###### 实验九 蓝藻的光谱分析和图像识别

###### 实验报告

1. 实验目的和要求

体验机器学习：无监督算法和有监督算法。

1. 实验原理

现代科学发展的一个主要趋势是学科之间的交叉与融合，学科交叉促使各个学科向其他领域渗透、发展，成为现代科学研究的创新源泉。蓝藻可进行光合作用，细胞内含有大量的色素分子如叶绿素和藻胆素等以吸收光能，这些色素分子可提供丰富的光谱信号，因此利用常用的光谱学分析方法如UV-Vis、三维荧光、拉曼及红外光谱，可得到不同种类蓝藻的“指纹光谱”，结合蓝藻重要的形态学特征，可对蓝藻进行快速鉴定。

此外，对该方法进行一定程度的拓展后，可对混种蓝藻在未经纯种分离和培养的条件下进行快速鉴定，突破传统蓝藻的鉴定方法。

1. 无监督学习、监督学习

1.无监督学习：直接从数据中学习某些模式

2.监督学习：从数据中学习提前指定的模式

1. PCA：主成分分析——无监督学习算法

PCA是最常见、最简单的一类无监督学习算法，可以对数据进行降维，本质是基变换。

1. PCA的原理：

多维数据可视作线性空间中的一个向量

例：数据（0.1, 0.5, 0.2, 0.4）是R4中的一个向量

PCA的目标：在这个线性空间中寻找一组有如下特点的正交基

（1）这组基的第一个向量的方向就是使得所有数据在上面投影的方差最大的那个方向

（2）这组基的第n至最后一个向量张成线性空间V’，所有数据点投射到V’后组成点集P’，第n个向量就是所有V’中向量里使得P’在其上的投影方差最大的那个向量

这组基中的每个向量都被称为一个主成分

