**第十四周报告**

1. **实验概览**

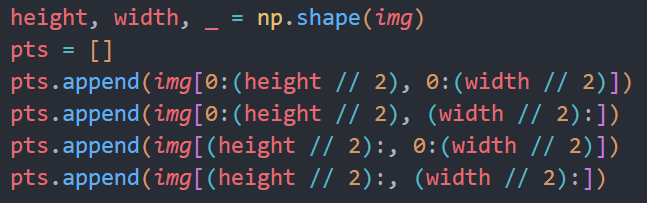
本实验旨在于学习LSH（Locality-sensitive Hashing）的基本原理，理解LSH在检索中的优势和思想，并在实验中实现对于图片的预处理、哈希函数和检索过程。

1. **实验环境**

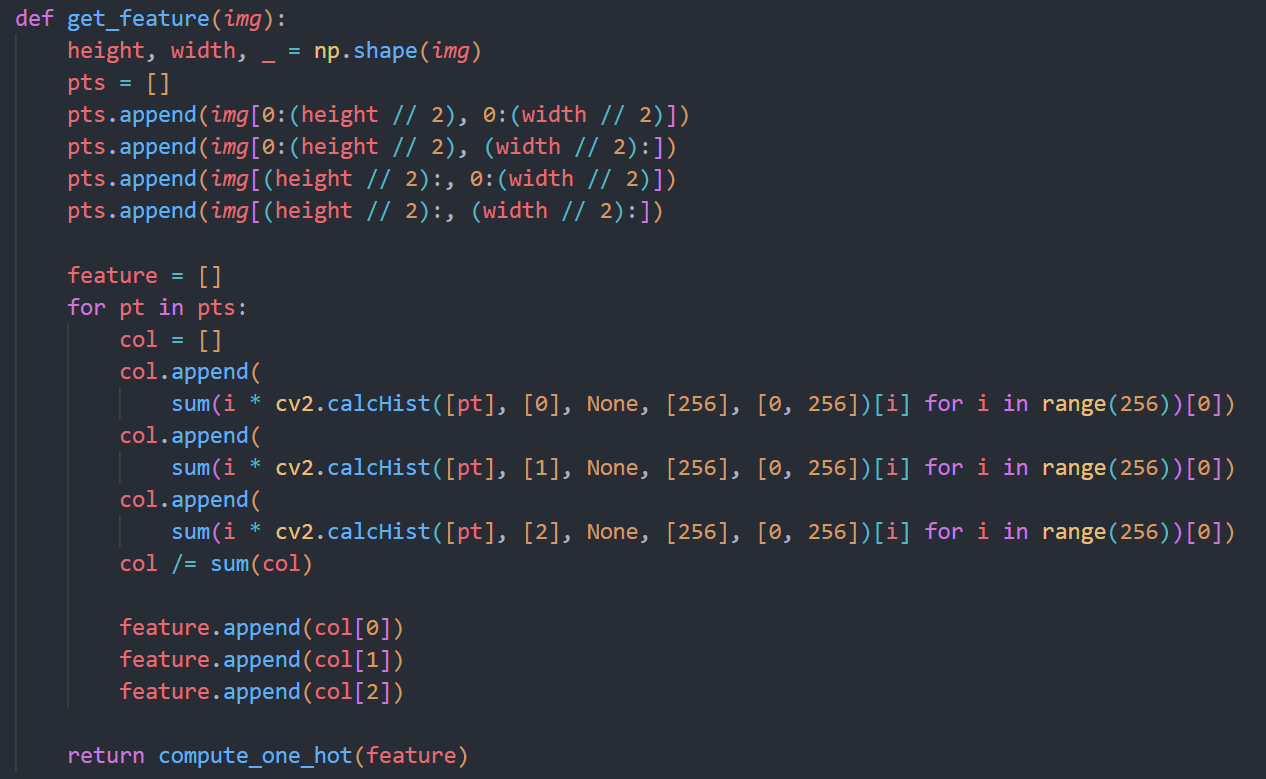
Docker: sjtumic/ee208

1. **解决思路**
   1. **图片的预处理**
      * **提取颜色直方图**

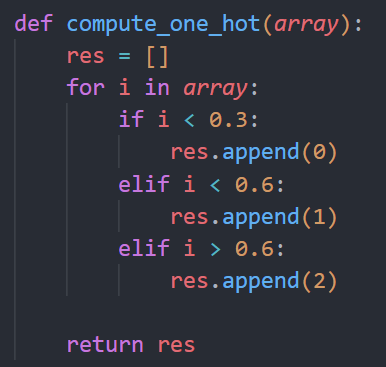
首先对图片分块处理



基于之前lab的结果，对图片进行分块提取颜色直方图拼接为12维图片特征

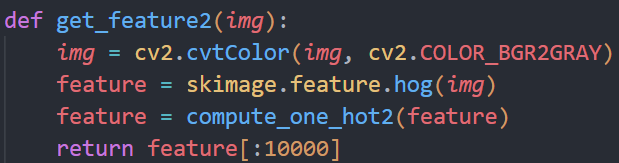


根据特征每一维特征值整数赋值得到最终特征

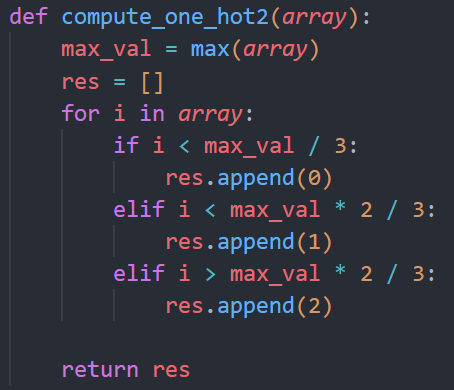


* **利用HOG算法进行特征提取**

实验中，作为补充，基于skimage库利用了hog方法对图片计算特征，得到10000位的图像编码作为特征

