

带细粒度属性的人像数据集FFHQ-Attr介绍

1 背景

2 数据集介绍

2.1 FFHQ亚洲人脸数据集

2.2 FFHQ-Attr细粒度亚洲人脸数据集

3 细粒度标注方法

3.1 基于CN_CLIP+CNN模型的预标注

3.2 人工double check

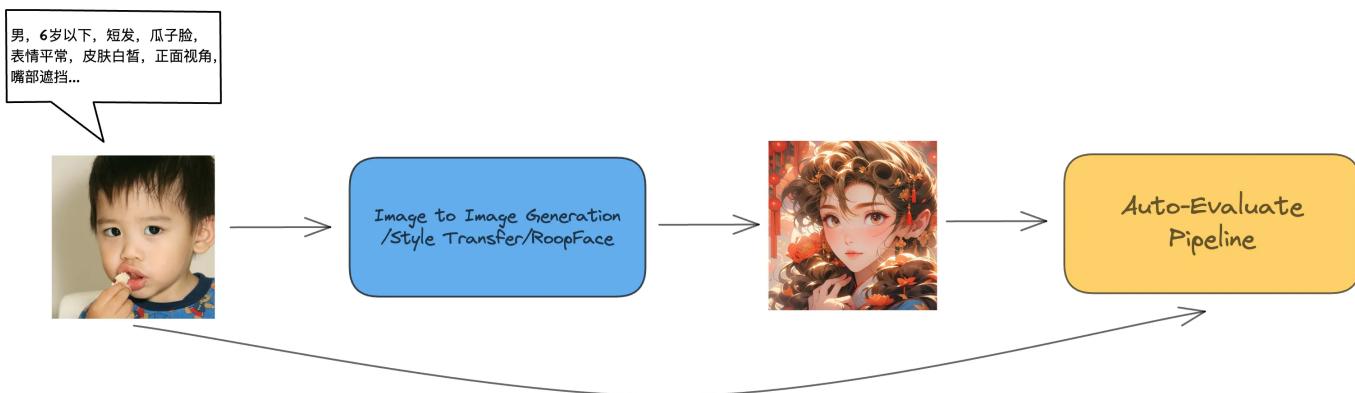
3.2.1 各细粒度指标的预标注结果

3.2.2 double check的时效对比

4 资源demo下载

1 背景

随着AI绘图能力的逐渐强大以及AI与人像交互业务的火爆，越来越多的工作已经从关注如何让AI生成、优化生成效果转移到如何对AI生成的质量进行评测并反向推进AI生成效果的优化。为了实现高效评测，benchmark（输入）的建设显得尤为重要，下面列举了图片风格转化场景的生成-评测流程图，其中FFHQ-Attr中的图片作为蓝色区域生成模型的输入，可以更加针对性地校验模型生成效果（假如我想知道模型对于小孩子生成效果如何等等）。



2 数据集介绍

2.1 FFHQ亚洲人脸数据集

FFHQ亚洲人脸数据

数据来源	简介	条数
FFHQ亚洲人脸数据	在年龄、种族和图像背景上丰富多样且差异明显，在人脸属性上也拥有非常多的变化，拥有不同的年龄、性别、种族、肤色、表情、脸型、发型、人脸姿态等，囊盖普通眼镜、太阳镜、帽子、发饰及围巾等多种人脸周边配件。	13061条