2.PCA的用途：去除数据多余的维数，将最明显的特征放于更低的维度上。

（1）降维

（2）降噪

（3）复杂算法的预处理步骤，如聚类的预处理

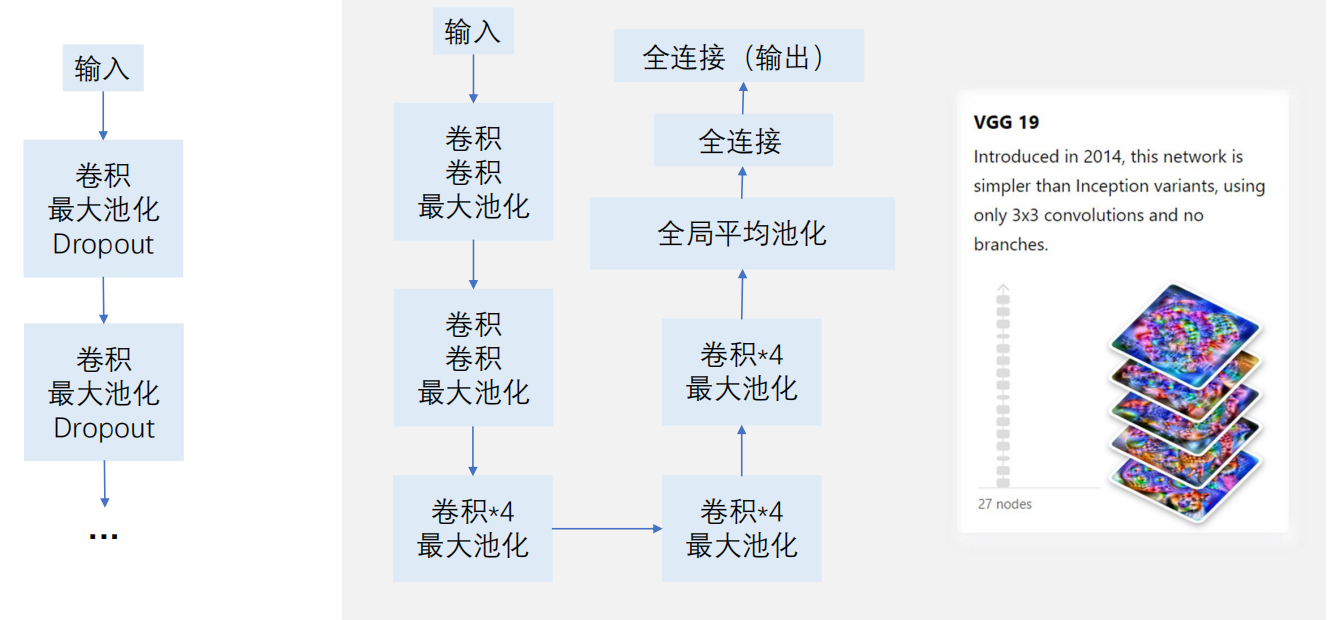
（4）提取特征

（5）体现数据的高维特性

本次实验中利用PCA进行蓝藻的光谱分析

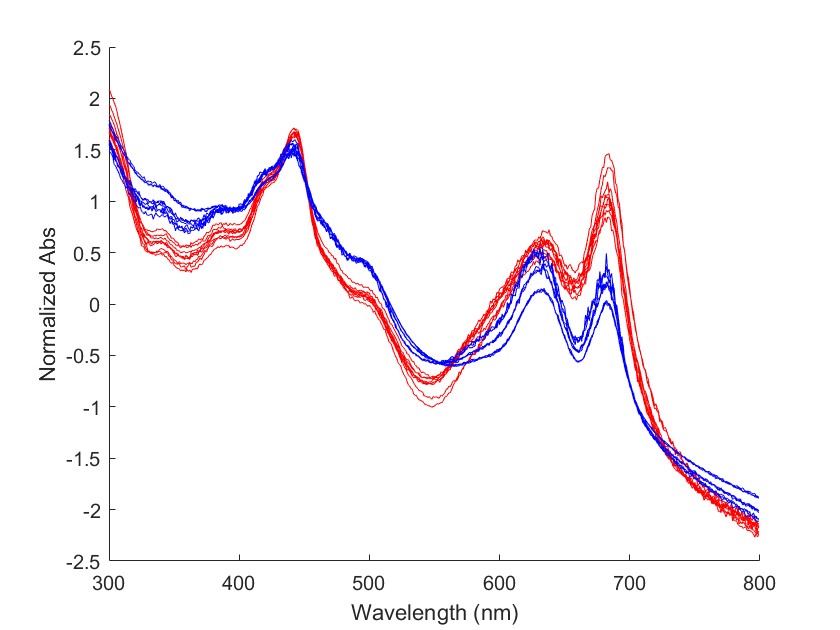
1. CNN：卷积神经网络——监督学习算法

CNN是比较常见的监督学习算法，可用于图像分类。

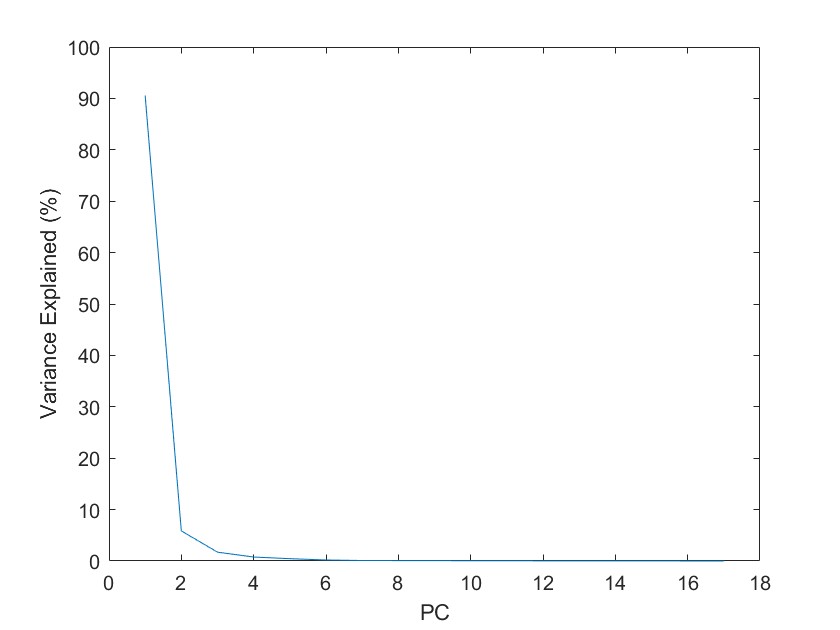
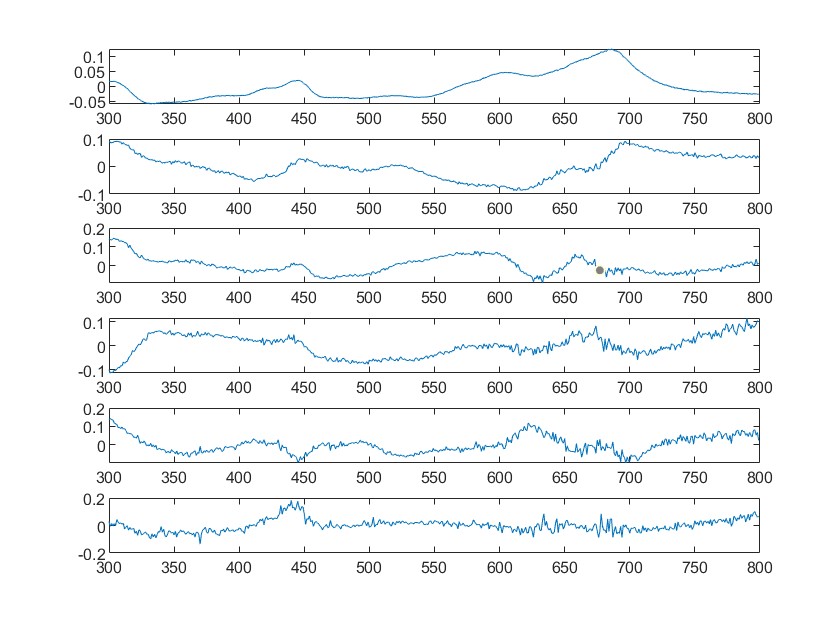
典型的卷积神经网络结构：  


