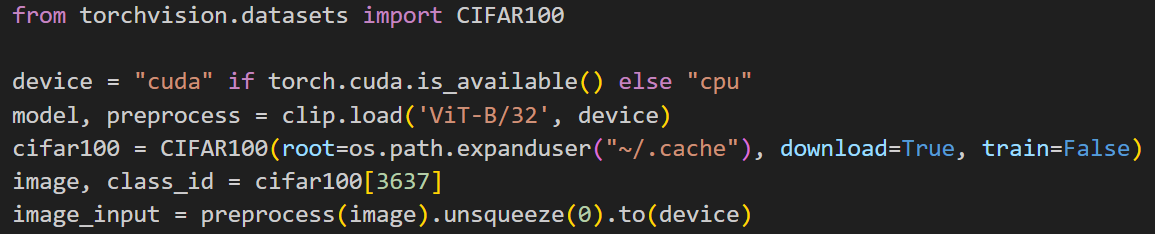
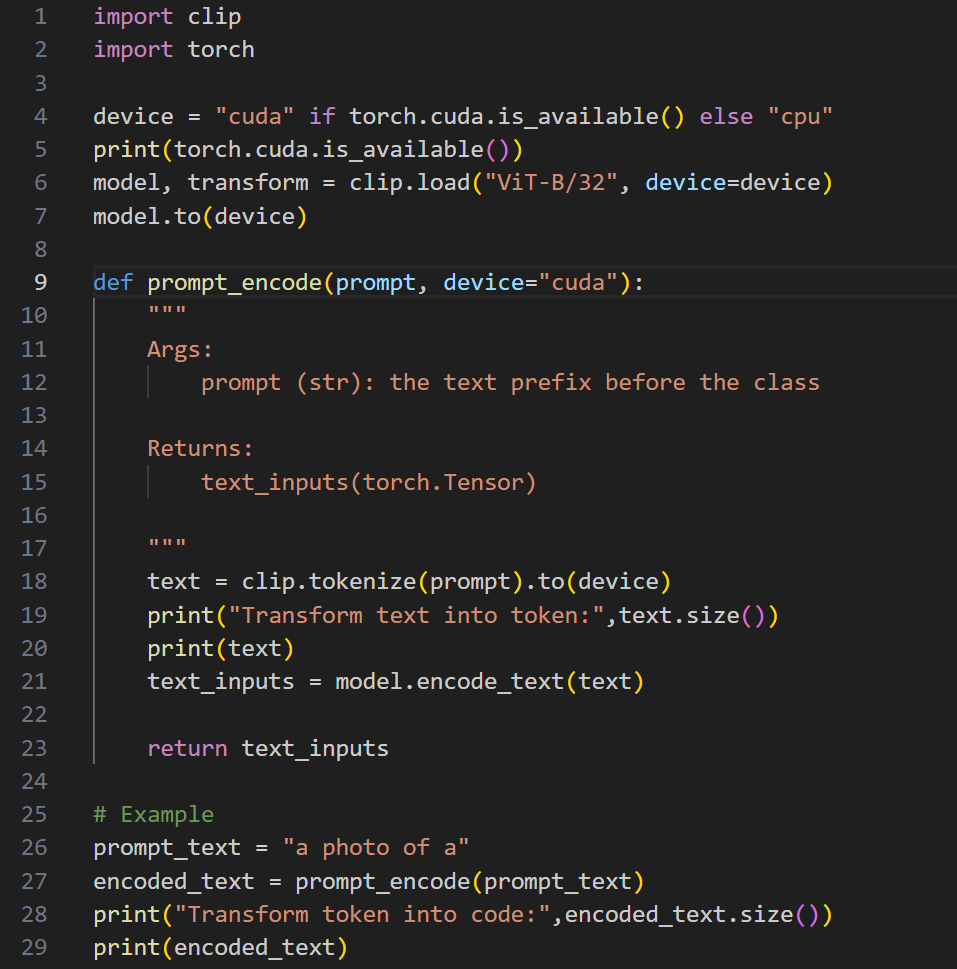
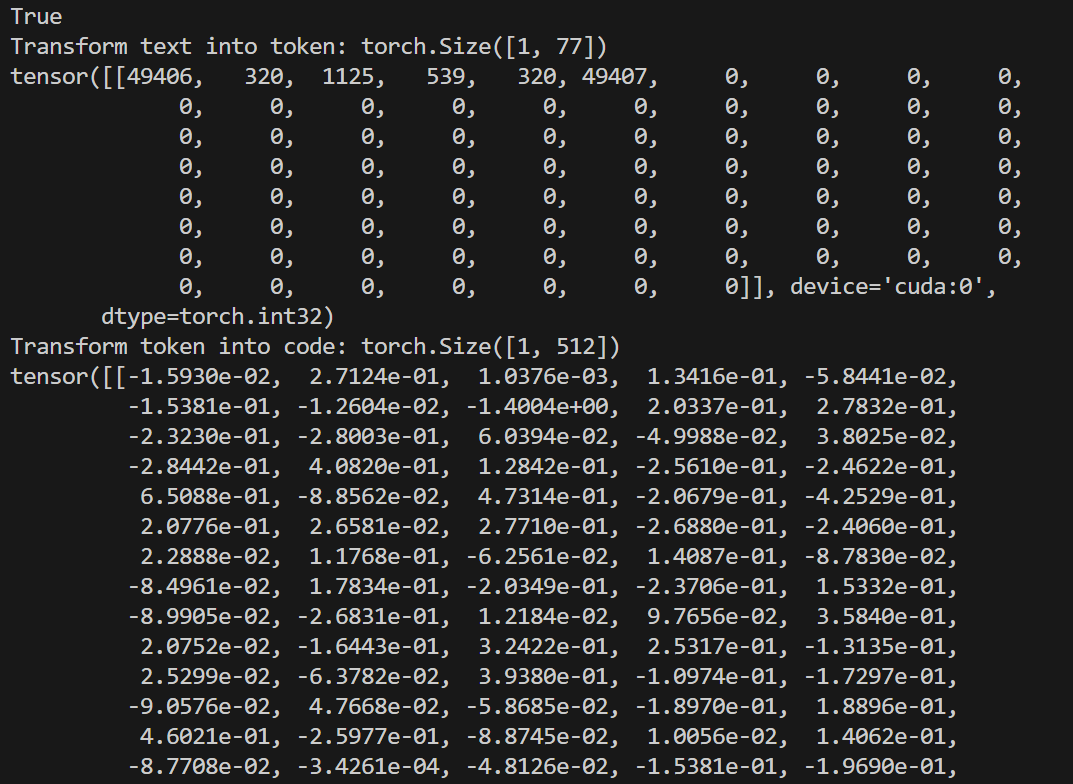
测试数据集使用CIFAR-100，以验证CLIP的zero-shot特性，模型作出略微调整，由原来的ResNet50更改为Vision Transformer B32,本实验使用cuda进行并行计算，以加速图片和文字的处理速度