部分数据demo展示：



2.2 FFHQ-Attr细粒度亚洲人脸数据集

FFHQ-Attr数据集是基于13061条FFHQ亚洲人脸数据进行细粒度属性标注后得到的作为AI生图输入的人像评测benchmark,

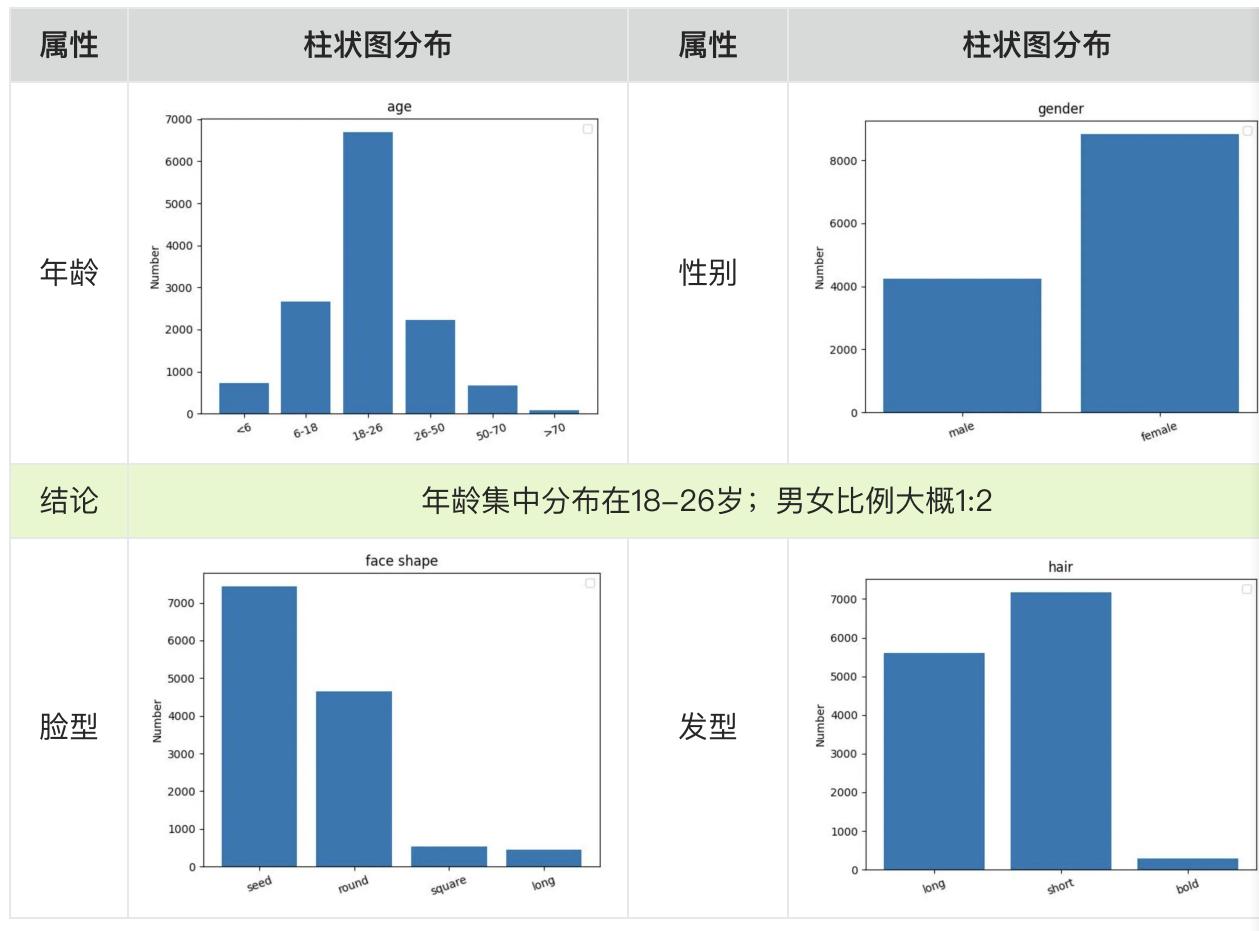
属性							
年龄	性别	脸型	发型	表情	肤色	视角	遮挡情况
<6	男	圆脸	长发	开心	黝黑	正面	额头厚刘海
6–18	女	长脸	短发	忧郁	白皙	侧面	额头帽子
18–26		瓜子脸	光头	悲伤	黄色		普通眼镜
26–50		方脸		愤怒			墨镜
50–70				平常			嘴巴遮挡
>70				惊讶			面部其他位置
							无遮挡

