

2022.03.20周报

纪新龙

本周计划任务

等待Faceswap网络训练完成。
做重光照。

具体完成情况

大部分时间花在优化模型训练效果上，重光照没怎么看。

下周计划任务

下周一定看重光照，不在模型训练上多耗时了。

具体完成情况

原本我选择的是faceswap提供的众多模型中的默认模型“original”，但是累计训练约30+小时后发现loss可以降到0.01以下但是换脸效果不好。所谓的换脸效果不好，一方面指的是A换成的B感觉上不那么像B，另一方面是仍然模糊。这时候我尝试勾选了 *warp tp landmarks*，似乎训练一段时间后A和B区分度更大了。

总结这周在训练中学到的东西：

- 模型：尝试中感受到Dfaker的模型换脸的结果中，A和B更有区分度。所以模型换成了Dfaker。
- 硬件：分别在三处机器跑过模型，对云计算的租用、依赖环境搭建、遇到问题及处理方法积累了经验。
 - 远程登陆、守护进程、传文件。
 - 与github的两种连接方式。
 - 查看显卡、CPU、内存、I/O的情况综合判断当前哪部分硬件是模型速度的瓶颈，合理利用

算力。

- 训练方法：

- 调整batch来充分利用硬件设施和促进训练
- 对于faceswap，用Dfaker模型、选择“warp to landmarks”换脸的区分度比original模型好。后期调整到“No warps”有望迅速降低loss，让网络学习到细节。
- faceswap的loss指的是A编码解码后还原成A的差别，所以并不是一个衡量换脸效果的绝对判断标准。模型有没有训练好还得加上肉眼主观判断。
- faceswap将一部分“对齐文件”里的信息（主要是特征点、掩膜等只关乎面部图片本身的信息）编码入了照片的png格式头部，这使得在训练的时候无需输入“对齐文件”了，所有训练时候需要的信息都存储在面部图片里的。“对齐文件”的偏重将人物照片和面部图片联系起来，比如convert时候需要。
- 从视频直接抽取面部时，假如视频中大量帧都没有人脸，这时候除了之前抽取完排序清洗的方法，还可以直接用照片过滤的方法。就是上传该人物的清晰图片作为过滤器，只抽取帧中该人物的面部图片。当阈值设的很小的时候，可以获得大量符合要求的该人物的面部图片，既没有混入别人的图片，也没有混入非人物图片。在算力允许的情况下，这样做简便了很多。

- 数据集的调整：

在训练的过程中可以小幅度调整数据集，这也是训练的一种方法。

为了得到更好的结果，我注意到视频的质量一般比照片更高，而且有人脸的帧数很密集。所以我再次清洗了数据，在原先基础上引入两个古天乐年轻时候影视作品的混剪视频、一个安倍晋三的演讲。

由于原先古天乐数据集包含年轻时候和年老时候的古天乐，样貌特征差别巨大，我认为这种B人物照片差别大的情况会影响到模型的训练。所以我彻底抛弃原古天乐数据集，将人物B限定为“年轻时候的古天乐”，并用收集的两个混剪视频仔细制作了新古天乐数据集。