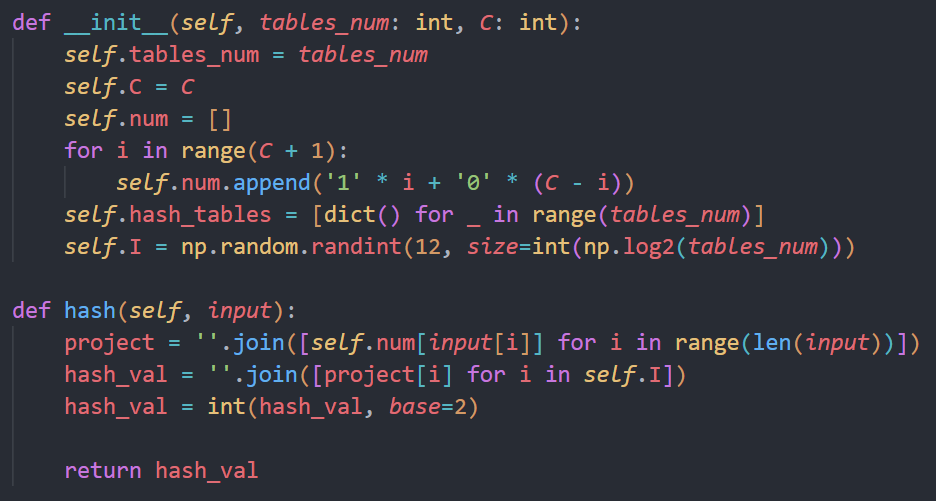


映射到整数域函数如下：

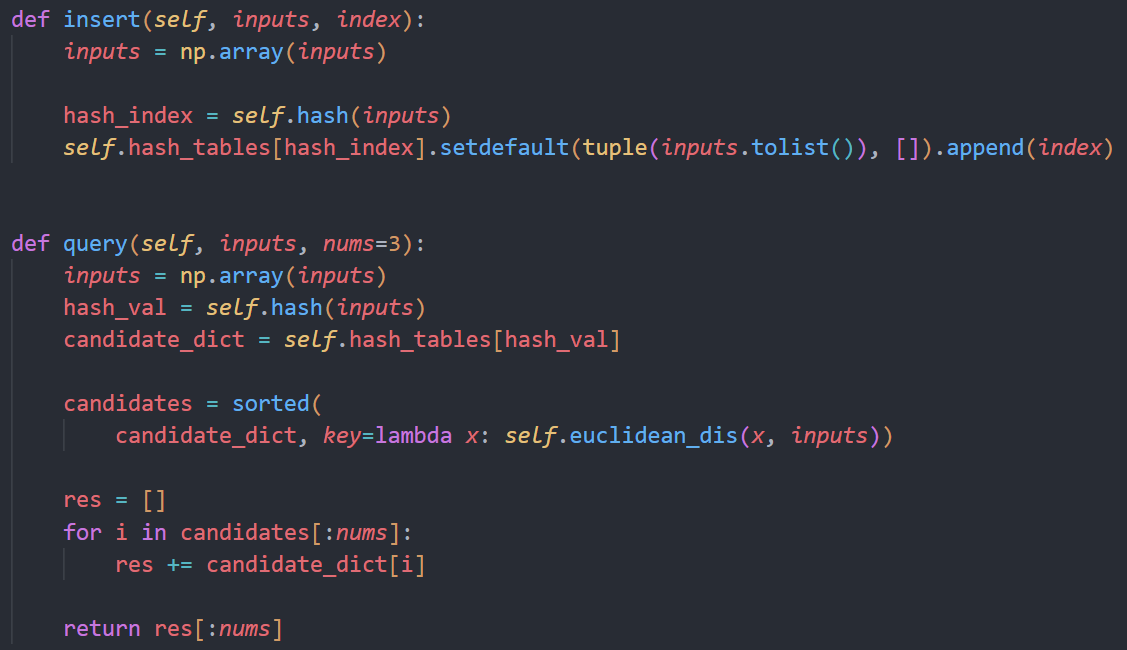


* 1. **构建LSH算法**
     + **定义hash函数**

根据参考得到hash函数，其中投影集合在每次初始化时随机取得



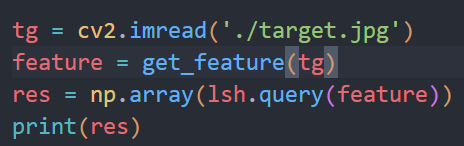
* + - **定义插入数据和查询操作**



其中，每一类初始化为字典是防止出现feature相同的情况（以颜色直方图作为特征容易发生）；利用特征之间的欧几里得距离作为定义图像相似度的指标

1. **代码运行结果**

基于颜色直方图的结果：





其中38.jpg和target.jpg图像一致



而12.jpg图像内容和target为同一类，且整体颜色分布和target相似度较高，可以发现以颜色直方图作为特征，将提取出和目标图片颜色分布类似的图片，但无法对细节进一步细化



基于HOG特征提取结果：

target.jpg



在实验中对于数据库中每一张图都进行了检索操作，整体上HOG得到的特征可以将数据库中的图片较好地区分开来