

色彩规律可视化：

基于 k -Means聚类算法

生成敦煌壁画色彩网络模型

二零二四年·一月拾日

小组成员：蒋星茹、张嘉怡、倪好、罗钰薇、葛瀚泽、黄泽君



样本采集

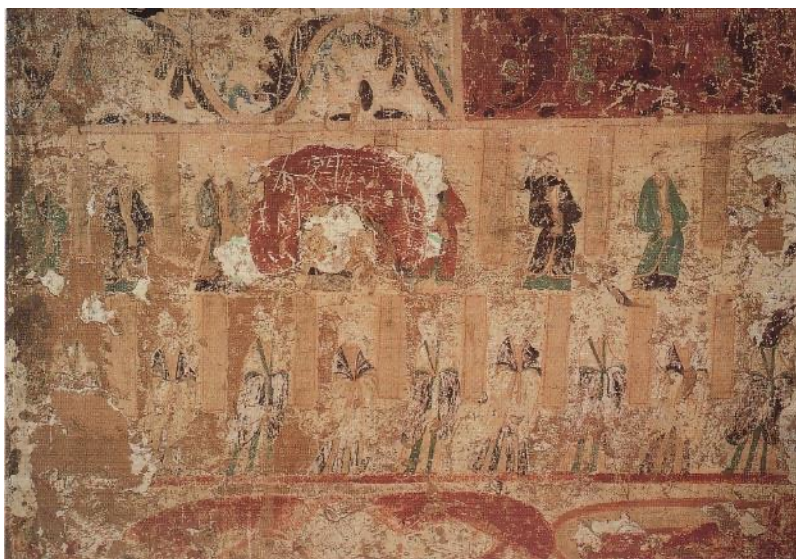
sample collection

共收集原始敦煌壁画实拍图片资料2495张，图片来源包括：吉美博物馆馆藏、大英博物馆馆藏、敦煌莫高窟-009窟、敦煌莫高窟-014窟、敦煌莫高窟-045窟、敦煌莫高窟-085窟、敦煌莫高窟-112窟、敦煌莫高窟-154窟、敦煌莫高窟-290窟、敦煌莫高窟-420窟、敦煌莫高窟-428窟、敦煌莫高窟-安西榆林窟、敦煌莫高窟-克孜尔以及部分壁画扫描稿

样本筛选

Sample Screening

对收集到的2495样本数据进行筛选，去除：1. 风化侵蚀严重的样本 2. 雕像、书法、线稿等不相干样本 3. 颜色单一的样本，如黑白色样本，保留1419张样本



图片预处理

Image Preprocessing

利用photoshop对保留的样本进行预处理，去除图像的背景部分，统一保存为png格式
利用Topaz DeNoise AI对保留的样本进行中值滤波去噪



K-Means 聚类

K-MEANS clustering
1. 将k的取值设定为8，对每张样本进行第一次聚类，生成1419张饼图（这里修改代码没用肘部法是因为样本量过大，慢）

2. 将k的取值设定为8，将1419张饼图最终聚类为一张饼图

Choose Images

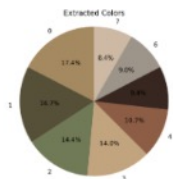
Drag and drop files here
Limit 200MB per file

