

神经网络作业一实验报告

洪亮 181240019

问题描述

有48x48像素的面部灰度图像组成的数据集。面部已自动对齐，一般ROC应该居中，并且在每个图像中脸占据大约相同的空间。任务是根据面部表情中显示的情感 classify为七个类别。

主要难点在于本次数据集精度较低而且标注质量一般，以及表情较为复杂细腻难以区分。

说正事前的一些嘀嘀咕咕

不得不吐槽一下这个数据集，[3995, 436, 4097, 7215, 4830, 3171, 4965]这是七个种类各自的样本数量。disgust居然只有436个样本？？惊了。第一反应是数据集需要增广，但别的数据也少得可怜啊啊啊！

另外依然是关于数据，似乎有不少异常角度的图片，即极端的侧脸/俯视仰视，我以为标注的时候需要把这些和正常样本区分开来。

模型搭建

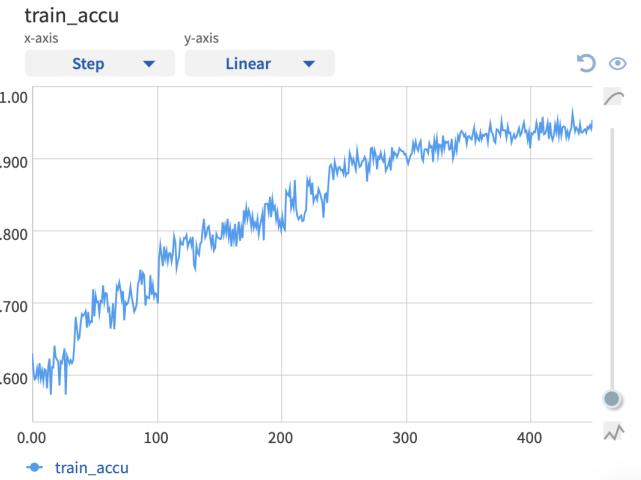
主干网络选取

尝试了以下几种网络backbone：

- vgg
- resnet
- mobilenet

由于三者的预训练模型都是 $228*228*3$ 的输入，我最开始尝试的时候是直接将 $48*48*1$ 作为输入且学习率选取较大，所以预训练模型基本没有起作用，后续就没有再做迁移。由于数据量较小，所以只有vgg13, vgg18, resnet18基本能达到预期。

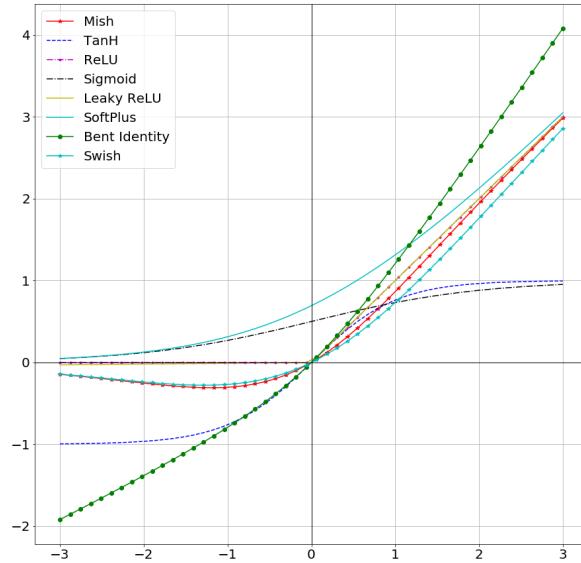
尝试了将resnet18参数量压缩到 $1/3$ ，在笔记本上训练（GTX1050Mobile）但是测试精度只能徘徊在0.59附近。



尝试添加的模块

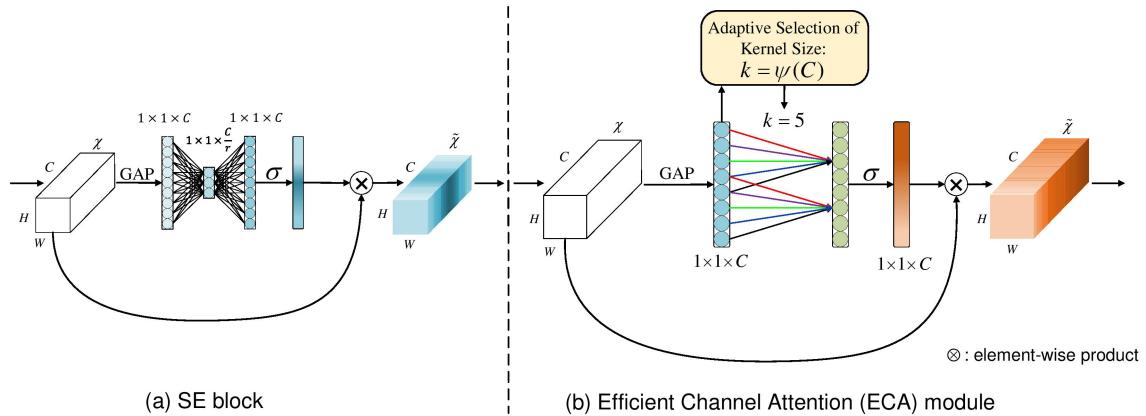
主要尝试了两个模块，Mish¹和ECA²

因为ReLU直接丢弃了负值过于极端，所以尝试采用了Mish，其函数如下。



结果上来看，训练/测试更加稳定（玄学警告⚠️⚠️）

第二个模块为ECA(Efficient Channel Attention)，是2020年CVPR一篇文章，作为对2017年Momenta提出的SENet的改进，ECA-Net将SE-Module改进为更加类似self-attention，且去掉了全连接层减少了参数量。



