

2018 TAIL CAMP AI实战训练营



图像项目

学习材料指引

2018-01-20

Week 2: 人脸表情识别

总体目标

- 熟悉数据处理与数据增强在深度学习中的作用
- 利用网络数据集，实现基于CNN的人脸检测和人脸表情识别

学习指引

Day 1: 表情识别是什么？难在哪里？

表情是人不可观测的内心活动在可观测空间内最直接的体现，通过分析人的表情我们可以判断一件产品的质量，一部电影的水准，或

者是一次交流的内容。在好未来课堂上的教学场景中，学生的表情往往代表着他们对老师与课程潜意识的实时反应，从而成为我们评估教学质量的重要标准之一。

1. 表情识别很难吗？

1971年，心理学家Ekman与Friesen的研究最早提出人类有六种主要情感，每种情感以唯一的表情来反映人的一种独特的心理活动。这六种情感被称为基本情感，由愤怒（anger）、高兴（happiness）、悲伤（sadness）、惊讶（surprise）、厌恶（disgust）和恐惧（fear）组成。本次训练营采用的数据图片均为网络爬虫收集+人工清洗（因此可能会出现少量错误数据），考虑到恐惧与厌恶的比例极少，我们去除了这两种表情，同时加入了中性（neutral）图片，所以最后是一个五分类问题。

表情识别的难点在于：1.表情受人脸姿态及环境光照的影响较大，鲁棒性不强；2.各种表情的程度都有强弱之分，程度强的表情变化较大，比较容易识别，而弱表情甚至微表情的识别效果仍然需要提高。总而言之，当前在非受限环境下表情识别的精度能达到80%已经是相当不错的成绩了，感兴趣的同学可以查阅现有的几个challenge了解相关情况：

FER2013: <https://www.kaggle.com/c/challenges-in-representation-learning-facial-expression-recognition-challenge/leaderboard>

Sighthound: <https://arxiv.org/pdf/1702.04280.pdf>

EmotiW: <https://sites.google.com/site/emotiwchallenge/home>

2. 觉得不够难吗？那就来制造点儿困难吧！

表情识别的难度还是有滴，但本质仍然是一个分类问题，这样岂不是和第一周没什么实质区别了？怎么可能这么简单呢？让我们来看看此次用到的数据长啥样：



发现问题了吗？是的，本周提供的数据将不再像上周一样已经为小伙伴做好了事先检测的工作，一个个字码得整整齐齐严丝合缝，而是需要大家自己来打败隐藏boss——人脸检测。预知后事如

何，且听下回分解。

Day 2&3: 实战！人脸表情识别之前传！

人脸检测是一切人脸分析（包括但不限于表情识别、人脸识别、年龄识别、性别分类、美颜...）所依赖的基础能力，好的人脸检测算法可以为后续的功能提供坚实的基础与更高的起点。

抛开Viola-Jones，HOG+SVM等传统检测算法不谈，近年来基于深度学习的人脸检测算法取得了长足的发展与进步，我们既可以采用性能优异的通用检测模型将其应用于特定目标——人脸上，比如：

Faster-RCNN: <https://arxiv.org/abs/1506.01497>

SSD: <https://arxiv.org/abs/1512.02325>

也可以看到一系列专攻人脸检测的高速靠谱网络，比如：

MTCNN: https://kpzhang93.github.io/MTCNN_face_detection_alignment/

FaceBoxes: <https://arxiv.org/abs/1708.05234>

SFD: <https://arxiv.org/abs/1708.05237>

请注意！在这次活动中，人脸检测作为一项额外的challenge，会给大家提供训练数据和验证数据（数据来源：FDDB & WIDER FACE，同学们也可自行下载训练），但并不会对检测的效果进行量化考察（当然检测的精度会一定程度地影响表情识别的精度）！这意味着大家在这一步有三条支线剧情可以触发接下来的表情识别任务，依难度划分如下：

1. 调研现有的人脸检测方法，自己训练模型或finetune现有模型
2. 调研现有的人脸检测方法，直接使用现有模型进行人脸检测
3. 不使用人脸检测，直接将原图片作为下一步表情识别的输入

上述三种方法肯定都是可以达到最终的目的地的，哪一种是最优解也并非绝对，如何选择取决于小伙伴们对自身能力的了解程度、对检测任务的熟练程度以及对全局时间的掌控程度。当然，无论选择哪种方法，希望小伙伴们如实地将自己的思考过程与处理步骤写入技术文档，作为后期回顾和评价的珍贵资料。

P.S. 在我们的表情数据库中，某一张图片里可能包含了多张人脸，在这种情况下，需要挑选出检测器给出置信度最大的那一张人脸作为该图片的唯一输出；相反的，如果你的检测器没有能够检测出任何一张人脸，建议直接使用原图作为整个人脸输入。

Day 4&5: 实战! 人脸表情识别之正传!

经过人脸检测的处理后，每一张图理想的BBOX随即生成，人脸区域的提取就已经完成（学有余力的小伙伴也可以尝试做一下face alignment对比最后的结果），接下来就可以进入我们熟悉的表情识别环节啦~

表情识别作为一个和OCR类似的分类问题，小伙伴们可以freestyle地来进行数据处理和网络构建。不过，大家在观察数据的时候就会发现：五个类别之中，不同类别之间的样本数量是有着很明显的差别的，有的甚至差了一个量级。没错，这就是表情识别任务里自带的彩蛋：**不均衡样本分布!** 那么，这就需要大家来尝试解决这个问题了，是通过欠采样来平衡样本呢？还是进行部分数据增强？亦或是考虑引入样本权重？或者你有什么其他的脑洞？尽管大胆地尝试吧~

P.S.1. 其实在真实场景下样本不均衡的问题要远比此次活动中提供的数据里呈现的要更加严重，这里只是简单地模拟一下。另外，由于样本均衡度带来的问题我们在测试中也会考虑到，**本周的最后精度计算方式与上周肯定是不一样的，请大家知悉。**

P.S.2. **重要!** 训练集中的label就包含在文件名里，但请大家制作数据集时一定要注意采用统一的label: 'anger':0, 'happiness':1, 'neutral':2, 'sadness':3, 'surprise':4, 这样才能和ground truth进行比较。

Day 6: 实战! 人脸表情识别之后传!

和第一周的文字识别一样，最后一步也是对我们的test集进行验

证。在这里，只需要把test的每一张图片输入训练好的网络，输出预测的label值，并将图片文件名和label写入指定格式的json文件（参考我们提供的submission_sample.json文件），**上传提交到平台上等待结果就可以啦！**

（此处重复第一周，重要的事说两遍）其实，相对数字化的结果输出，更重要的是整个学习和研究的过程。因此，除了最终的submission，也请同学们静下心来总结第二周的心得体会：从数据处理到网络设计，从模型训练到超参数调整，可以是train和val的loss曲线，也可以是你在训练中遇到的给你带来最大困扰的bug，把所有的亮点和难点都记录成一份简单的技术文档吧。

最后需要提交的文件包括：

- code（必选）
- 技术文档（必选）
- submission_sample.json