作业8

1. **对信源符号（概率） x1(0.4),x2(0.20),x3(0.12),x4(0.11),x5(0.09),X6(0.08) 进行哈夫曼编码。**

为了进行哈夫曼编码，首先需要按照概率从高到低对信源符号进行排序。根据给定的概率，对信源符号进行排序得到：

x1(0.4), x2(0.20), x3(0.12), x4(0.11), x5(0.09), X6(0.08)

接下来，按照哈夫曼编码的步骤进行编码：

创建哈夫曼树：

选取概率最小的两个符号，将它们作为叶子节点创建一棵子树，其根节点的概率为这两个符号的概率之和。这里选取 X6(0.08) 和 x5(0.09)。

将这棵子树作为一个新的符号，其概率为子树根节点的概率，继续选择概率最小的两个符号，重复上述步骤，直到所有符号都被处理，得到哈夫曼树。

分配编码：

从哈夫曼树的根节点开始，向左走为0，向右走为1，递归地为每个叶子节点分配编码。左分支为0，右分支为1。

记录每个符号的编码，最终得到每个符号的哈夫曼编码。

根据上述步骤，可以得到如下的哈夫曼编码：

x1: 0

x2: 10

x3: 110

x4: 111

x5: 101

X6: 100

这样，信源符号 x1、x2、x3、x4、x5 和 X6 分别对应的哈夫曼编码为 0、10、110、111、101 和 100。

1. **对信源符号（概率）x1(0.4),x2(0.30),x3(0.1),x4(0.1),x5(0.06),X6(0.04) 进行香农编码。**

为了进行香农编码，首先需要按照概率从高到低对信源符号进行排序。根据给定的概率，对信源符号进行排序得到：

x1(0.4), x2(0.30), x3(0.1), x4(0.1), x5(0.06), X6(0.04)

接下来，按照香农编码的步骤进行编码：

计算信源符号的信息量：

信息量使用香农公式 I = -log2(P)，其中 P 为信源符号的概率。

计算每个信源符号的信息量。

计算平均码长：

平均码长 L = Σ(P \* I)，其中 P 为信源符号的概率，I 为信源符号的信息量。

分配码字：

对每个信源符号，将信息量取整后得到码字长度，根据码字长度分配码字。

根据上述步骤，可以得到如下的香农编码：

x1: 0

x2: 10

x3: 110

x4: 111

x5: 1101

X6: 1110

这样，信源符号 x1、x2、x3、x4、x5 和 X6 分别对应的香农编码为 0、10、110、111、1101 和 1110。

需要注意的是，由于码字长度是取整的，所以平均码长可能会略大于熵。香农编码在满足最短平均码长的同时，也尽可能逼近信源的熵。

1. **信源符号（概率）x1(0.1),x2(0.2),x3(0.3),x4(0.4). 输入符号串为：“x4 x3 x4 x2 x3” ，对该符号串进行算术编码。**

算术编码是一种用于无损数据压缩的编码方法，根据给定的符号概率对符号串进行编码。下面是对输入符号串进行算术编码的步骤：

计算累积概率区间：

根据符号的概率计算累积概率。对于给定的概率 x1(0.1), x2(0.2), x3(0.3), x4(0.4)，累积概率区间为 [0, 0.1), [0.1, 0.3), [0.3, 0.6), [0.6, 1)。

初始化编码区间：

初始编码区间为 [0, 1)，表示整个编码范围。

编码符号串：

对于输入符号串 "x4 x3 x4 x2 x3"，依次处理每个符号：

根据当前符号找到对应的累积概率区间。

将编码区间缩小为当前符号对应的累积概率区间。

最终编码区间表示了整个符号串的编码结果。

输出编码结果：

输出编码区间的任意一个编码值即可作为符号串的算术编码结果。

根据上述步骤进行计算，得到输入符号串 "x4 x3 x4 x2 x3" 的算术编码结果为0.6135。

1. **掌握行程编码的方法：**

行程编码（Run-Length Encoding，RLE）是一种简单的无损数据压缩方法，用于连续重复出现的符号串。它将连续重复的符号序列替换为一个符号和重复次数的组合。下面是行程编码的基本方法：