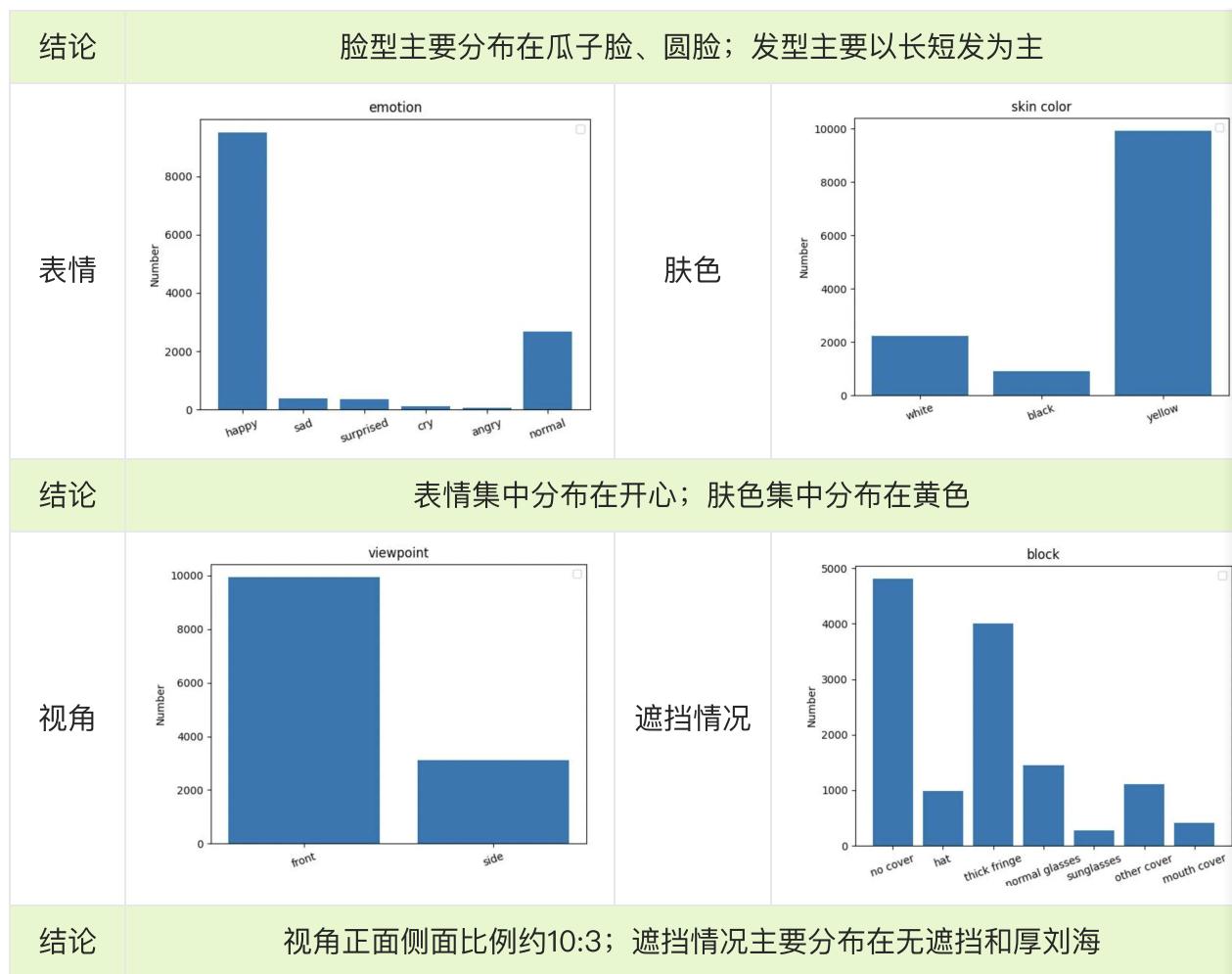
部分数据demo展示：

用户图	细粒度属性	用户图	细粒度属性
	<pre>{ "url": xxx, "attributes": { "hair": "短发", "year": "6岁以下", "sex": "男", "skin_color": "白皙", "face_shape": "瓜子脸", "emotion": "平常", "vision": "正面", "block": "嘴巴遮挡" } }</pre>		<pre>{ "url": xxx, "attributes": { "hair": "短发", "year": "18-26岁", "sex": "男", "skin_color": "黝黑", "face_shape": "瓜子脸", "emotion": "平常", "vision": "正面", "block": "眼镜-普通眼镜" } }</pre>
	<pre>{ "url": xxx, "attributes": { "hair": "长发", "year": "26-50岁", "sex": "女", "skin_color": "黄色", "face_shape": "瓜子脸", "emotion": "平常", "vision": "侧面", "block": "无遮挡" } }</pre>		<pre>{ "url": xxx, "attributes": { "hair": "长发", "year": "18-26岁", "sex": "女", "skin_color": "白皙", "face_shape": "瓜子脸", "emotion": "平常", "vision": "正面", "block": "额头厚刘海" } }</pre>

