

实验报告

181830196 匡亚明学院 吴启龙

实现

之前也自己搭了一个很简单的网络，但是准确率只有百分之四十几，因此决定找一个已经搭建好的（也为了取得更好的思考题的可视化效果）。

网络搭建基于 <https://github.com/phamquiluan/ResidualMaskingNetwork>，自称在 FER2013 数据集上是 state of the art (SOTA)。但是它用到了额外的数据集。因此，我对它进行的改造，实现了 `split_data.py` 将仅有的 `train.csv` 划分为训练集（90%）和验证集（10%）。思考题的实现中，除了最后一题已经有现成的之外，其他的都是我基于网上的资料整合到作业中的（lime 参考的官方文档，其他参考的博客，知乎等，都需要一定的改造）。需要注意的是目录中只有少量文件与本作业有关，用的网络模型基于 AlexNet（其他的是一些其他的模型，其他的训练方式之类）。（最后添加了数据清洗，可能会提升准确率（原为 0.68514），但是这里来不及写了；readme 是原来目录的）

思考题

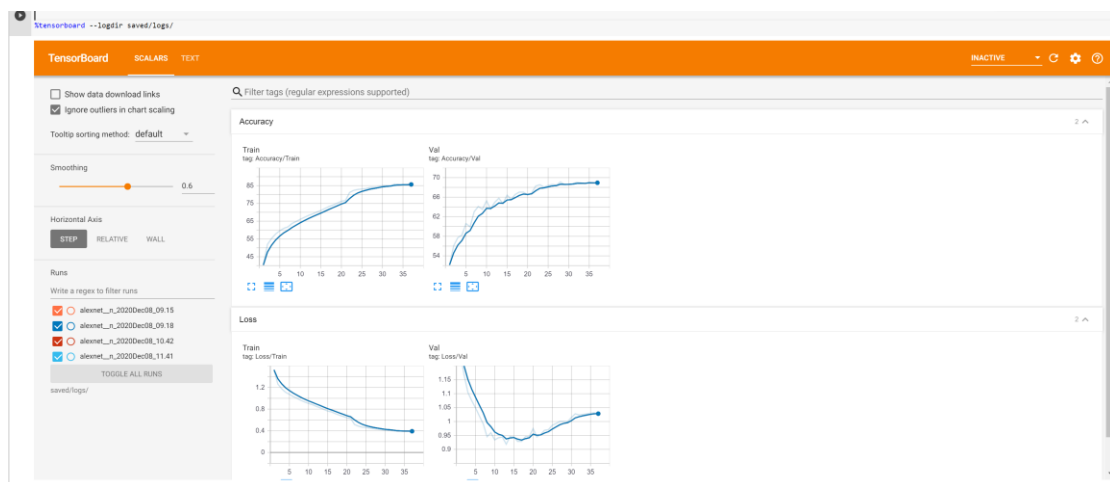
前三题的结果在 `result` 目录下。最后一题结果请用 `tensorboard --logdir saved/logs/` 在项目目录下查看。值得注意的是，不知道为什么（可能和 `import` 库有关）会产生大量的页面文件，我 24GB 内存+32GB 虚拟内存才跑起来（占用的内存并不多，但是打开任务管理器可以看到 `committed` 占用非常多，可达 50 多 GB）。可视化结果均基于验证集。我认为基于测试集比较合理，更反映实际情况，但是测试集的标注并没有给出。如果基于训练集可视化效果可能会更好，但是更加不能够反映真实情况。

1. 总共可视化出了验证集的十张图片（具体数量可调）。通过观察不难看出，模型关注的主要为面部，是比较合理的。有的甚至可以看出面部的五官。然而，也出现了一些例外的情况，例如第一张图片中，模型除了关注面部还关注了图片中的人的手和拿着的手机。可以推测正脸的照片的效果可能是最好的，因为最后一张关注的轮廓最清晰，而且集中在五官。

2. 初始化时用的是随机生成的噪声图片，选择的是第一层 CNN。在可视化出的 64 张图片中，相当一部分显示出有规律的条纹，以横向，纵向，约四十五度斜向为主，也有的呈现出有规律的斑点状。这与底层卷积神经网络捕捉图像低级（局部）特征相吻合。

3. 根据官方文档，lime 是利用轻微扰动来分析图像各部分对分类结果的重要性。发现数据集中有的数据标注的表情似乎并不正确。愤怒：基于五官的某一部分，看不出什么规律。厌恶：集中在眼睛和鼻子上部。恐惧：似乎与眼睛有关，但是也不太明显。快乐、忧伤、惊奇、中性：没有找到什么规律。总体来看大部分情况下网络注意的区域在面部，比较合理，但具体的部位并不精细，需要提高。

4. 利用 TensorBoard 可视化，可以得到训练期间的 loss, accuracy 等变化，还可以比较不同次训练的效果，非常的方便（开组会时看到学长在科研中用）。注意到 TensorBoard 对曲线进行了钝化处理，可以更好地反应趋势而不受小的波动的影响。可以看到刚开始的时候 loss 和 accuracy 变化较快，而后逐渐放缓。虽然验证集上的 loss 在中后期有反弹，似乎有过拟合的倾向，然而 accuracy 可以继续上升。loss 和 accuracy 不一致说明 loss 的设计还有改进的空间？通过日志信息可以得知学习率的动态调整和训练终止都是基于验证集准确率的。每当验证集准确率在几轮训练后不再上升，就降低一次学习率；学习率降低数次后训练停止。学习率下降可短暂地加速 loss 的下降或 accuracy 的提升。



其他

这次作业因为中间有很多事情耽误了，比较仓促，望谅解。