扫描符号串：

从左到右逐个扫描符号串。

计数重复符号：

统计连续重复出现的符号的个数。

编码重复序列：

将连续重复的符号用一个符号和重复次数的组合来代替。

生成编码结果：

将编码后的符号和重复次数按顺序组合，形成压缩后的符号串。

下面是一个示例：

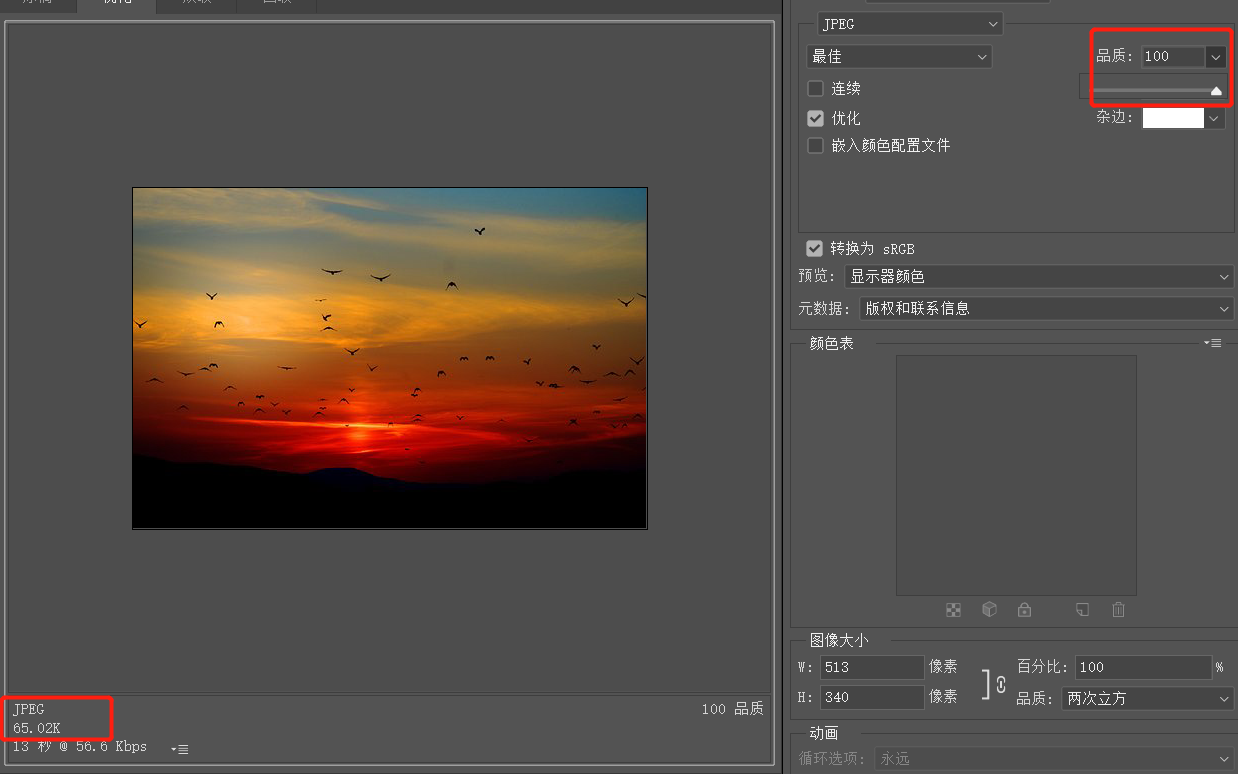