实验一源代码



实验一结果：

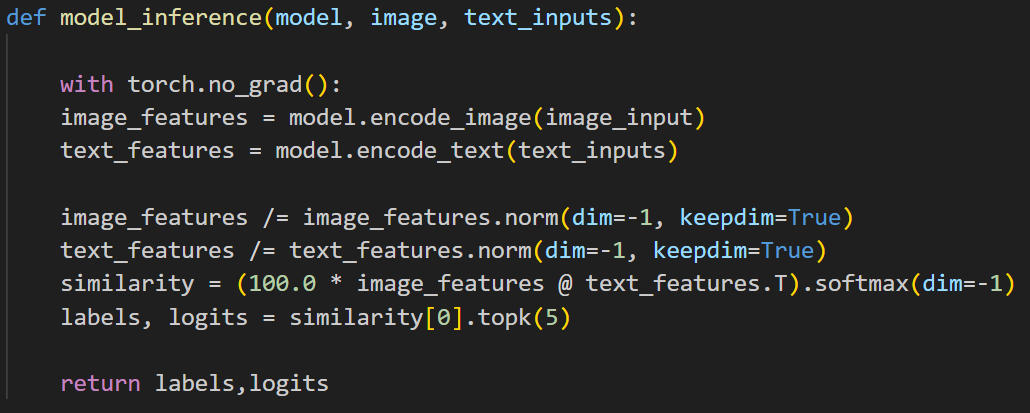


说明文本被转化成token，是一个1\*77的张量