<pre>{ "url": xxx, "attributes": { "hair": "短发", "year": "6岁以下", "sex": "男", "skin_color": "黝黑", "face_shape": "瓜子脸", "emotion": "平常", "vision": "正面", "block": "无遮挡" } }</pre> 	<pre>{ "url": xxx, "attributes": { "hair": "短发", "year": "70岁以上", "sex": "男", "skin_color": "黄色", "face_shape": "长脸", "emotion": "平常", "vision": "侧面", "block": "眼镜-普通眼镜" } }</pre> 
---	--

具体属性数据占比分布情况：

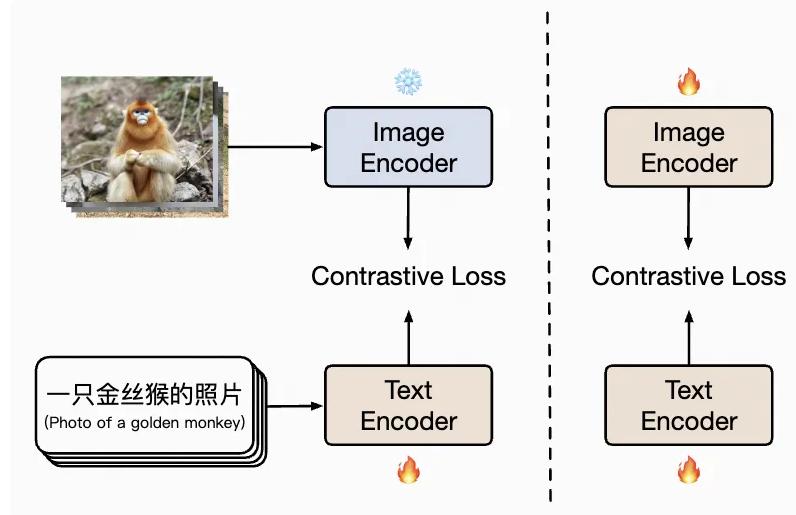




3 细粒度标注方法

3.1 基于CN_CLIP+CNN模型的预标注

首先，介绍下CN_CLIP，它是由阿里达摩院联合北航出品的大规模中文CLIP，训练数据使用了高质量的2亿图文对~，基于这2亿的大规模原生中文图文数据，通过如下图所示的二阶段预训练策略（一阶段仅训练文本侧，二阶段同时训练），实现了CLIP模型的中文化版本，即CN_CLIP；



其次，根据支付宝业务上的评测需求以及常识上的认知，设定了8个细粒度属性对应的共33个细粒度文本标签（见2.2），因此可基于CLIPScore的思路来对FFHQ亚洲人像数据集中的人像图做细粒度的预标注；最后，针对年龄、性别、视角三个和图片语义关系稍弱的属性，还额外采用了基于CNN的模型对细粒度属性进行精细化修正，当CNN模型的置信度较高时，会修改CN_CLIP的结果，对最终预标注的精度进行了提升。

3.2 人工double check

虽然上述3.1方案可以对FFHQ亚洲人脸数据继续预标注，但为了标注精度进一步提升的同时也能验证上述方案的实际效果，仍然对FFHQ预标注后的结果做了人工的double check，得到了如下的结论：

3.2.1 各细粒度指标的预标注结果

属性		精确率	属性		精确率	属性		精确率
年龄	<6	99.8% (585条)	发型	长发	58.4% (8766条)	遮挡情况	额头厚刘海	-
	6–18	86.1% (3044条)		短发	85.1% (2903条)		额头帽子	-
	18–26	93.2% (6912条)		光头 (标准问题)	17.6% (1372条)		普通眼镜	69% (226条)
	26–50	97.7% (1790条)	表情	开心	81.9% (11493条)		墨镜	64.3% (226条)

	50–70	96.8% (659条)		忧郁	47% (591条)		嘴巴遮挡	—
	>70	94.4% (71条)		悲伤	35.5% (200条)		面部其他位置	91.2% (1456条)
性别	男	99.9% (4222条)		愤怒	— (47条)		无遮挡	41.9% (11098条)
	女	99.8% (8839条)		惊讶	41.5% (777条)	视角	正面	87.9% (10763条)
脸型	长脸	43% (242条)		平常	— (2671条)		侧面	79.3% (2298条)
	方脸 (70%与瓜子脸混淆)	15% (1225条)	肤色	黝黑 (与黄色混淆)	12.8% (6831条)			
	圆脸	66% (5573条)		白皙 (与黄色混淆)	43.2% (4111条)			
	瓜子脸	81.9% (6021条)		黄色	88.5% (2119条)			

