

使用PCA对人脸图片进行处理

此部分作业要求使用PCA技术对给定的人脸数据集进行处理，你需要在该文件中完成你的代码，并输出结果。

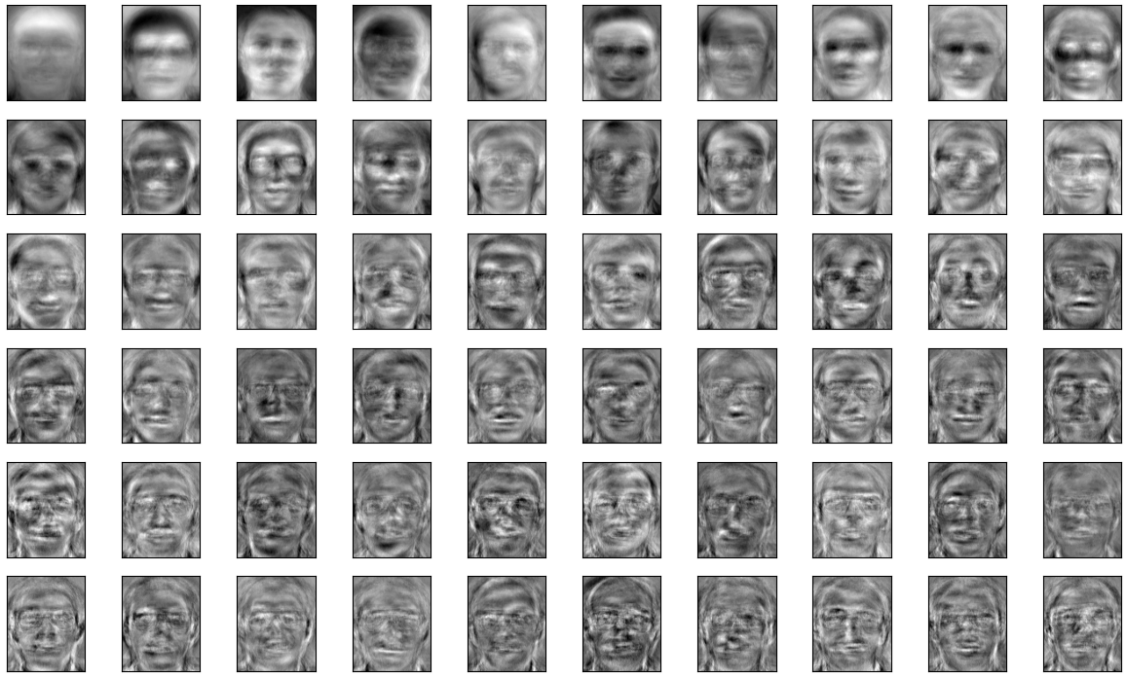
本次使用到的人脸数据集是 ORL人脸数据集，共包含40个不同人的400张图像。此数据集下包含40个目录，每个目录下有10张图像，每个目录表示一个不同的人。所有的图像是以PGM格式存储，灰度图，图像大小宽度为92，高度为112。

数据集中部分图像示例：



在本次作业中，你需要**按顺序**完成以下内容并且在该文件中**保留要求**的输出结果：

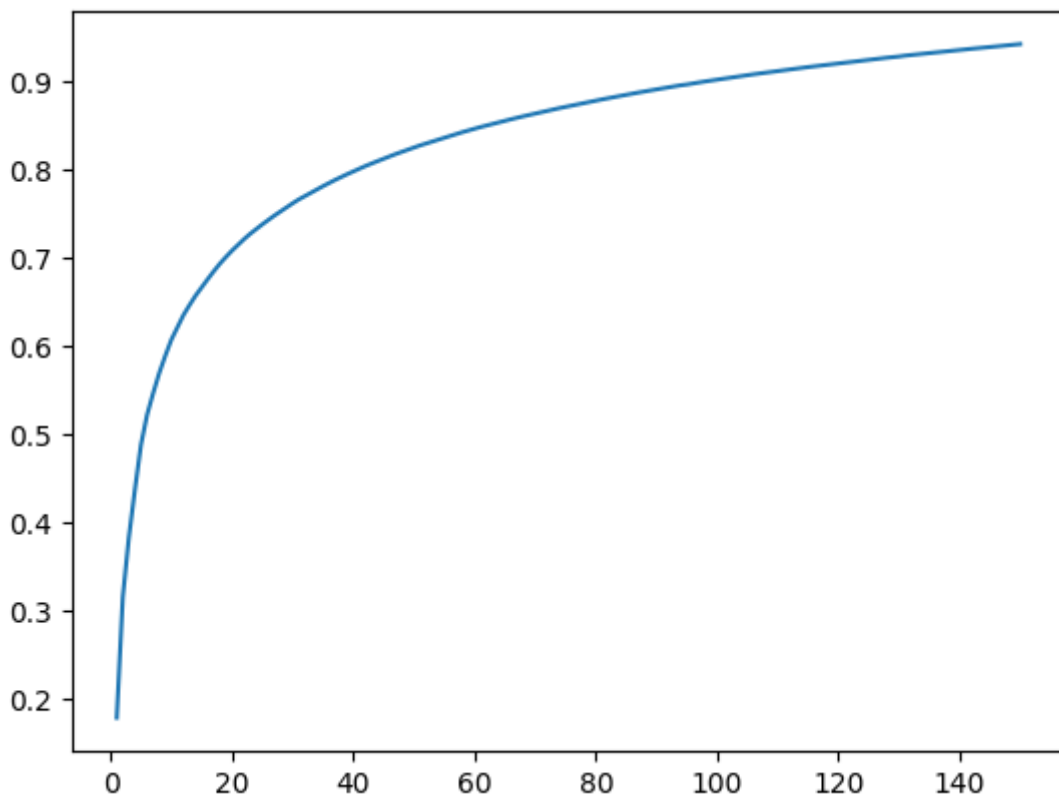
1. 将数据集划分为80%的训练集，20%的测试集，在训练集上使用PCA将特征维度降为100，即得到100个特征和其对应的特征向量，并使用训练得到的PCA将测试集维度也压缩到100，**输出：压缩后的训练集维度和测试集维度、经过PCA得到的特征向量维度。**
2. 得到100个特征向量后，**使用这些向量重建特征脸并输出。** 示例：