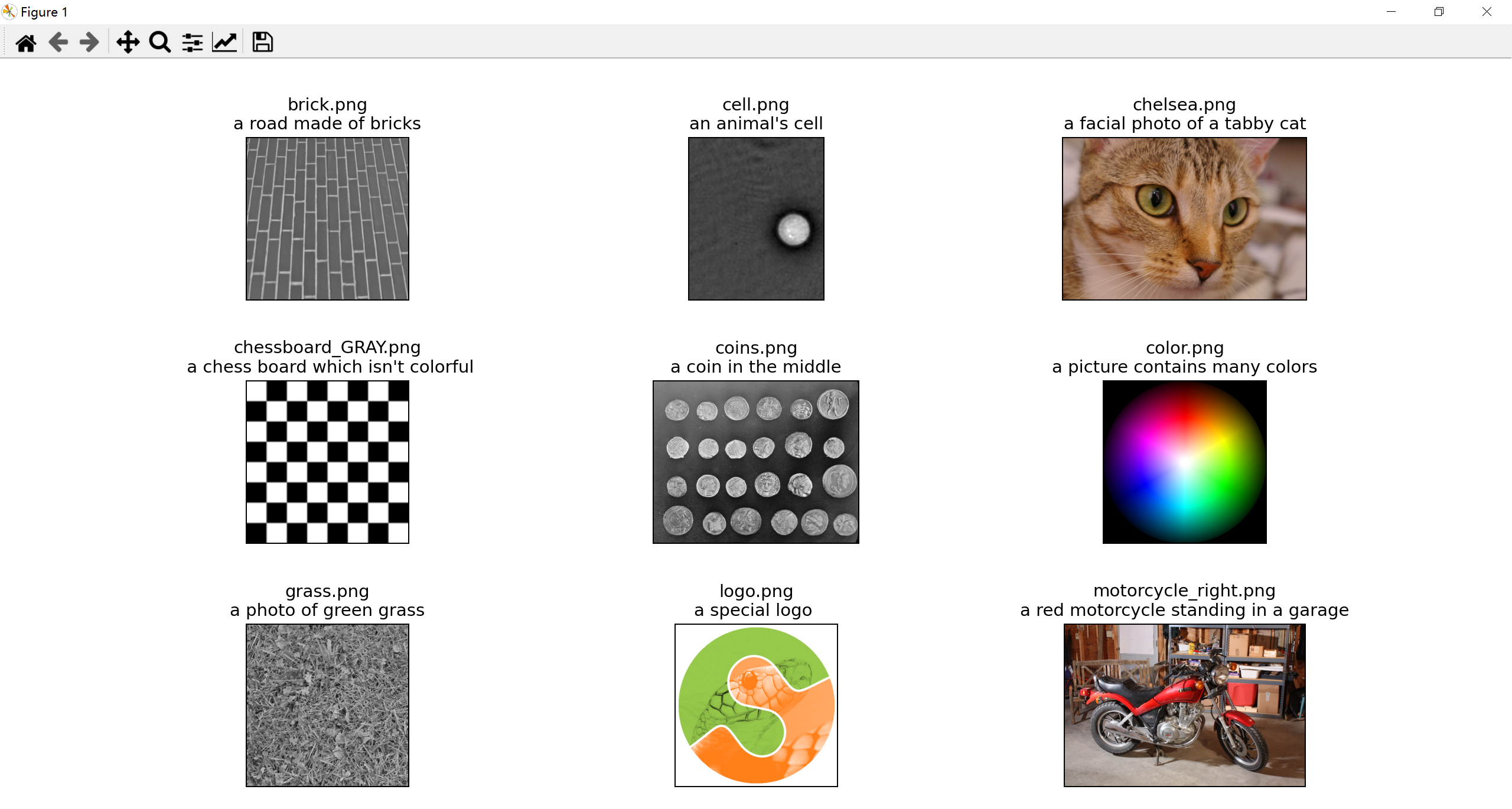
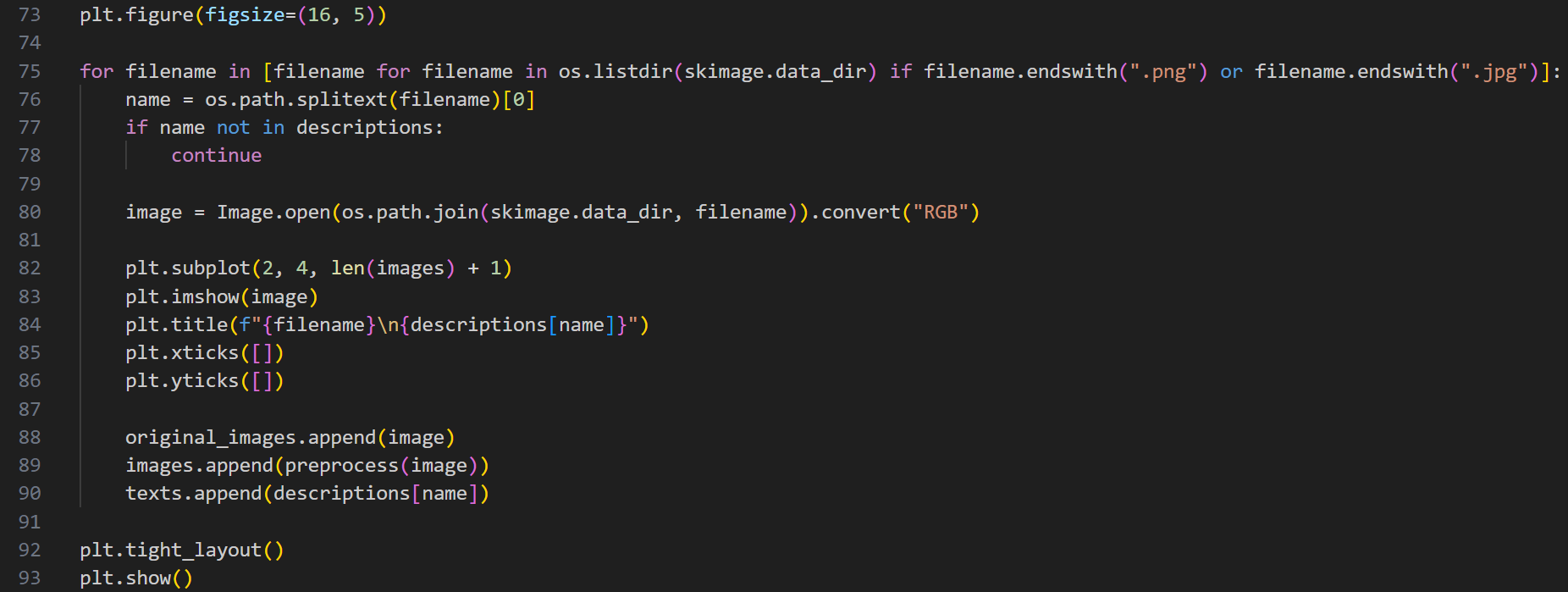
之后token被编码为一个1\*512的张量