ECA 模块可以完美整合进Resnet，能带来2%的精度提升，但同时使得训练变得不稳定。

模型训练

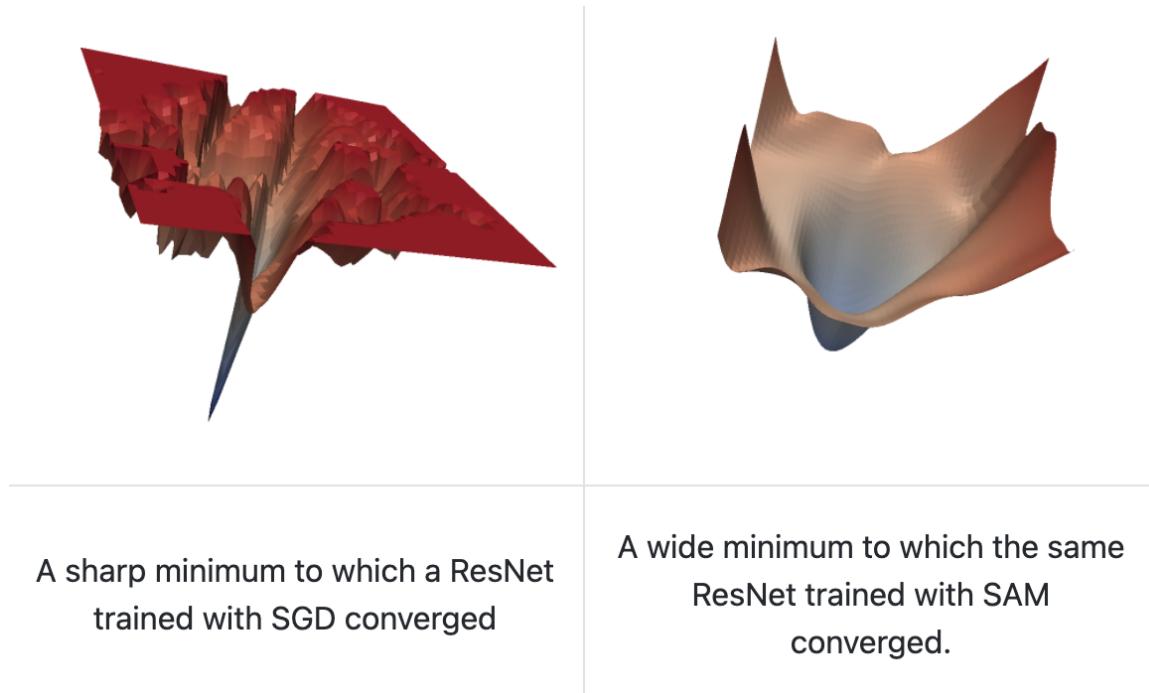
损失函数及优化器的选取

首先关于输出是否要归一化以及损失函数是采用交叉熵还是对数熵，就实验结果来看几乎没有区别。

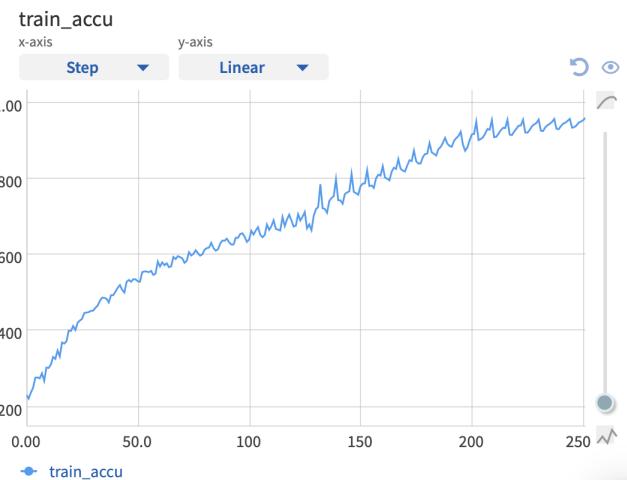
优化器选择上，Adam收敛性较差，故使用了最简单的SGD，结果如首页图。

后来找到一个好东西，谷歌ai部门的一篇文章SAM³(Sharpness-Aware Minimization)，此优化器能非常有效地防止过拟合，但需要手工调整搜索区域，目前只在简化版的ECA-Net上进行了调整，在后续模型上仍未收敛。