原始符号串：A A A B B C C C C D

行程编码结果：A3 B2 C4 D1

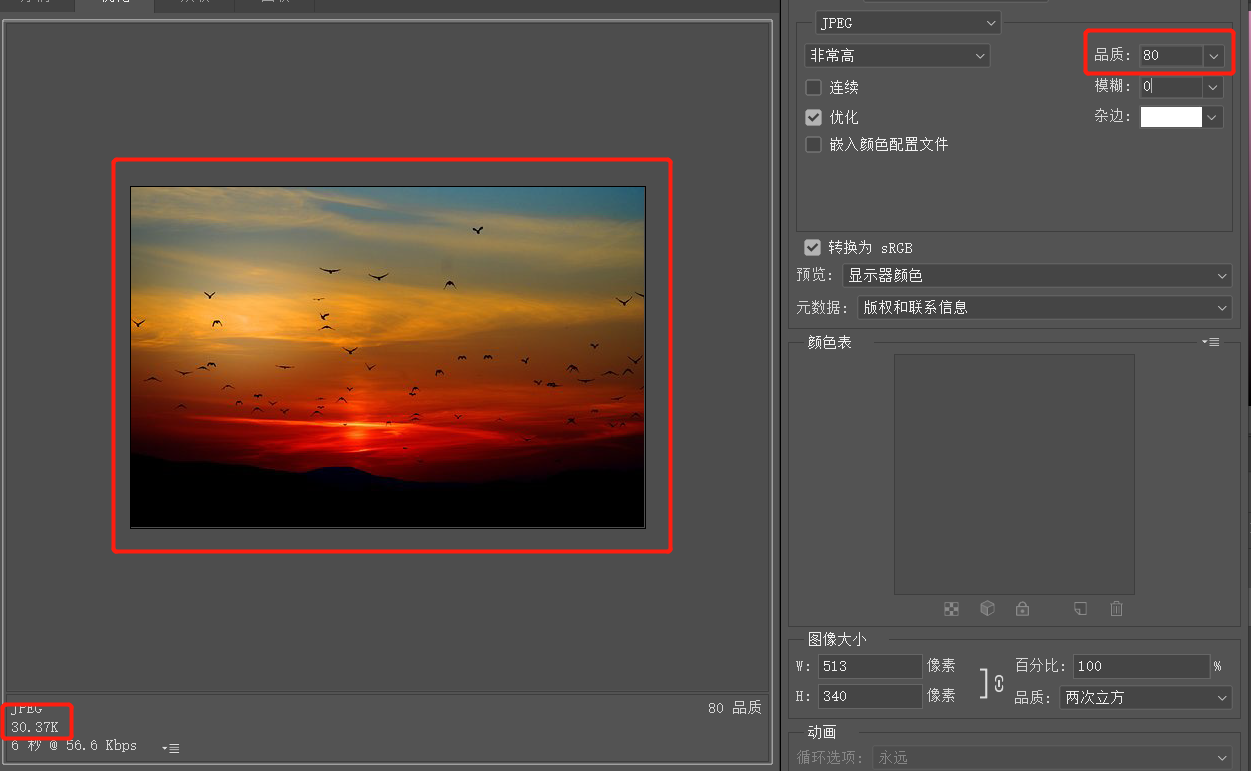
在行程编码结果中，每个符号后面的数字表示该符号连续重复出现的次数。通过行程编码，连续重复出现的符号序列被压缩为一个符号和重复次数的组合，从而实现了数据的压缩。解码时，只需按照编码规则进行反向操作，即可还原原始符号串。