实验二源代码：

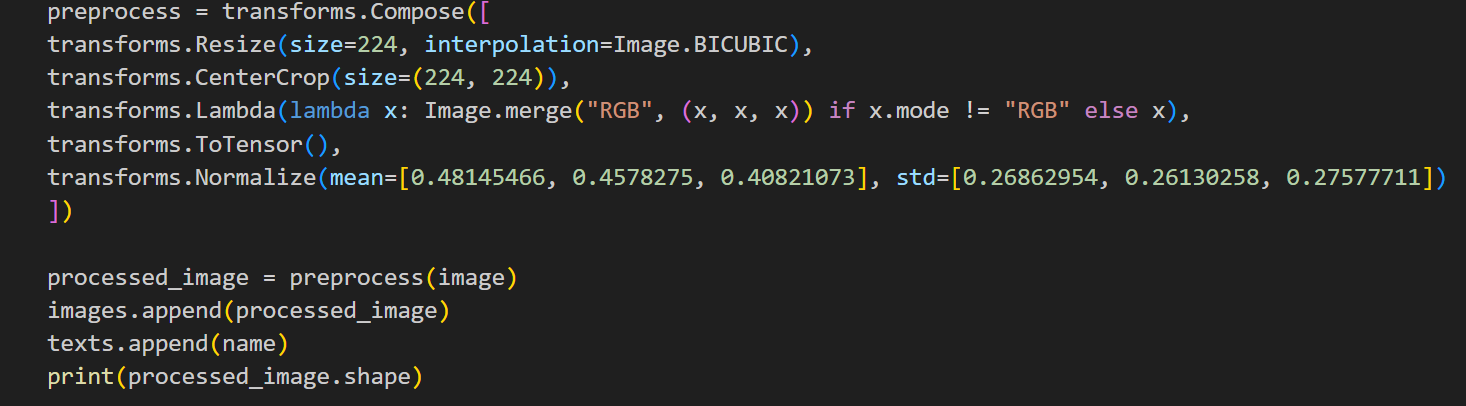
函数定义如下



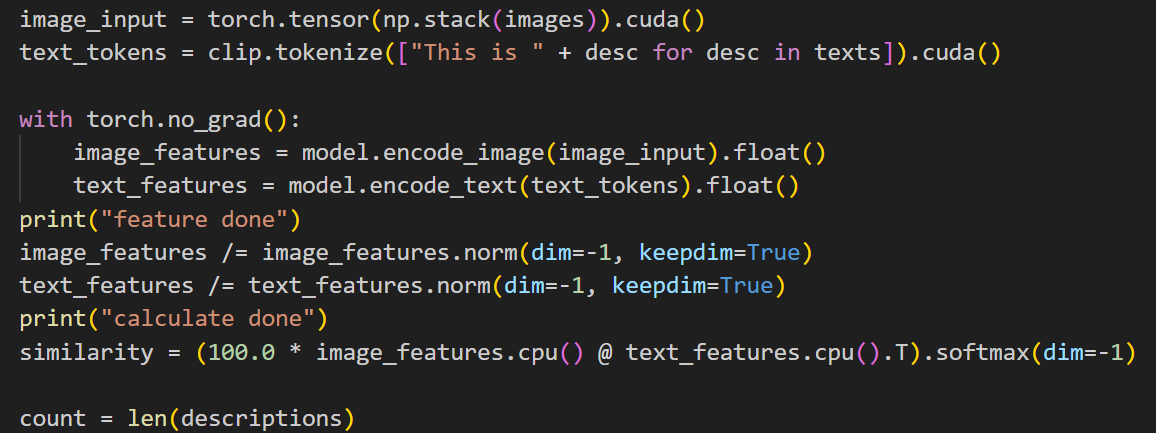