我实现的SAM是基于SGD的，相比SGD，SAM就像名字一样强制损失函数空间变得平滑，即并非向周围最低处移动而是向周围最高点最低的区域移动，这样能有效防止训练中的剧烈抖动以及过拟合。废话不多说上图。



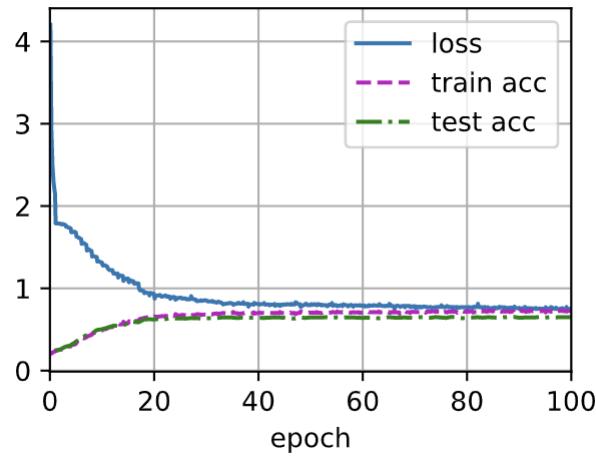
对比首页的resnet18，调整过的模型加入了ECA模块并使用了SAM作为优化器。在18, 28 epoch设置了step lr decay。28 epoch处的lr以及momentum还需要进一步调整，但之前的训练精度和验证精度基本保持在了6%以内。





fine-tune

直接挑重点说了。一件非常奇怪的事：vgg的效果优于resnet。经高人提点，图片应当进行上采样到与ImageNet的输入一致。但是啊，你妈的输入这么大，我的batch size就上不去了啊啊啊。最后折中用了 $128 \times 128 \times 1$ 。



此vgg与SAM联用，loss 0.752, train acc 0.723, test acc 0.648。但是由于我完全不明白为什么vgg会有这样的performance，所以最终还是弃用了。详细代码请见段首链接，version 1.

然后我就发现了一个令人眼前一亮的方案！！！

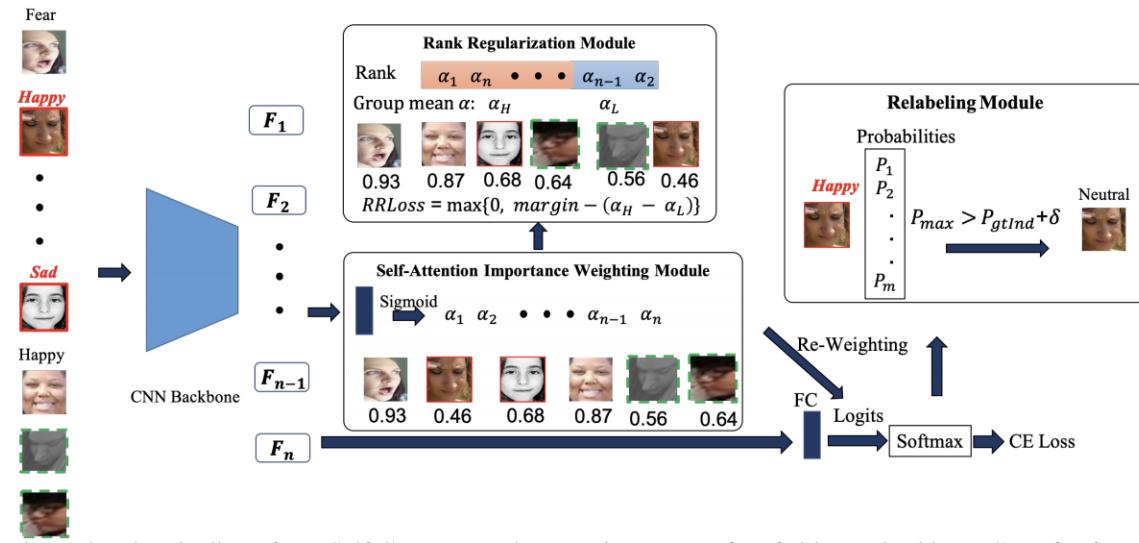
Self-cure network⁴ 是2020年CVPR的一篇文章，重点解决表情识别标注不确定性的问题。



Figure 1: Illustration of uncertainties on real-world facial images from RAF-DB. The right samples are extremely difficult for machines and even human which are better to be suppressed in training.

目前已有的方案是对单张图片进行多标注，也能达到不错的效果，但这样对于数据集的重新标注着实比较为难人，于是self-cure network就着力解决这个问题，即，认为标签存在错误，在一定epoch的学习后，尝试将错误标签修改为正确标签，增加收敛性能。

self-cure network基于resnet，我用了resnet18进行了实现。文中模型与resnet18的区别主要在于其最后的一个self-attention，是对于一个epoch内的图片对于参数更新的贡献加了attention。网络结构如下：



随后，会用attention权值划分区间，低于指定值的图片会被重新标记。

Training SCN with original labels on the RAF-DB dataset



Training SCN with the synthetic corrupted labels on the RAF-DB dataset without relabel module.