本次实验中利用CNN进行蓝藻图像识别。

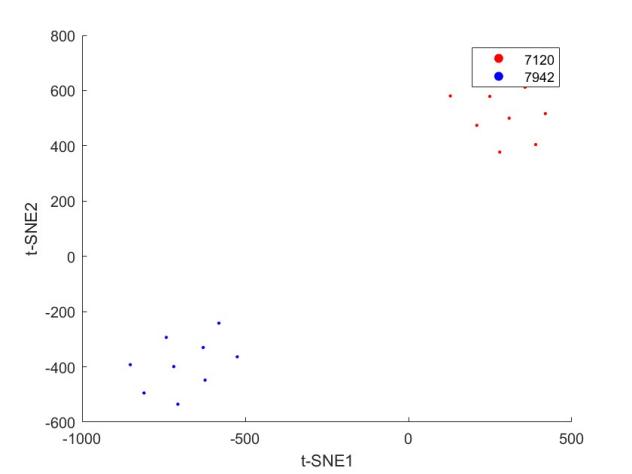
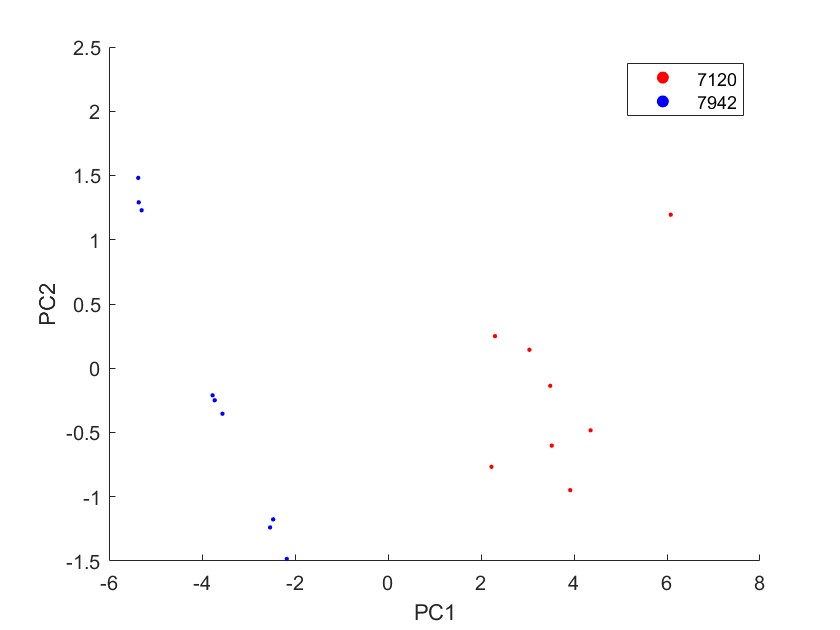
1. 实验步骤
2. 蓝藻视频拍摄及UV-Vis(240-800nm)光谱的测定；
3. 蓝藻的UV-Vis光谱PCA分析；
4. 利用CNN进行蓝藻图像识别。
5. 实验结果
6. 蓝藻的UV-Vis光谱分析



图一 两种蓝藻的光谱图

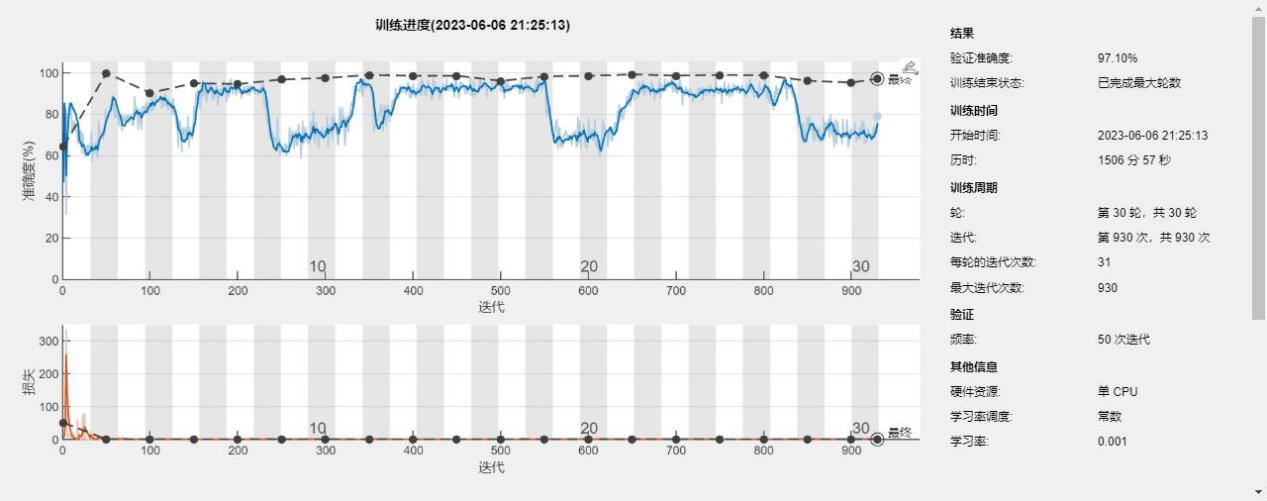


图二 PCA前六种主成分 图三 PCA主成分解释方差的情况



图四 两种蓝藻光谱数据的PCA分析 图五 两种蓝藻光谱数据的t-SNE分析

1. 蓝藻图像识别



图六 蓝藻图像识别CNN人工智能训练图