可视化结果如下



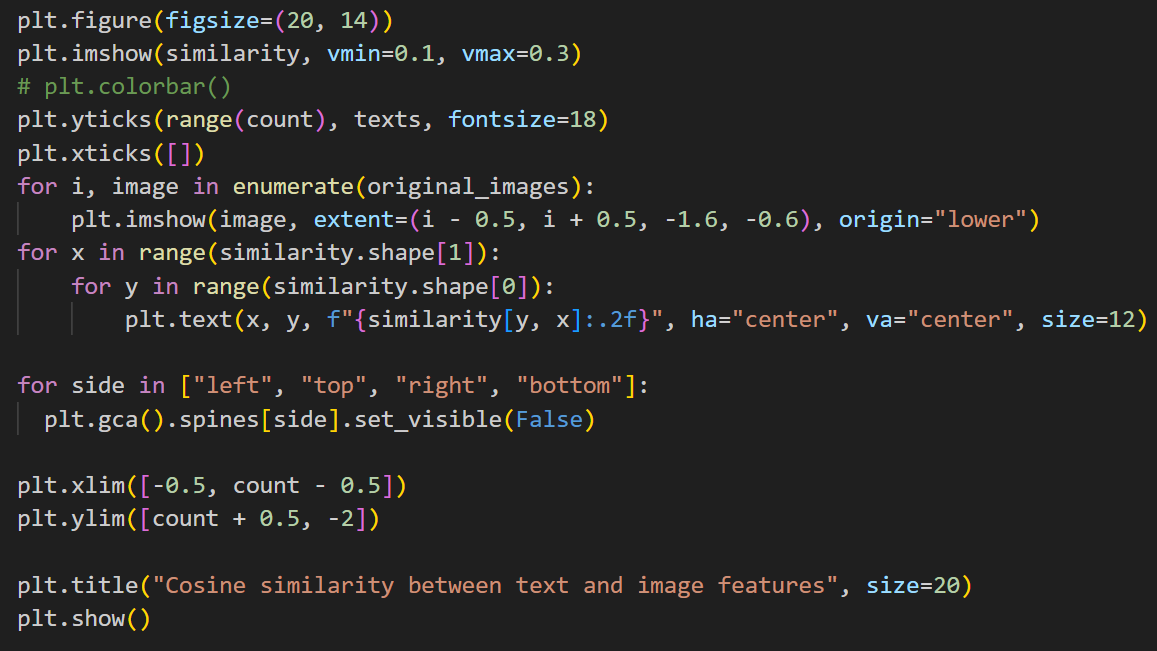
数据预处理