Training SCN with the synthetic corrupted labels on the RAF-DB dataset with relabel module.

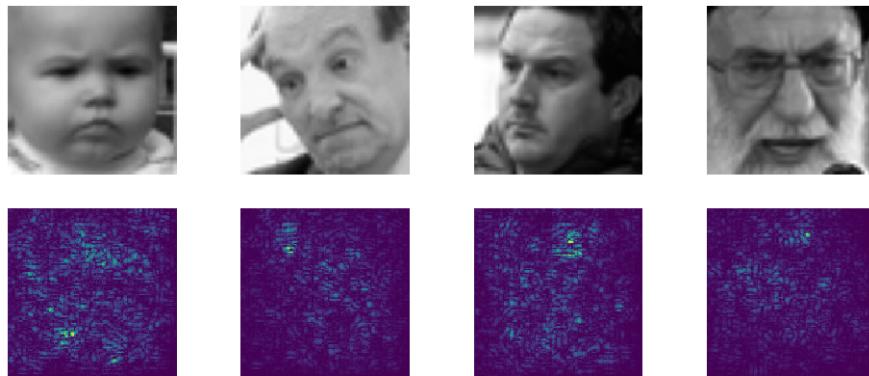


但是，我尝试将ECA模块和SAM拼接上去使用，暂时未能收敛到最优，可能是增加了ECA模块后相比简化版的resnet18，完全参数的需要有更好的初始化。SAM已知需要手工多次调整搜索范围，待有空再玩。

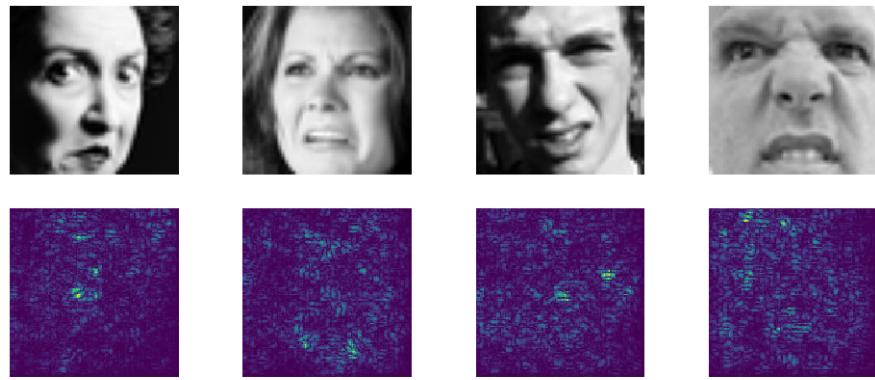
任务一

绘制Saliency map。此处分7类各自绘制：

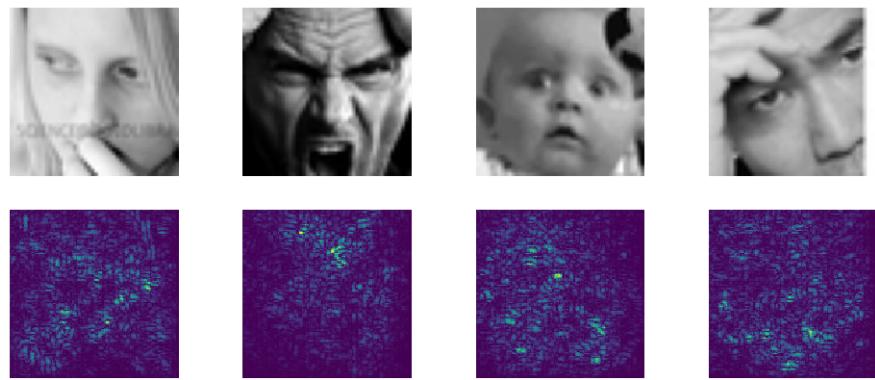
Angry:



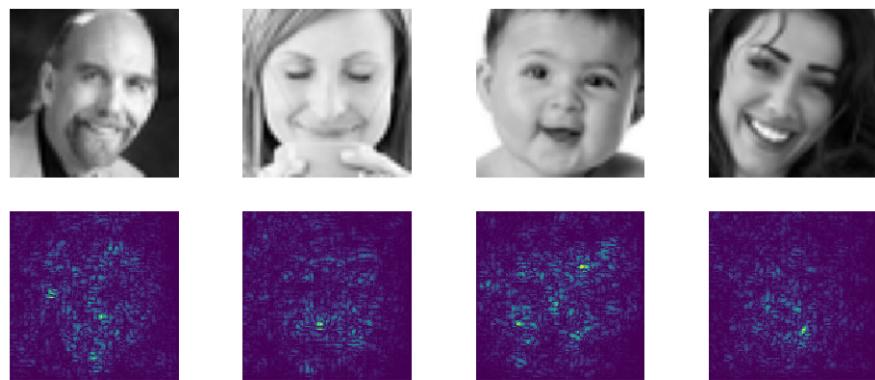
Disgust:



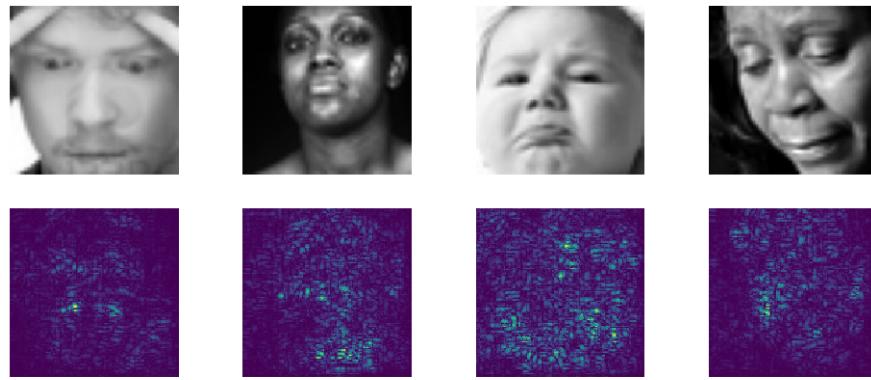
Fear:



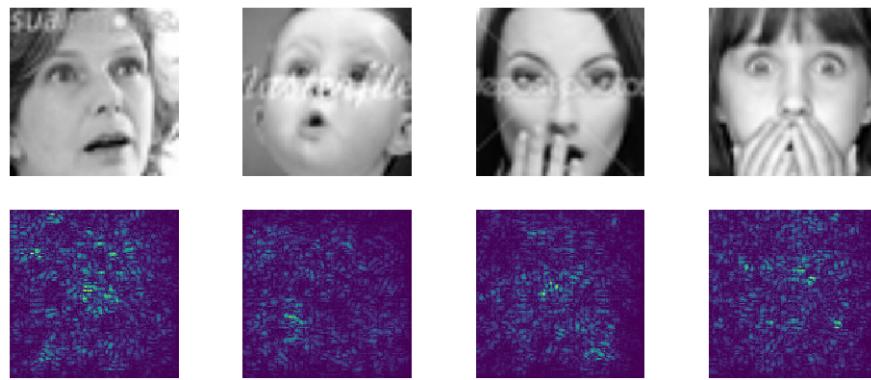
Happy:



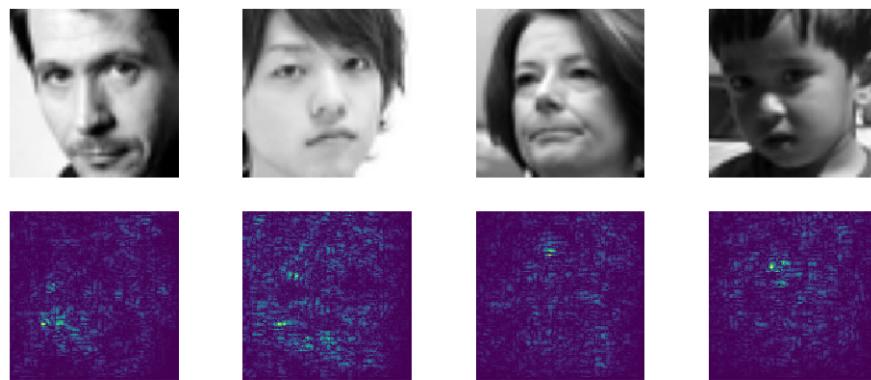
Sad:



Surprise:



Neutral:

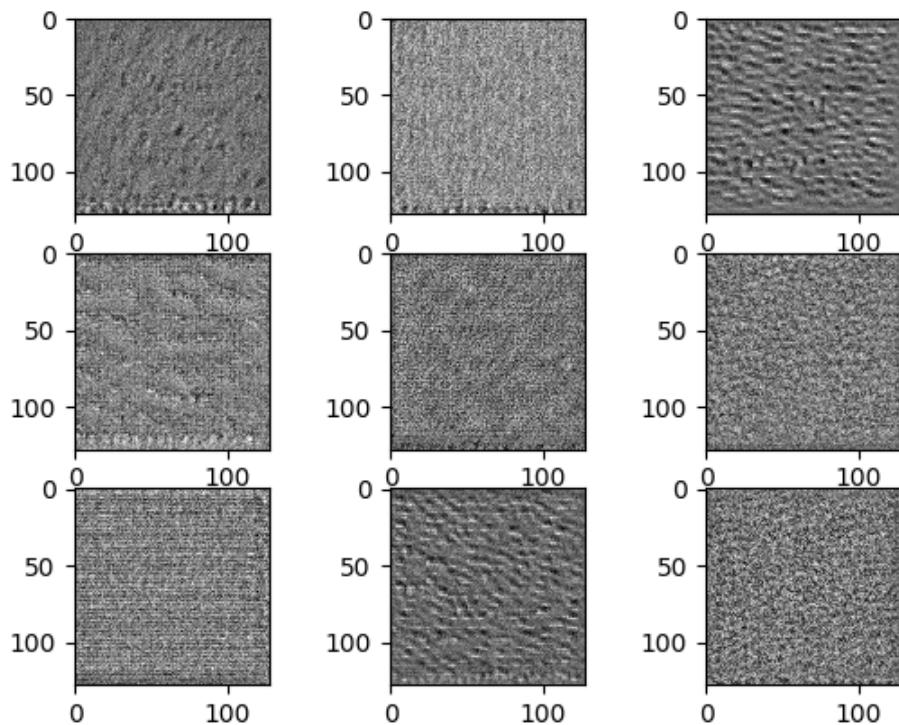


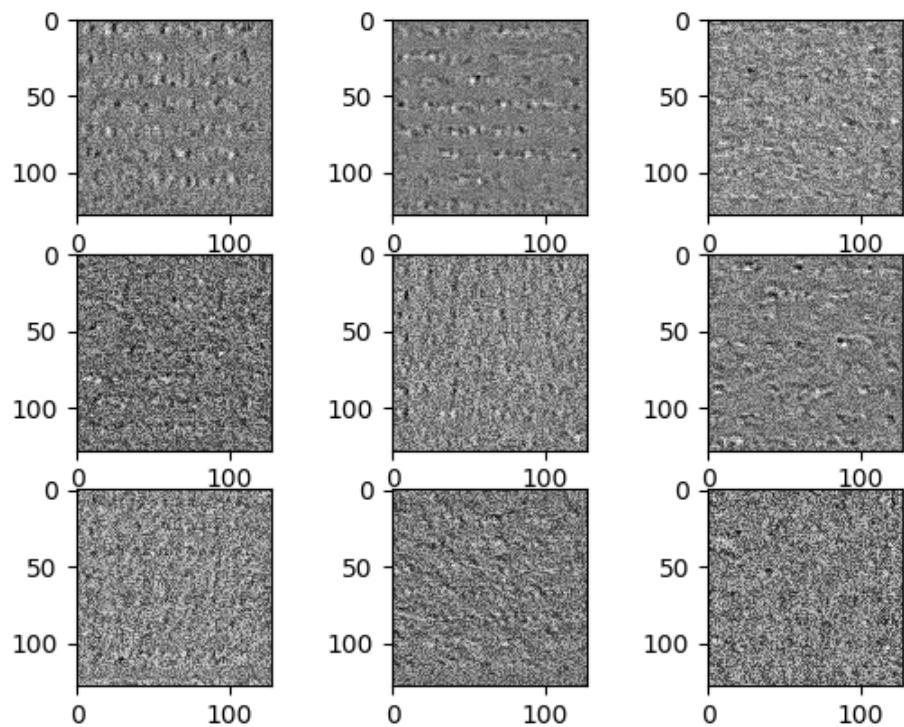
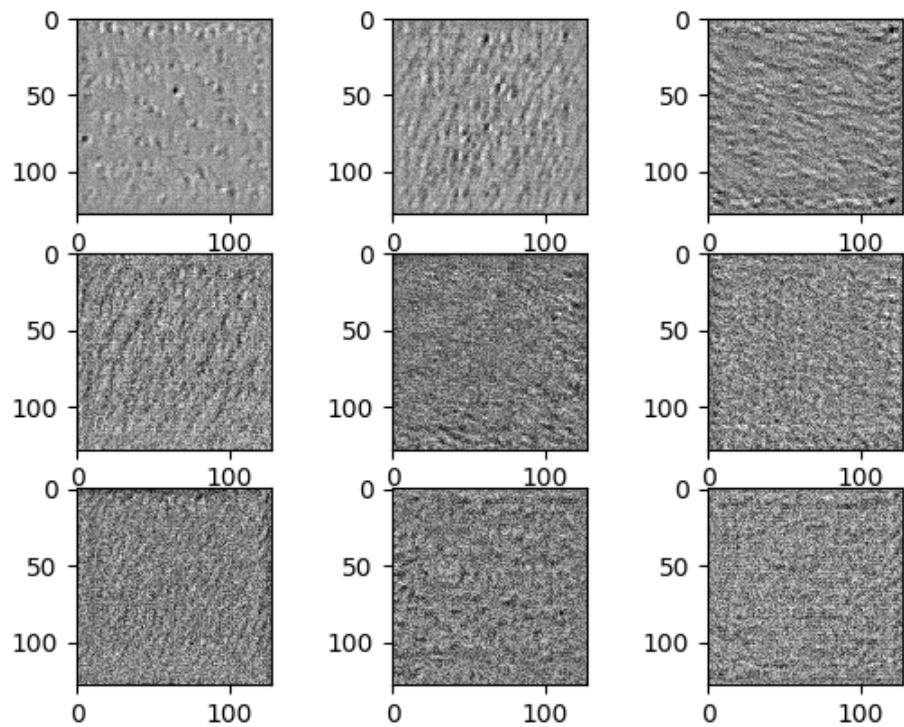
可见主要是对眉毛，鼻子和嘴较敏感。