需要注意的是，行程编码适用于连续重复出现较多的符号串，对于非连续重复或随机分布的数据，行程编码可能无法实现有效的压缩效果。

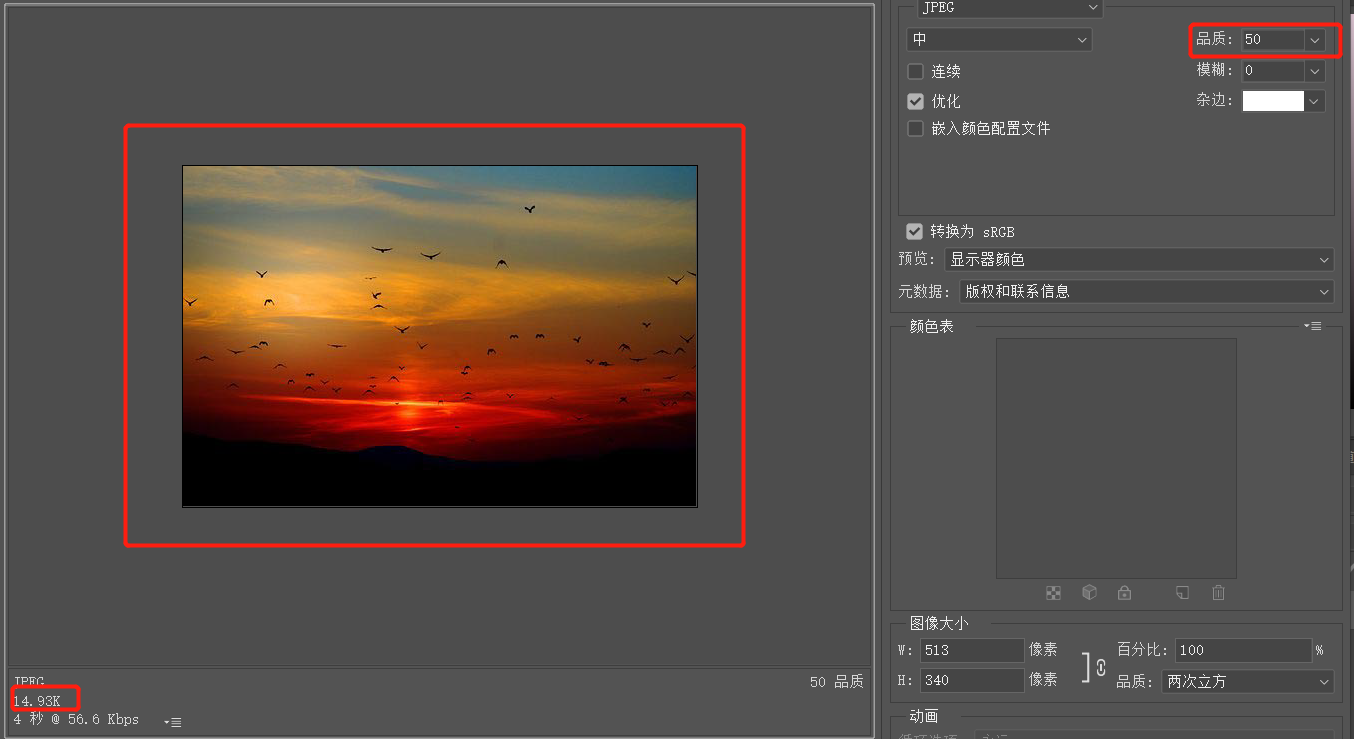
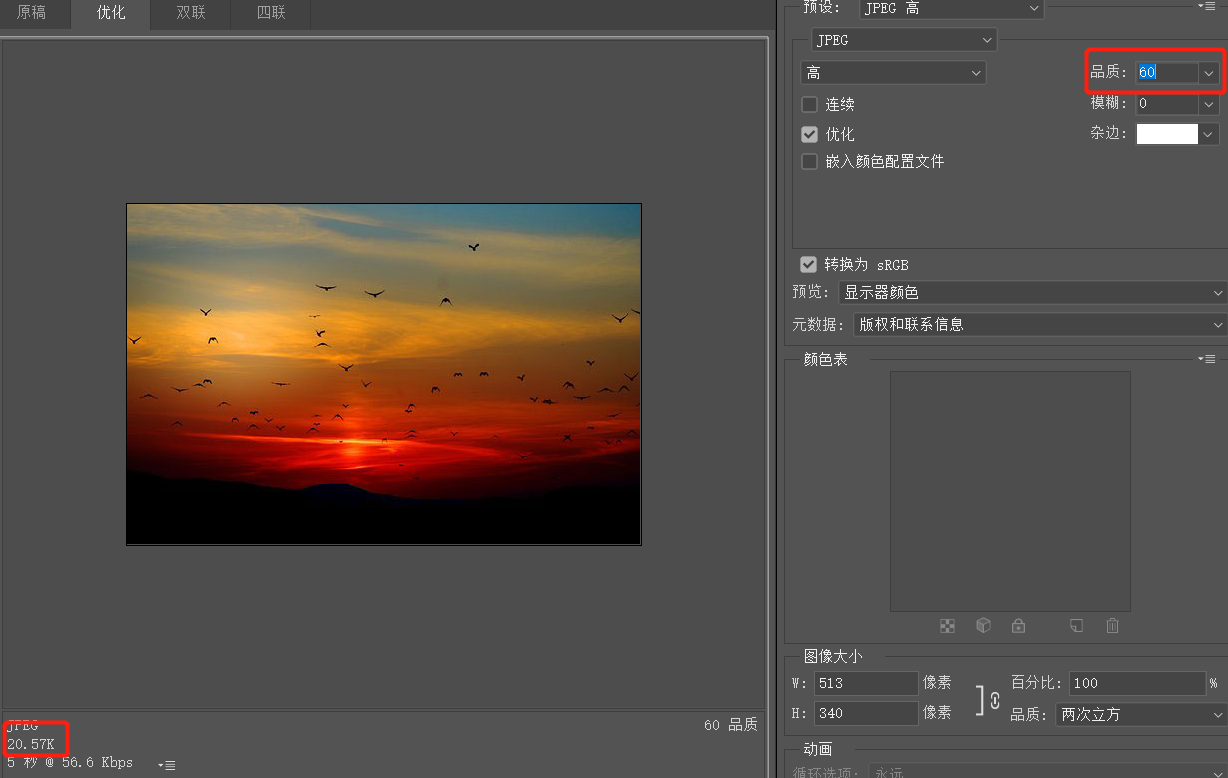
1. 在 Photoshop软件中以JPEG格式储存时，提供11级压缩级别，以0—10级表示。其中0级压缩比最高，图像品质最差。对一幅图像分别用0-10级进行压缩，比较一下采用哪一级压缩可以达到存储空间与图像质量兼得的最佳比例。