Browse files

未命名作品.png 6.6MB
未命名作品 127.png 2.8MB
未命名作品 126.png 2.3MB

Showing page 1 of 473

Running `extract_colors(...)`



一次聚类

Choose Pie Charts

Drag and drop files here
Limit 200MB per file

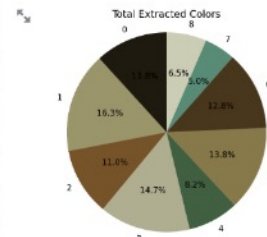
Browse files

pie_18.png 27.8KB
pie_17.png 29.8KB
pie_16.png 29.5KB

Showing page 23 of 24

Enter the no. of colors to be extracted in the Total Pie Chart

9



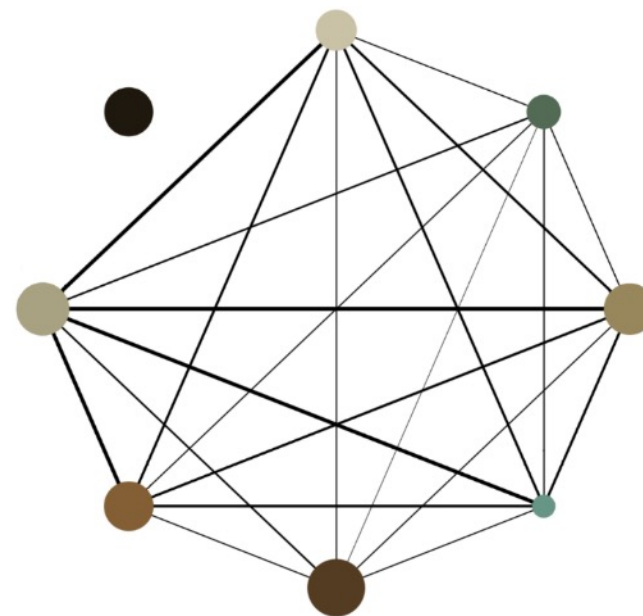
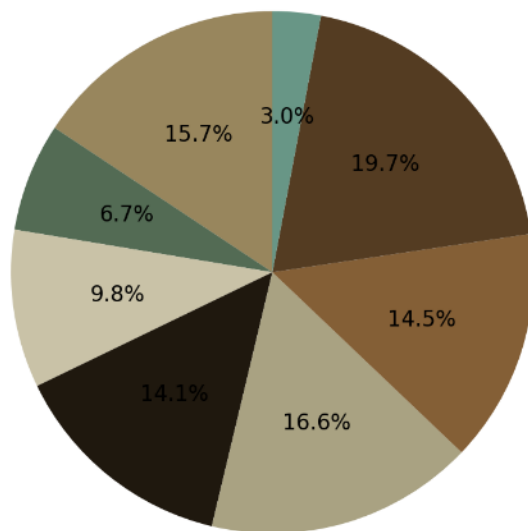
二次聚类

生成色彩网络模型

Generating Color Network Models

根据第一次聚类生成的饼图建立色彩网络模型，提取前八种出现频率最高的颜色作为色彩网络节点，连线则表示被连接的颜色以一定频率多次共现而产生搭配关系，连线越粗表示共现频率越大

Total Extracted Colors



K-Means算法与其它算法的对比

Comparison of K-Means algorithm with other algorithms

Extracted Colors

Input Images

Enter the Max K of K-means

20

Choose Images



Drag and drop files here

Limit 200MB per file

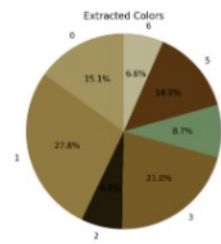
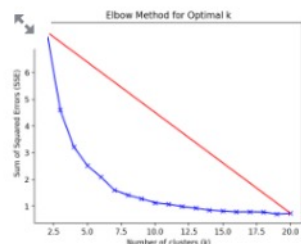
Browse files



DSC_1927.jpg 1.3MB



Image 1



Download All Pie Charts

DBSCAN (Density-Based Spatial Clustering of Applications with Noise):

优于K-Means:

它不需要事先指定簇的数量，这一点比K-Means更有优势

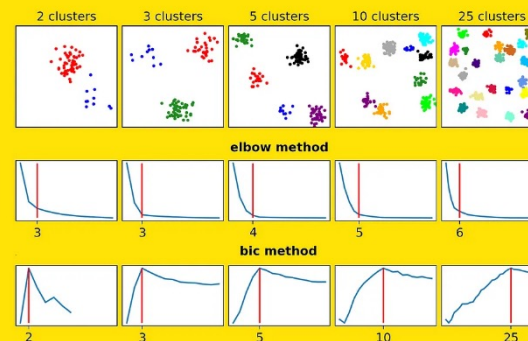
劣于K-Means:

1. K-Means比DBSCAN快，计算效率高，尤其是对于大型数据集。这是因为K-Means的复杂度与数据点的数量呈线性关系，而DBSCAN的计算成本会随着数据点的数量增加而增长
2. K-Means算法易于实现

为什么不采用bic method确定K值

Why not use the bic method to determine the K-value

对我们的数据而言过拟合



创新点及应用场景

Innovation points and application scenarios

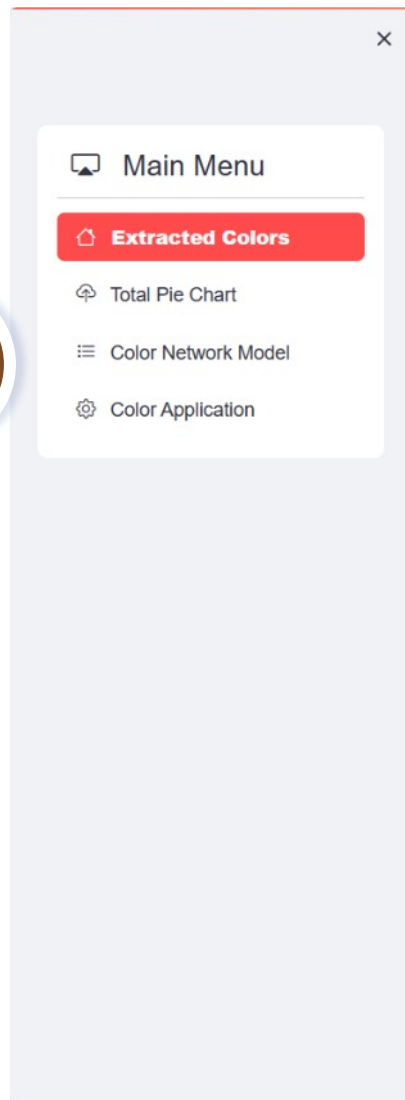


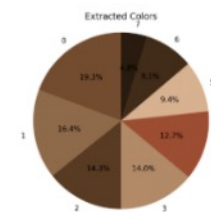
Image 1419

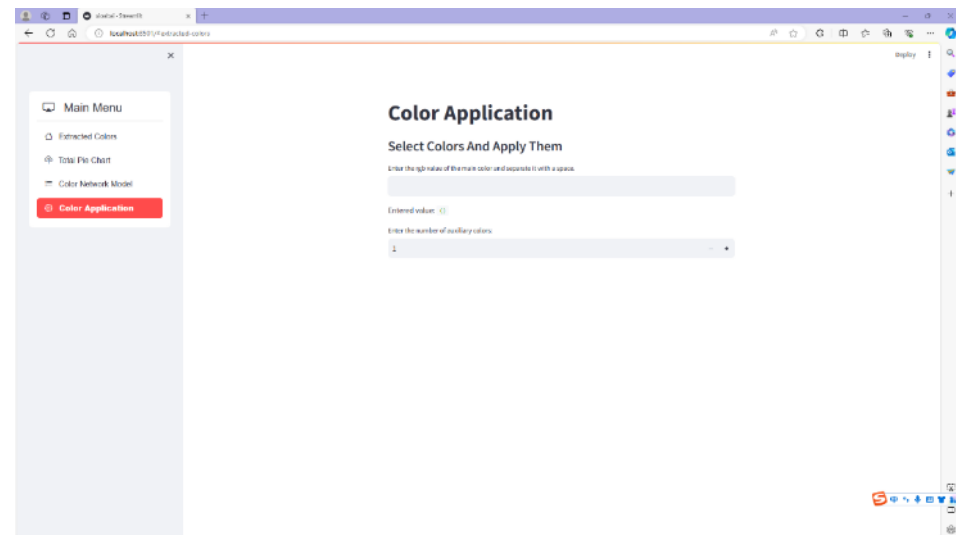


1. 基于Streamlit库做了可交互的Web界面，可在Streamlit官网上在线运行也可以在本地图像服务器运行，即使是不懂代码的设计师也可以提取某一风格的色彩网络模型，降低了使用、学习成本

2. 每一功能部分分开：

1. 将图片导入生成饼图，肘部法自动获取最优K值
2. 将饼图聚类生成总饼图，用户可填写需要的饼图色彩数量，显示每一色彩在饼图中所占比例
3. 根据聚类结果生成色彩网络模型，用户可手动填写阈值来调整色彩间的关联度
4. 用户选取色彩网络模型中某一色彩为主色，填写想要的辅色数量，根据共现频率为用户提供色彩搭配，用户可上传图片，将图片颜色替换为用户选择的色彩搭配（优化中）