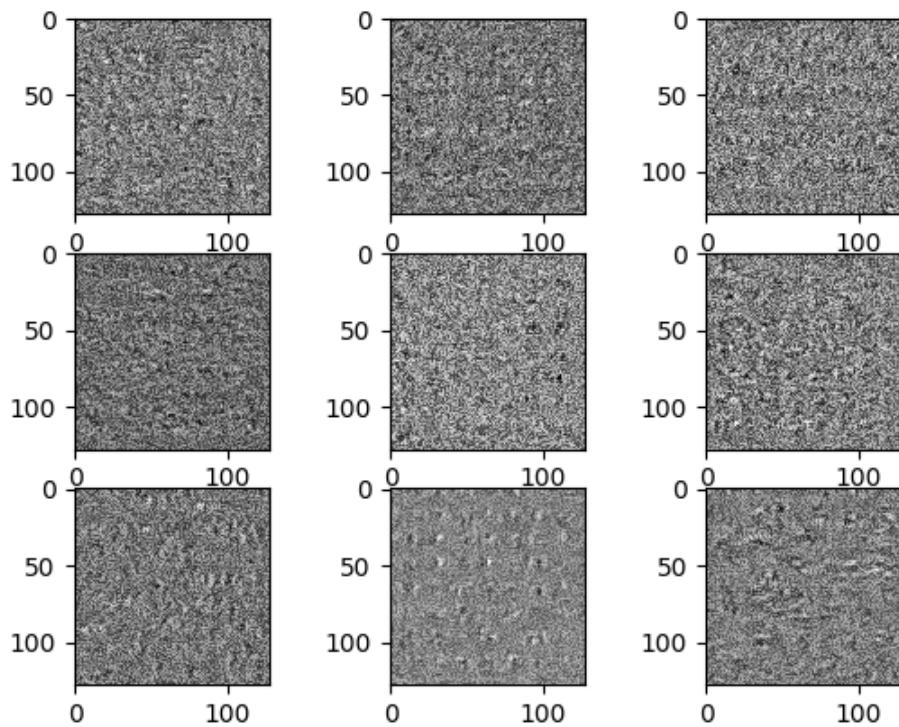
任务二

Gradient Ascent 方法

resnet18共4个basic block，每个第一个层卷积取前9个channel可视化







任务三

LIME要RGB图才行。。。灰度图强行转三通道很奇怪的说。

任务四

观察CNN模型的训练

emmm没太懂啥意思，我就可视化一下一张图是怎么被分类器分辨的吧。

ScoreCam⁵ 是CMU的一篇关于CNN可视化的文章，对比以往的Cam方法最大优势在于不需要梯度。pipeline如下：

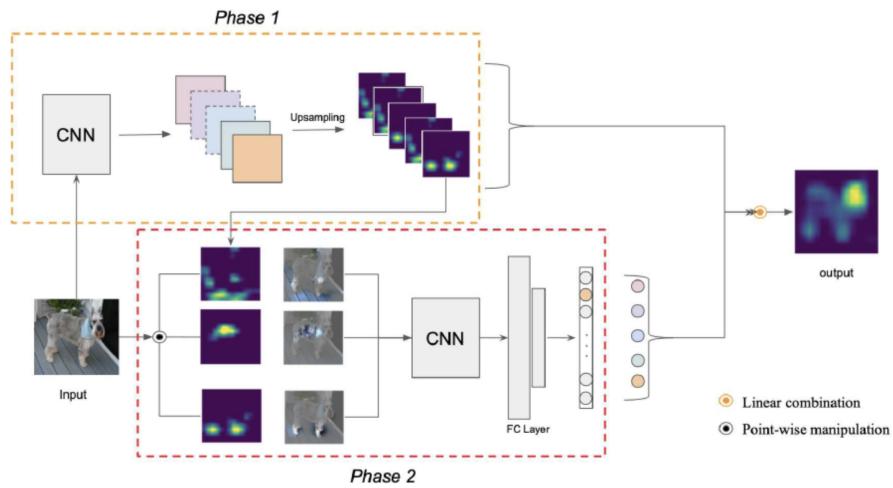
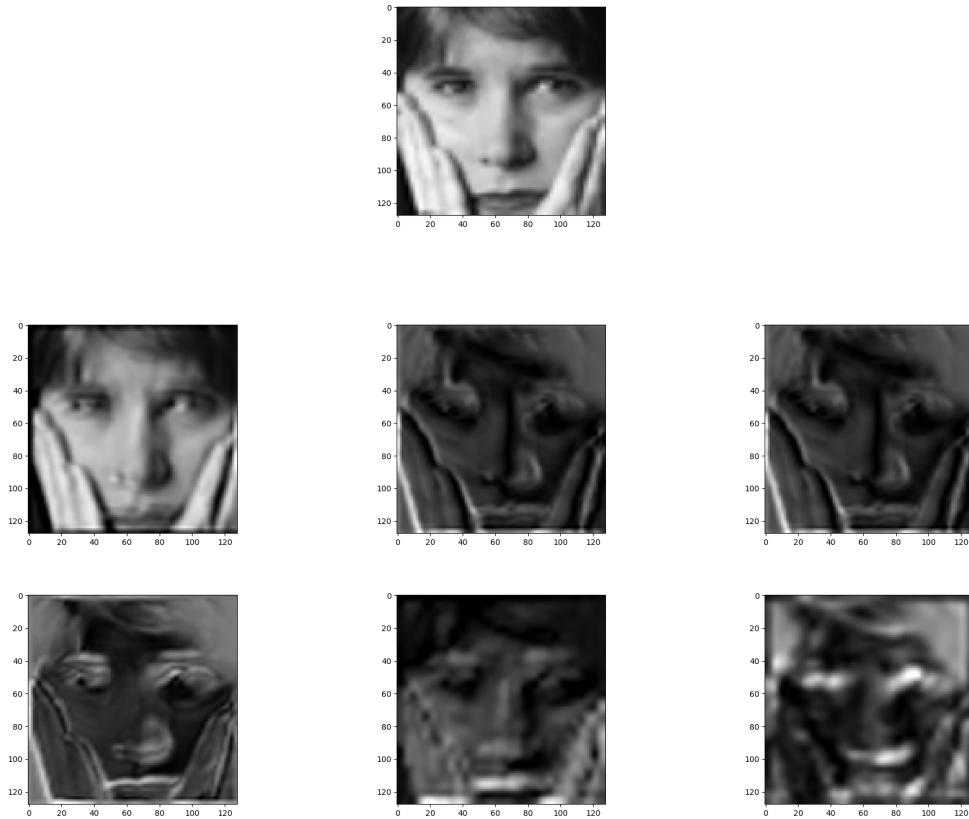


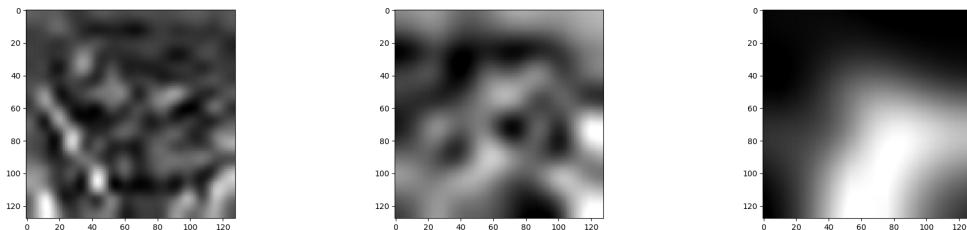
Figure 3. Pipeline of our proposed Score-CAM. Activation maps are first extracted in Phase 1. Each activation then works as a mask on original image, and obtain its forward-passing score on the target class. Phase 2 repeats for N times where N is the number of activation maps. Finally, the result can be generated by linear combination of score-based weights and activation maps. Phase 1 and Phase 2 shares a same CNN module as feature extractor.

https://blog.csdn.net/fq_34813925

然后奉上特征提取模块各层CNN的激活图：

从原图开始





讨论