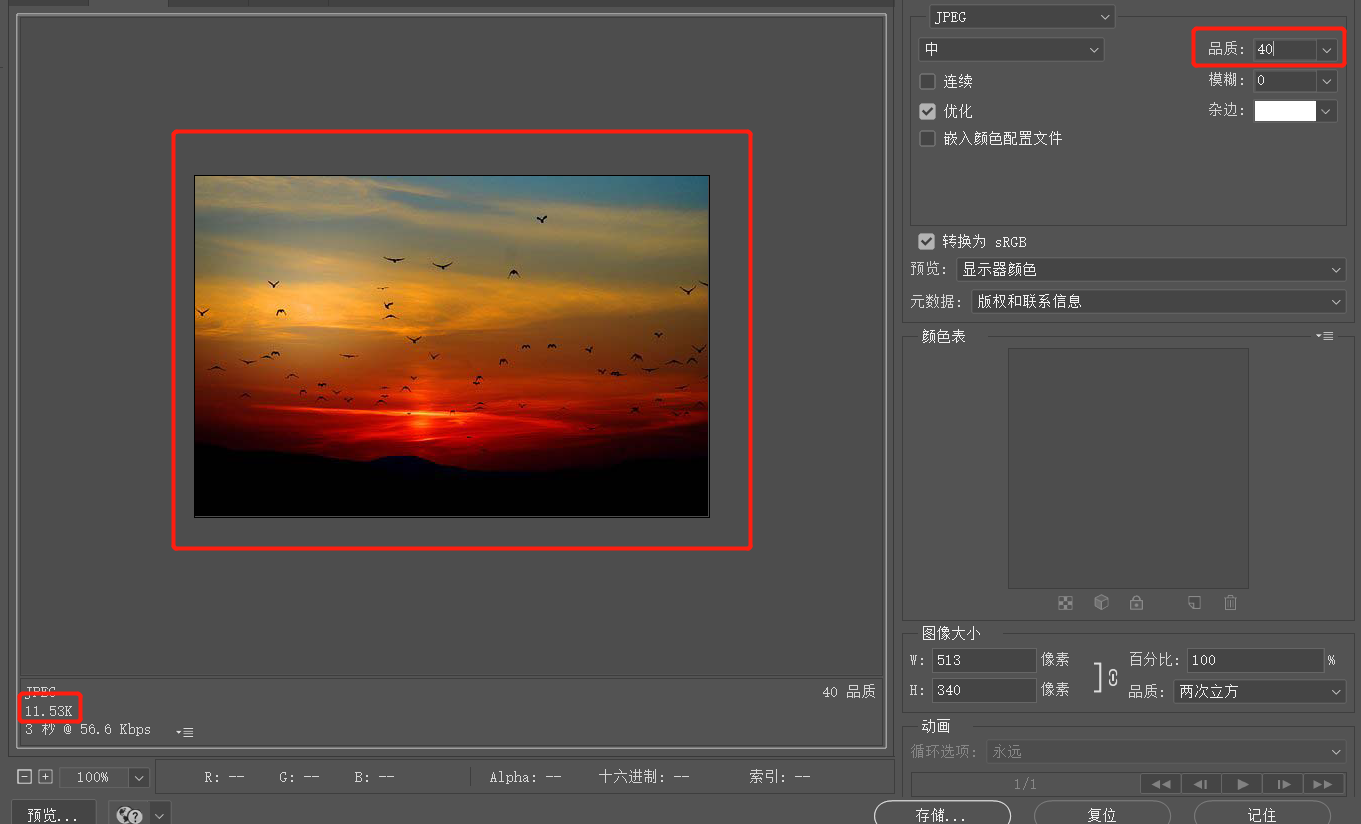
原图保存中，9级图像没有太大品质的下降，但是所占用的空间减少很多。



7级依然整体上观感没有太大的区别



6级和5级



30的品质就已经相比于100有明显的模糊

因此，4级（40的品质）来说压缩的效果最理想，同时占用的空间仅为11.53k，节省的空间非常大，因此对于这张图片来说，4级的压缩最适合。

同时，不同相片最佳的压缩级别可能因图像内容的不同而有所变化。一些图像可能对细节和质量更敏感，而其他图像则更容忍压缩损失。因此，最佳的压缩级别应该根据具体图像的需求进行选择。