

# CBIR Report

范道宇 2019013273

## 一、实验概述。

1.使用 python 完成实验，源代码为 /src/main.py，环境为 python3.8，需要额外安装的 python 包如下：  
(pip 安装即可)

```
Pillow
matplotlib
tdqm
numpy
pickle
os
```

因为 python 和 matlab 相似，都需要借助特定的解释器来运行代码，因此这里没有将其打包成 exe 程序。

2.QueryImage.txt 放在 /data 文件夹下，运行 main.py（命令为 python3 main.py），将计算的结果输出到 /result/xxbins/yy/ 文件夹下，其中 xx 为 16 或 128，分别表示 16 和 128 bins，yy 为 Bh，HI，L2，分别表示不同的距离计算方式。文件夹中的 res\_ImageName.txt 和 res\_overall.txt 含义与 PPT 中的介绍相同。

3.大致的实验思路如下：

- 首先读取图片，根据每个点的 RGB 值将其划分到不同的 bin 中，最后所有 bin 中的像素点的个数形成一个 16 维或 128 维的向量。
- 对于数据集中的每一张图片，计算它们的 bin 向量，存储成 pkl 文件，以后每次查询时不比再重新计算一遍。
- 对于每一个 query，计算它对应图片的 bin 向量，与数据集中的所有图片的 bin 向量 计算 L2, HI, Bh 距离。
- 根据距离对数据集中的图片排序，选择前 30 张图片，将其图片名和距离输出到结果文件中，计算这 30 张图片中有多少张与 query 图片属于同一类，计算所有 query 的平均准确率。

## 二、结果分析。

### 1.实验结果。

	L2	HI	Bh
16 bins	0.2975	0.3370	0.3877
128 bins	0.3778	0.4568	0.4901

### 2.距离指标。

可以看出，在 bin 的个数相同的情况下，选择不同指标对准确率的影响很大，总的来说，Bh 指标好于 HI 指标，HI 指标好于 L2 指标。其中 L2 指标不准确是很容易理解的，因为 bin 向量的某一个分量的变化对整体指标的影响是平方关系的，误差的影响被大大放大。事实上，这一点是有依据的，例如将 L2 指标更换为 L1 指标（绝对误差之和），实验结果如下：

	L2	L1
16 bins	0.2975	<b>0.3370</b>

	L2	L1
128 bins	0.3778	<b>0.4568</b>

可以看到 L1 指标明显比 L2 指标效果更好，证明了上述结论。另外，注意到很有意思的一点，即 L1 指标与 HI 指标完全相同，这实际上是因为 HI 指标和 L1 指标是等价的，虽然具体算出的距离结果不同，但排序之后的结果是相同的。

而 Bh 指标是最好的，可能是因为 Bh 指标对误差的容忍度更高，例如 P 和 Q 的每一个分量  $p_i$  和  $q_i$  最终对距离的贡献是  $\sqrt{p_i q_i}$ ，既与两个分量的相似程度有关，又与这两个分量的绝对大小有关（分量的绝对大小代表了落到这个 bin 中的像素点的个数，可以认为是对最终距离的权重），综合考虑了距离和权重两个因素，效果最好。

### 3.Bins 的个数。

从实验结果可以看出，更多的 bin 意味着更好的结果，这一点也是容易理解的，更多的 bin 相当于图片的色域被划分得更细，16 bins 的模型中红色只有 2 种，而 128 bins 的模型中红色有四种，模型对色彩的敏感性提高，自然可以取得更好的效果。为了验证这一点，增加 1024 bins 的实验（R:G:B=8:16:8）和 8192 bins 的实验（R:G:B=16:32:16），结果如下：

	L2	HI	Bh
16 bins	0.2975	0.3370	0.3877
128 bins	<b>0.3778</b>	0.4568	0.4901
1024 bins	0.3568	<b>0.4926</b>	<b>0.5012</b>
8192 bins	0.3074	0.4815	0.4901

可以看到，在 1024 bins 的实验中，以 L2 为距离指标的精度有所下降，以 HI 和 Bh 为距离指标的精度有提升，在 8192 bins 的实验中，三种指标的精度都有下降。这说明，适当地增大 bin 的个数可以提高模型对色彩的敏感性，有利于提升精度。但 bin 数量过多会使得模型对色彩过于敏感，容错能力变差，忽略了宏观的色彩特征，模型精度变差。

### 4.错误分析。

除了上面分析到的距离指标和 bin 的个数对实验结果的影响，其实使用颜色分布直方图这种方法本身可能就有一定的局限性，它完全忽略了像素点的空间位置信息，而这一点对图片的语义也许是很重要的，例如以下两张图片（分别为 beach/110 和 football/Image14）的颜色分布非常相似，但语义却截然不同，上图的蓝色是天空和大海，而下图的蓝色是足球场的草坪图案，这种对空间位置信息的忽略给这种图像检索方式带来了天生的局限性。





## 5.改进方案。

通过上面的分析，可以得出以下几种改进方案：

- 选择合适的距离指标
- 选择合适的 bin 数量
- 使用考虑空间位置的模型，例如 CNN

另外，实际上深度学习方法已经将图片分类做的极其完善，使用在 ImageNet-1k 数据集上的准确率已经接近 91%，但是，基于传统计算机视觉方法进行图像检索仍然有其意义，可以在数据集极小，没有预训练模型权重的情况下完成图像检索，这种很强的人为的归纳偏置也许会在未来发挥重要的作用。

最后，感谢老师和助教的悉心指导！