

- 14周：
  - 看懂并小组交流transformer的原理，和vit的结构（草稿纸和笔记）
  - 学习交流了jupyter，google colab的使用等等
  - 简单交流了一些tnt等变体
- 15周：根据vit就先写出了vision-transformer的代码，并在mnist数据集上跑，准确率最高的是97%
  - 普通的CNN都有99%，就觉的有点低
- 16周：改造了一下vit模型的部分代码，使它在raf-db数据集上跑
  - 但可能改的有问题，准确率只有38%
- 17周：小组成员集中开始尝试各种vit变体
  - reformer
    - 训练vit的成本太高，为了提高效率，reformer的思路是减轻原版vit计算的成本
    - 没有预训练好的模型，自己预训练条件不足，所以先暂时搁置
  - lite-transformer：代码量过于大和复杂，先搁置
  - swin-transformer：跑通了模型，并且利用imagenet预训练好的swin-T和swin-B在raf-db上跑，准确率38.6%，十分疑惑
  - cvt：利用预训练的模型在raf-db上跑通了，准确率最高的是77%
- 反思和疑惑：
  - **准确率低的问题：**

可能有如下原因：模型太大了泛化能力不强，数据集不好（数据量太小）；代码本身有问题等等

如何提高准确率？
  - **预训练的问题：**

我们的机器可能不能支撑预训练，一个imagenet就有100多个g

用人脸和不是人脸的预训练思路可不可以，只有人脸标签的数据集应该比较好找
  - **对于大创进行疑惑的地方：**

有没有现成的vit相关的模型代码或者资源可以参考？现状是改进谈不上，连一般的准确率不知道为什么都达不到

目前不太确定transformer是否真的适合fer，简单的与训练好的resnet18跑出来都有百分之80多的准确率

目前试过的结果让我们摸不着头脑，所以对后续vit的研究，对后面的时间任务安排，不是很清晰，寒假应该做哪些努力？中期五月份之前要有怎样的目标？