名 输入标签, 按Enter键添加
确定 取消

全部标签 2

标签名称	标签数量	操作
fake1	68	编辑 删除
real1	272	编辑 删除

3. 训练完成后查看训练结果并部署测试

The screenshot displays the Huawei Cloud ML training console interface. At the top, there is a navigation bar with the Huawei logo, '华为云 控制台', a search bar, and various utility links like '费用中心', '资源', '工单', '企业', '开发工具', '备案', '支持与服务', '中文(简体)', and a user profile 'hid_tj5eyoe6hod021'. Below the navigation bar, the main content area is titled 'exeML-1123' and includes a breadcrumb '< 返回自动学习'. Three progress steps are shown: '1 数据标注' (completed), '2 模型训练' (active), and '3 部署上线' (available).

On the left, the '版本管理' (Version Management) section shows a table with one entry:

状态	版本ID	创建时间	操作
已完成	V002 (fba9f18a-9cdf-4100-8475-...)	2021/09/17 09:06:35 GMT+08:00	部署, 删除

The '训练详情' (Training Details) section is divided into three panels:

- 训练状态:** 已完成. 开始时间: 2021/09/17 09:07:21 GMT+08:00. 训练时长: 00:02:15.
- 评估结果 (highlighted in red):**
 - 召回率: 0.708
 - 精确率: 0.861
 - 准确率: 0.883
 - F1值: 0.752
- 训练参数:** 训练时长不大于 (min): 60. 计算规格: 增强计算型1实例-自动学习 (GPU).

Below the training details, the '模型' (Model) is identified as 'exeML-1123_ExeML_331edbc0_0.0.2'. A '分类统计表' (Classification Statistics Table) is provided with a search bar '请输入关键字查询':

标签名	F1值	精确率	召回率
fake1	0.571	0.833	0.435
real1	0.932	0.888	0.981

4. 上传测试图片（也可以是自己PS的图片）

华为云 控制台

搜索 费用中心 资源 工单 企业 开发工具 备案 支持与服务 中文(简体) hid_tz5eyoe6hod021

exeML-1123 < 返回自动学习

1 数据标注 ✓ 2 模型训练 ✓ 3 部署上线

评价 使用指南 建议反馈

版本管理

2021/09/17 09:19:50 GMT+08:00 停止

运行中 47分钟 后停止 | 设置 删除

服务测试

当服务状态为“运行中”时，才能进行预测。

服务名称: exeML-1123_ExeML_1631841587860646082

选择需要预测的文件 上传 重新预测



预测结果

预测结果: real1

```
1 {
2   "predicted_label": "real1",
3   "scores": [
4     [
5       "real1",
6       "0.984"
7     ],
8     [
9       "fake1",
10      "0.016"
11    ]
12  ]
13 }
```

URL接口 接口调用指南

<https://9866b3c2d0d04c5dbc41957c9fd17c70.apig.cn-north-4.huaweicloudapis.com/v1/infers/7aa71566-253b-4504-8f54-7e8ed26c601b>

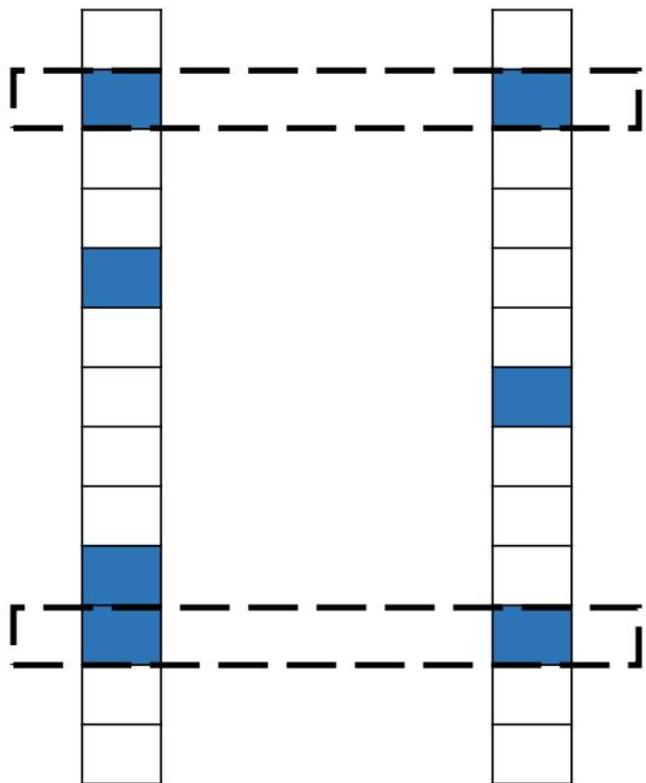
要求

- 理解召回率 (recall), 精确率 (precision), 准确率 (accuracy), F1值 (F1-score) 的含义。
- 尝试使用不同数量的训练图片, 尝试用不同比例的真实图片和伪造图片, 观察并分析验证集 (validation set) 上的性能。
- 观察并分析测试图片的预测结果, 总结失败案例 (failure case)。
- 提交至少5页A4纸, 五号字体 (12pt) 的实验报告, pdf格式。

要求

■ :1 伪造图片 □ :0 真实图片

ground-truth \mathbf{y} predicted $\tilde{\mathbf{y}}$



精确率 (precision): $\frac{|\mathbf{y} \cap \tilde{\mathbf{y}}|}{|\tilde{\mathbf{y}}|} = 0.66$

召回率 (recall): $\frac{|\mathbf{y} \cap \tilde{\mathbf{y}}|}{|\mathbf{y}|} = 0.5$

F1值 (F1-score): $\frac{2|\mathbf{y} \cap \tilde{\mathbf{y}}|}{|\mathbf{y}| + |\tilde{\mathbf{y}}|} = 0.57$

$F_1 = \frac{2}{\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}}$ 调和平均