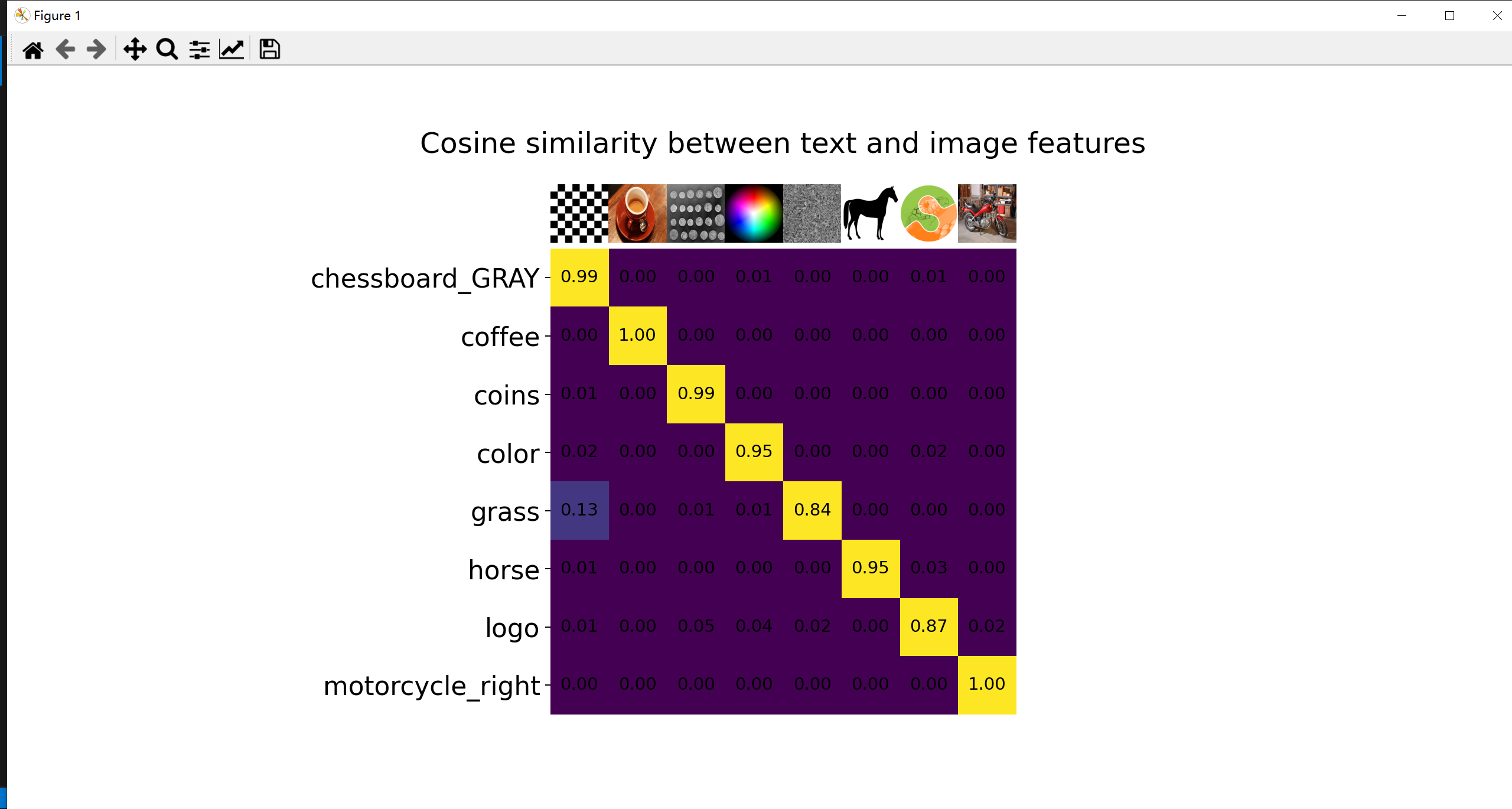


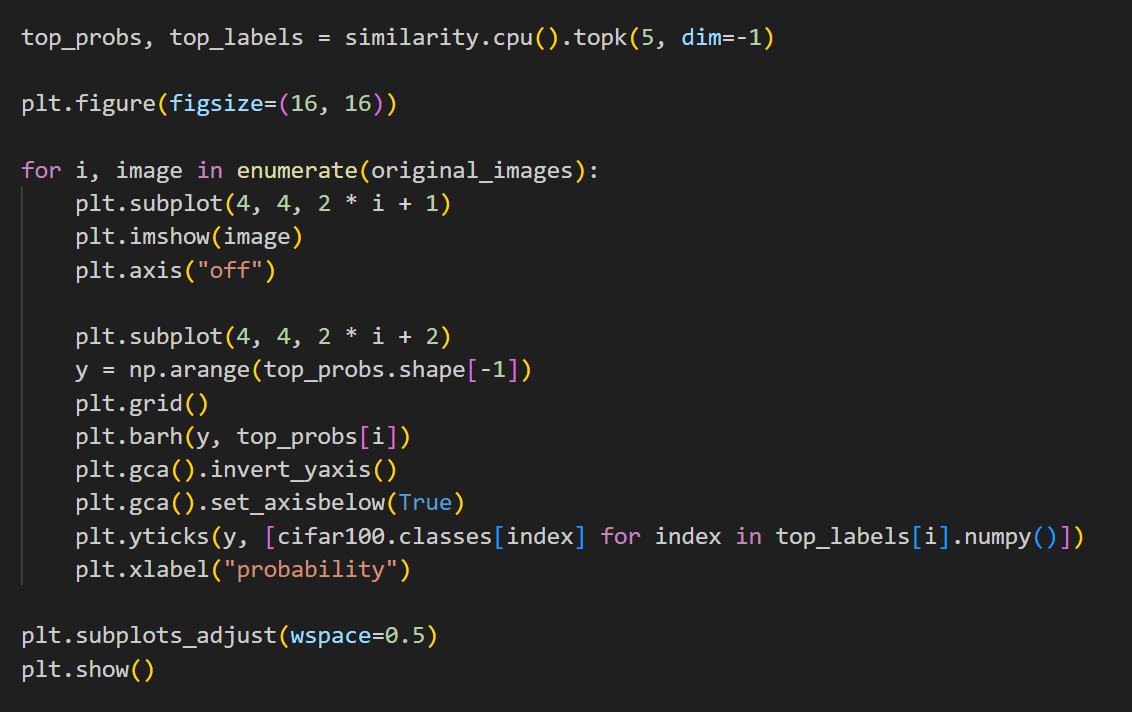
相似度计算