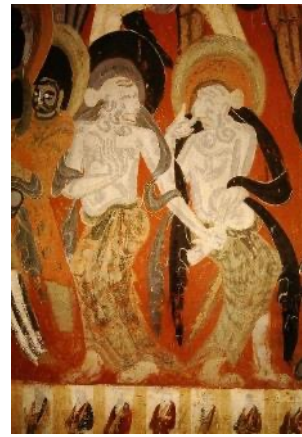
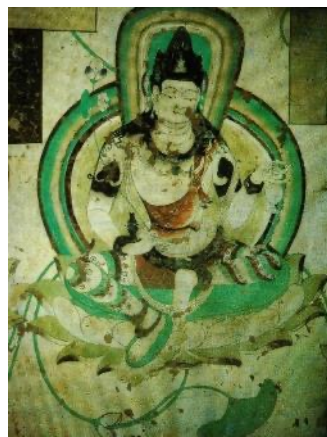
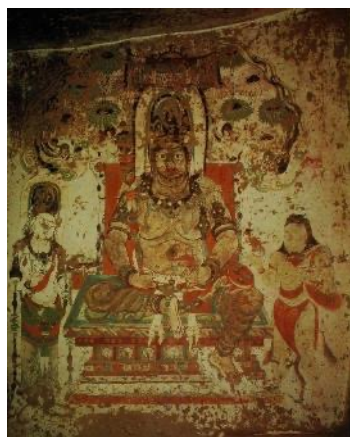




不足与未来改进方向

Shortcomings and future directions for improvement

1. 图片样本因采集时的时间、角度、光照、设备都不同，不统一，故得出的色彩网络模型颜色较灰，具有偏差
2. 莫高窟不同编号的窟壁画风格不同，本次实验将所有窟的样本统一生成1个色彩网络模型，未来可以对每一个编号的窟单独生成色彩网络模型



3. 可以对算法进行优化、替换
4. 完善为图片替换色彩的功能
5. 可以把中值滤波代码也加上，就不需要人工处理图像
6. 可通过对各民族传统服饰及各年代代表服饰（例如唐代宫廷风格等）进行数据采集，聚类生成色彩网络模型，直接为设计师提供色彩搭配方案, 对中华优秀传统文化进行传承和创新

小组工作分工

Division of work of the Group

实验设计：蒋星茹、葛瀚泽、张嘉怡、罗钰薇、倪好、黄泽君（外援）

代码：黄泽君（外援）

样本采集：蒋星茹

样本筛选：蒋星茹、葛瀚泽、张嘉怡、罗钰薇、倪好（每人500张）

图片预处理：蒋星茹、葛瀚泽、张嘉怡、罗钰薇、倪好（每人284张）

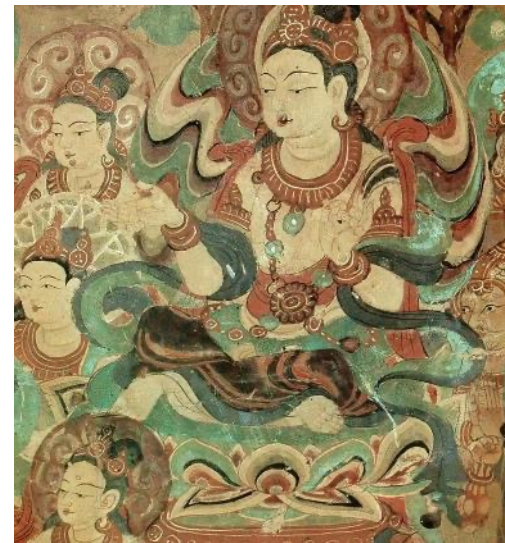
第一次聚类：蒋星茹、葛瀚泽、张嘉怡、罗钰薇、倪好（每人284张）

最终聚类、生成色彩网络模型：葛瀚泽

已有案例和技术（文献）分析：蒋星茹、葛瀚泽、张嘉怡、罗钰薇、倪好

课堂思政：倪好

PPT制作：蒋星茹、葛瀚泽、张嘉怡、罗钰薇、倪好



汇报结束
感谢聆听

