

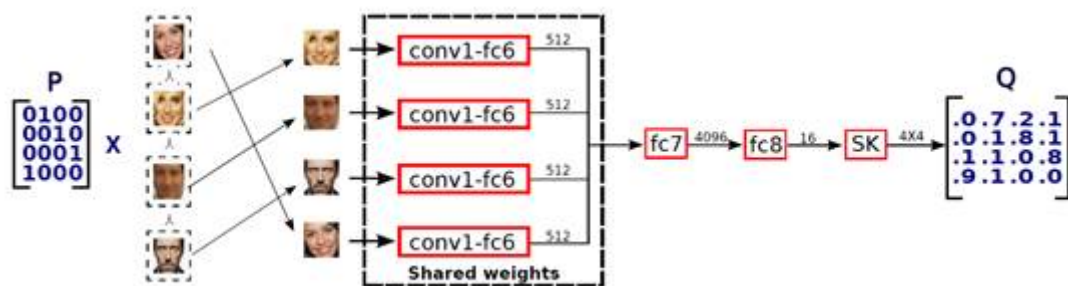
AI3607 深度学习 大作业

本次深度学习课程大作业基于“视觉”与“图”两大深度学习的重点研究方向，对于一个视觉问题中的应用（拼图问题），在参考已有论文的基础上，提出自己的解决方案

目标

在 CIFAR-10 数据集上，我们将在图像分类任务以外研究图片拼接这一任务，即对于一张图片，我们将其切分为若干张子图片，例如水平竖直分为四份。接着，我们将这几张子图片打乱顺序作为输入，目标就是设计一个模型可以根据这些乱序的子图片还原出他们本来的排列位置。

相关参考文献：DeepPermNet: Visual Permutation Learning (<https://arxiv.org/abs/1704.02729>)



任务步骤

这里只是推荐的实现步骤，如果有不同的设计也可以自行实现

1. 修改作业二/作业三中的图片分类的代码框架，将输入图片分割为 N 张子图片，打乱它们的顺序作为输入，并将他们的正确排列作为 label，输出格式采用排列阵 P ，其中第 i 行第 j 列为 1 则代表第 i 个子图片应该排在第 j 个位置。注意，每行与每列均最多有一个元素为 1，具体切分方式可以自行定义
2. 设计一个针对图片拼接的神经网络模型，可以参考 DeepPermNet 的设计，即将每一个子图片输入一个 CNN 进行特征提取，将得到的特征进行拼接后通过若干全连接网络得到一个 $N \times N$ 的矩阵，使用 Sinkhorn 算法使其变成双随机矩阵 Q （即每行每列的和固定为 1）作为神经网络的预测值。最后将预测值 Q （双随机矩阵）与真值 P （排列阵）作为损失函数的输入，计算 Loss 后用于更新神经网络
3. 神经网络可以套用作业二/作业三中的代码，也可以自行设计新的网络。Sinkhorn 算法推荐使用 pygtools 中的函数

本作业侧重探索性质，对精度不做要求，没有固定的解决方案，言之有理即可