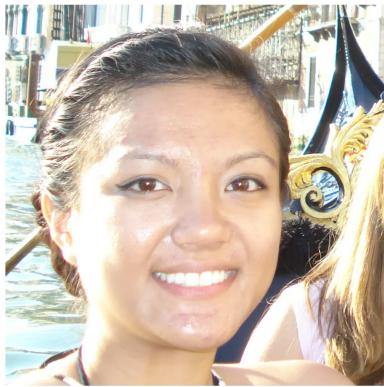
结论：修正过的属性标签识别精确率较高，包含语义的部分标签识别精确率低。

3.2.2 double check的时效对比

带有预标注标签的标注过程：

人像图: 	性别: <input checked="" type="checkbox"/> 人像-性别-男	年龄: <input checked="" type="checkbox"/> 人像-年龄-18到26岁
脸型: <input checked="" type="checkbox"/> 人像-脸型-瓜子脸	发型: <input checked="" type="checkbox"/> 人像-发型-长发	遮挡情况: <input checked="" type="checkbox"/> 人像-无遮挡
视角情况: <input checked="" type="checkbox"/> 人像-视角-侧面	表情: <input checked="" type="checkbox"/> 人像-表情-开心	肤色: <input checked="" type="checkbox"/> 人像-肤色-黝黑
<input checked="" type="checkbox"/> (单选) 性别判断是否有误 <input checked="" type="checkbox"/> 是 <input type="checkbox"/> 否		
<input checked="" type="checkbox"/> (单选) 年龄判断是否有误 <input checked="" type="checkbox"/> 是 <input type="checkbox"/> 否		
<input checked="" type="checkbox"/> (单选) 脸型判断是否有误 <input checked="" type="checkbox"/> 是 <input type="checkbox"/> 否		
<input checked="" type="checkbox"/> (单选) 发型判断是否有误 <input checked="" type="checkbox"/> 是 <input type="checkbox"/> 否		
<input checked="" type="checkbox"/> (单选) 遮挡情况判断是否有误 <input checked="" type="checkbox"/> 是 <input type="checkbox"/> 否		
<input checked="" type="checkbox"/> (单选) 表情判断是否有误 <input checked="" type="checkbox"/> 是 <input type="checkbox"/> 否		
<input checked="" type="checkbox"/> (单选) 肤色判断是否有误 <input checked="" type="checkbox"/> 是 <input type="checkbox"/> 否		

不带预标注标签的标注过程：

图片: 	(单选) 年龄 6岁以下 <input type="checkbox"/> 6-18岁 <input checked="" type="checkbox"/> 18-26岁 <input type="checkbox"/> 26-50岁 <input type="checkbox"/> 50-70岁 <input type="checkbox"/> 70岁以上以上
	(单选) 性别 <input checked="" type="checkbox"/> 男 <input type="checkbox"/> 女
	(单选) 脸型 圆脸 <input type="checkbox"/> 长脸 <input type="checkbox"/> 方脸 <input checked="" type="checkbox"/> 瓜子脸
	(单选) 发型 长发 <input checked="" type="checkbox"/> 短发 <input type="checkbox"/> 光头
	(单选) 遮挡情况 额头厚刘海 <input type="checkbox"/> 额头帽子 <input type="checkbox"/> 墨镜 <input type="checkbox"/> 普通眼镜 <input type="checkbox"/> 嘴部遮挡 <input type="checkbox"/> 面部其他部分 <input type="checkbox"/> 无遮挡
	(单选) 视角 <input checked="" type="checkbox"/> 正面 <input type="checkbox"/> 侧面
	(单选) 表情 开心 <input type="checkbox"/> 平常 <input type="checkbox"/> 忧郁 <input type="checkbox"/> 愤怒 <input type="checkbox"/> 惊讶 <input type="checkbox"/> 悲伤
	(单选) 肤色 黝黑 <input type="checkbox"/> 白皙 <input checked="" type="checkbox"/> 黄色

经过测算，在保证标注内容完全随机且没有重复的情况下，带有预标注标签的标注任务时效=285.71条/小时/人；

不带预标注标签的标注任务时效=180条/小时/人，可提效约**59%**。

4 资源demo下载

github: <https://github.com/created-Bi/VQA-GPT>