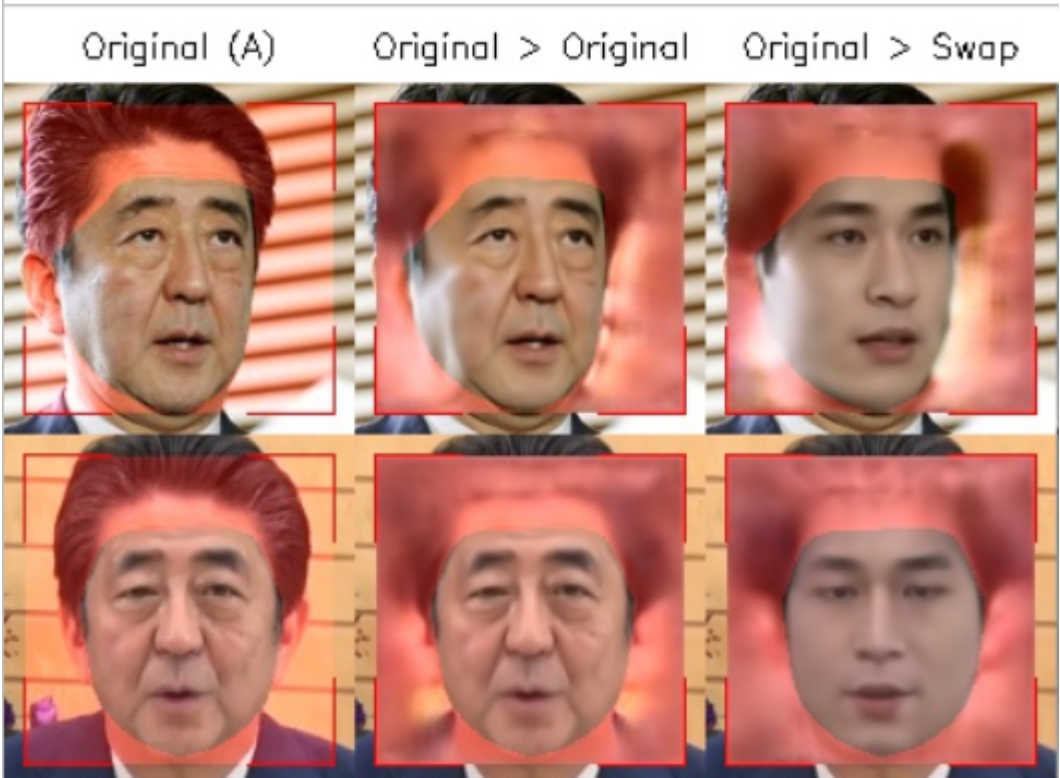
而原安倍晋三数据集质量要比原古天乐数据集质量更高一点，包含了各角度、各光照的情景，且从缩略图上看清晰度够的。前期我没有改安倍晋三的数据集。后来用了autodl算力跟上之后，loss降低变得非常缓慢，可能是到了训练的瓶颈了，我清洗了原安倍数据集里不清晰的、有遮挡的约200张，然后用安倍晋三演讲视频重新收集500+图片补充进去，提高安倍晋三数据集质量。