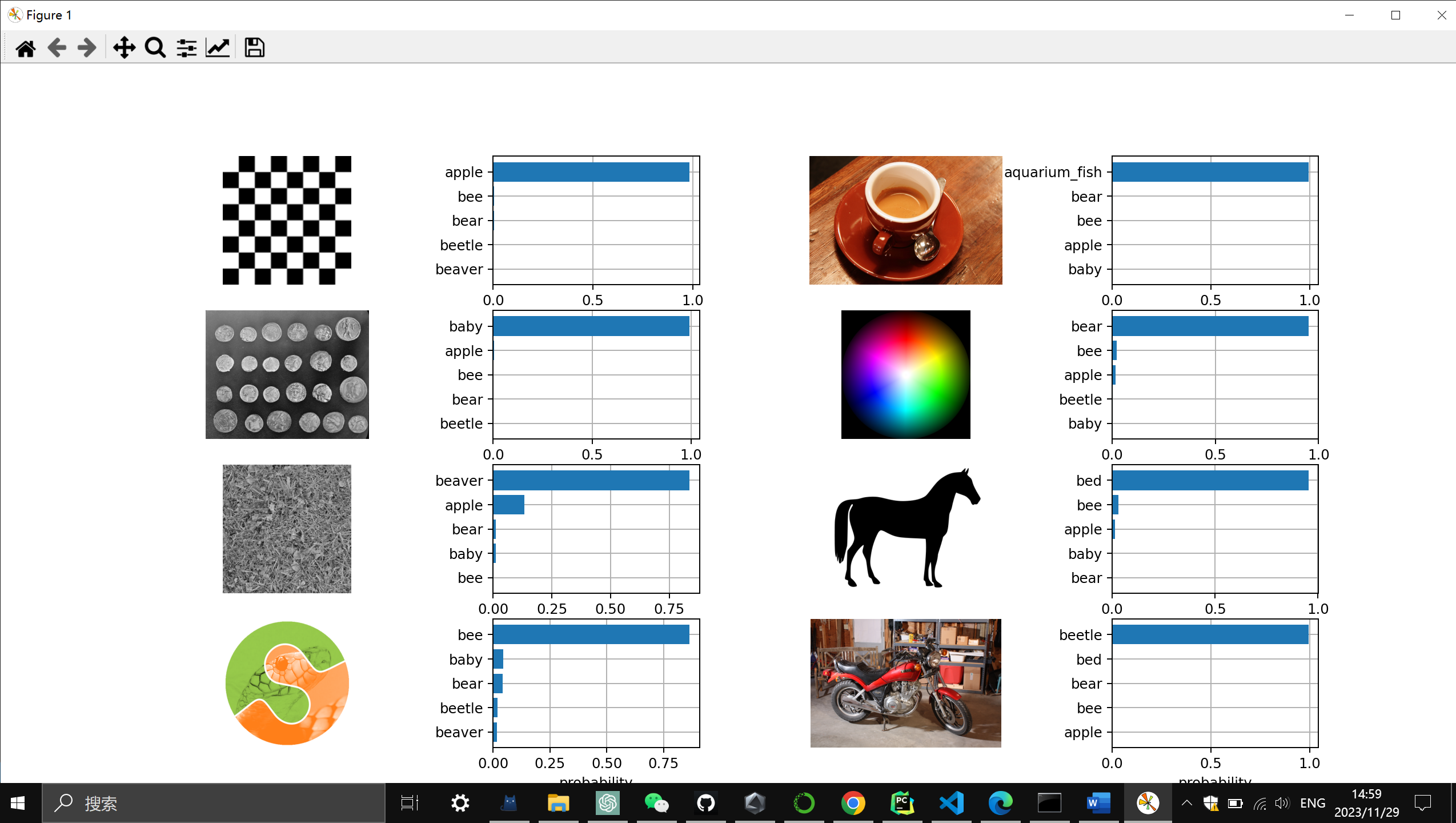


可视化









实验三，在分类任务上比较有意义

如果想搞的话，计算softmax的损失函数是交叉熵，