数据集的增广

上述提到的数据重新标注方法仍然需要优化重标注的比例，随机旋转允许的角度仍然未知，因为数据集中图片已经裁剪到只有人脸了，旋转后有效像素可能被裁减掉。目前只能使用flip。

多步骤识别

此处为yy，仅供娱乐。

感觉可以用人脸关键点定位的方法，训练类似图网络。可参见CVPR2018一篇文章 PCD-CNN⁶ .

参考文献

- [1] Misra, Diganta, Mish: Self Regularized Non-Monotonic Activation Function. BMVC 2020
- [2] Banggu Wu, ECA-Net: Efficient Channel Attention for Deep Convolutional Neural Networks. CVPR 2020
- [3] Pierre Foret, Ariel Kleiner, Sharpness-Aware Minimization for Efficiently Improving Generalization. ICLR 2021
- [4] Kai Wang, Suppressing Uncertainties for Large-Scale Facial Expression Recognition. CVPR 2020
- [5] Haofan Wang, Zifan Wang, Score-CAM: Score-Weighted Visual Explanations for Convolutional Neural Networks. CVPR 2020
- [6] A Kumar, Disentangling 3D Pose in A Dendritic CNN for Unconstrained 2D Face Alignment. CVPR 2018

补充说明

Github repository 中有四个分支: main, v1-minivgg, v2, v3-rb。main分支可以不看, v1是使用了缩小版的vgg, v2使用了ECA-Net, v3使用了Reborn-Net。