安倍数据集升级之后，原先用RTX3090训练了3个小时都没什么变化的A、B的loss都发生了降低，且A的loss一反常态，首次降低到比B还低一点。

此时训练效果如下：

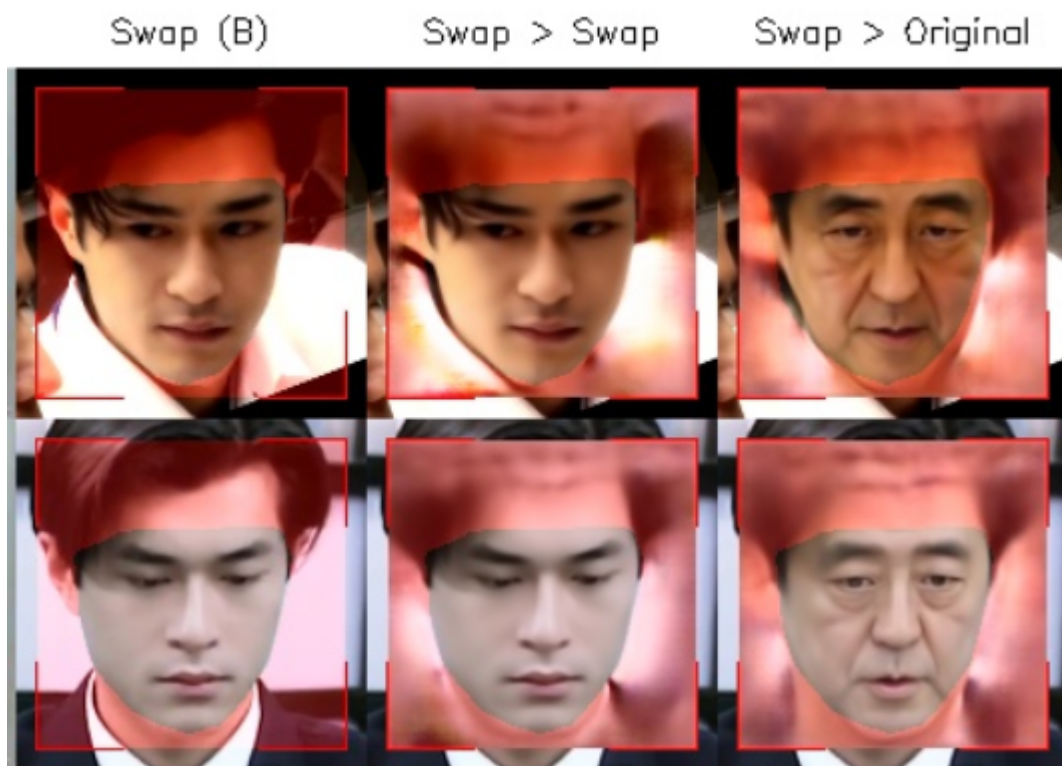
A换成B





B换成A





一个问题：安倍的额头有皱纹没有替换掉

这是遮罩（Mask）选择的问题。我之前尝试过extended遮罩和unet-dfl遮罩，这两种都是被推荐的遮罩，可是都没有留出额头。训练和转换的时候都会用到遮罩。训练时用遮罩时为了让神经网络集中精力学习某块区域，转换时的遮罩决定那些区域被替换。

所以我尝试在转换的时候提供一种新的能够显露额头的遮罩“bisenet-fp”，来把额头也替换掉：



能明显看到采用新遮罩转换后安倍额头的没有了。

bisenet-fp遮罩固然能够解决安倍额头的问题，但是毕竟比之前用的遮罩多覆盖了额头区域，而额头是个很危险的地方，因为它毗邻头发。假如遮罩稍微不准确，都会引入一些头发的黑色到额头上。还有就是，这种更复杂的遮罩算法没有之前两种更稳定。对比古天乐换安倍的视频分别采用bisenet-fp和unet-dfl的效果，明显能看到有些连续的帧间在古天乐的发迹线处出现了遮罩范围的跳跃。也就是说，对于“古天乐换安倍”这种本来就没有“额头皱纹难题”的替换，引入更细致的bisenet-fp遮罩在替换时没有改进，反而导致了帧间遮罩区域跳跃，是一种退步。

我总结视频换脸中的“额头皱纹问题”的解决方案如下：

1. 优先使用不覆盖额头的遮罩，例如extended遮罩和unet-dfl遮罩。这些遮罩更稳定，不会出现帧间不连续导致的失真。（例如古天乐换安倍，换完之后效果是不错的，没有出现额头皱纹导致的虚假）
2. 假如用extended遮罩和unet-dfl遮罩等遮罩后，替换后的视频中额头皱纹成为很显眼的瑕疵，这时候需要：
 - 1) 依然使用 extended类型遮罩，将替换后的视频再做模糊或磨皮处理（假如允许的话），这样可以掩盖额头的皱纹。
 - 2) 使用bisenet-fp类遮罩，为了防止帧间发迹线跳跃，最好用faceswap的手动调整工具修改一些问题帧的遮罩范围，使得帧间一致。这种，动手量也不会很大，因为大部分帧是好的。
3. 还有一种考虑就是，直接在bisenet-fp遮罩的数据集上训练出的模型或许会更好的替换bisenet-fp遮罩下的视频。

使用了bisenet-fp_face遮罩的Dfaker模型在刚开始训练时效果如下（新模型，尚模糊，截屏时loss在0.03左右，还未成形）：