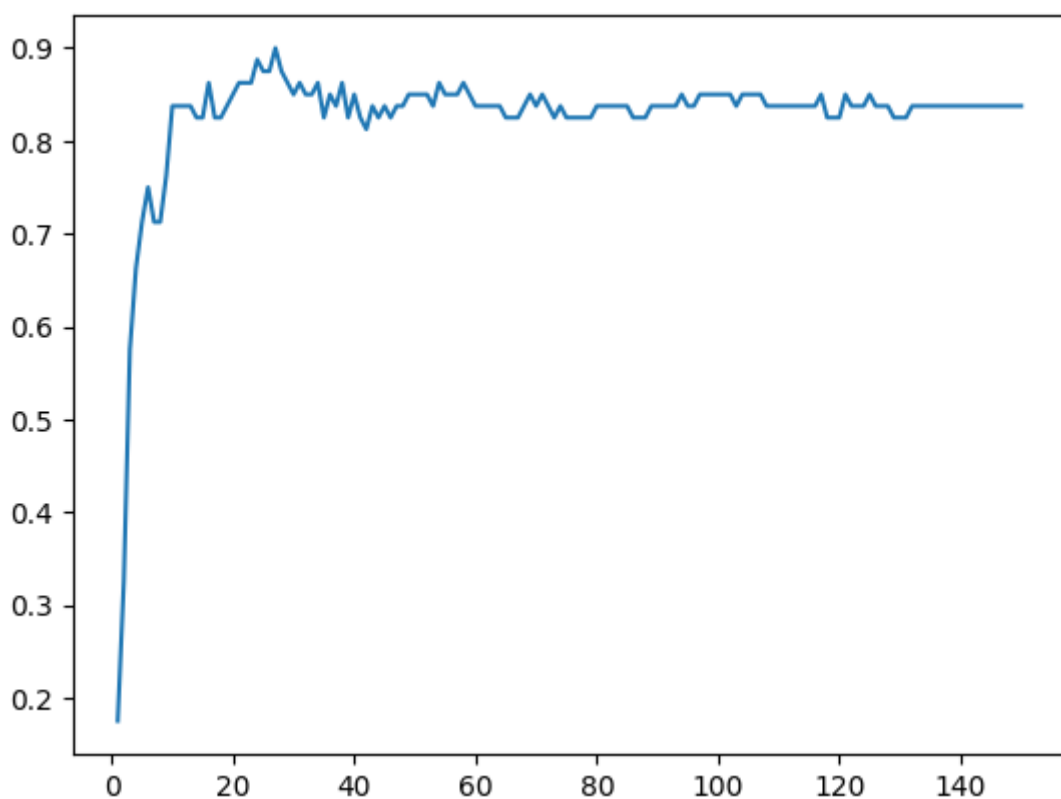
3. 使用这100个特征和对应的特征向量，自行在训练集和测试集中**分别**选择5张人脸图片进行人脸重建并输出对比图。示例：



4. **输出降维后每个新特征向量所占的信息量占原始数据总信息量的百分比，以及所有返回特征所携带的信息量总和是原始数据的多少；**
5. **画出特征个数和所携带信息数的曲线图，此处特征数上限设置为150，此处以150为示例：**



6. 保留特征数为150，得到的训练集每次保留前 ($n=n+1$, n 初值为1) 个特征，分别训练一个KNN分类器，KNN分类器的参数 $n_neighbors$ 为3，观察并验证测试集在使用其前 n 个特征时KNN分类的准确率。**注意：此处仅进行一次PCA**，输出：不同的特征保留数和准确率的曲线图。此处以 $n_neighbors=5$ 为例：



本次作业严禁抄袭，一经发现严肃处理

附加题

若感兴趣，可以自行拍摄自己或舍友的人脸照片、或者其他较为简单且特征明显的数据集进行PCA操作。