

《人工智能》课程大作业

1. 题目

用 BP 神经网络分类“室内景物”和“室外景物”图像

2. 背景

随着计算机技术和网络技术特别是因特网技术的发展，以及近年来消费型电子产品（如数码相机、带拍照功能的手机、电子摄像机等）的普及，使得数字图像等多媒体数据的规模急剧膨胀，每天都有巨量的数字图像产生、发布和共享。如何从浩如烟海的图像信息中及时、高效地寻找所需要的图像成为一个需要迫切解决的问题。基于内容的图像检索问题正是在这样一个背景下提出。

目前的商用 Web 图像搜索引擎，如 Google 和百度都是采用和文本搜索类似的技术路线，即采用 Web 网页中与图像关联的文字信息（即图像标签文字）来完成搜索任务。这样的方法带来的问题是：（1）**搜索精度有限**。因为和用户输入的关键词匹配的网页图像标签文字并不一定能很好地描述图像本身的“内容”；（2）**无法实现“以图搜图”功能**。即根据图像而不是文字来搜索图片。“以图搜图”功能具有重要的应用价值。基于内容的图像检索（content-based image retrieval, CBIR）是有望解决上述问题的关键技术。图像检索是属于图像分析的一个研究领域。其目的是在给定查询参数（如查询图像）的前提下，依据内容信息或指定查询标准，在图像数据库中搜索并查找出符合查询条件的相应图像。

对图像检索技术的根本解决依赖于对人类认知图像能力的本质理解与突破，从而准确和高效地实现诸如“查找所有图像中包含黑色狗的图像”或者“查找所有图像中包含快乐小孩的图像”等等检索要求。限于目前的技术，当前绝大多数图像检索系统采用的是“按例检索”框架（query by example, QBE），即预先按照某种算法抽取图像的低层特征（如颜色、纹理和形状）并形成特征库，然后计算查询图像的特征并与特征库进行匹配以寻找相似的图像。

对一个海量图像数据库（典型的例子是因特网）而言，每次查询均采用原始的 QBE 框架显然效率是很低的，因此对 QBE 的一个改进就是“**分层检索**”技术。在“分层检索”框架里，图像数据库事先被预分为若干个语义类（如“汽车类”、“鲜花类”、“飞禽类”、“哺乳动物类”等），每个语义类对应一个子数据库（当然这个类可以继续划分下去，比如“汽车

类”又可以划分为“轿车类”、“卡车类”、“SUV 类”等等)。当用户将一幅检索图像提交给系统后,系统先对该图像进行分类,比如将其分为“卡车类”,这样下面的检索就可以直接在“卡车类”这个子数据库中进行,而不必在整个数据库上进行,从而大大提高了系统的运行效率。图 1 给出了一个简单的“分层检索”框架示意图。

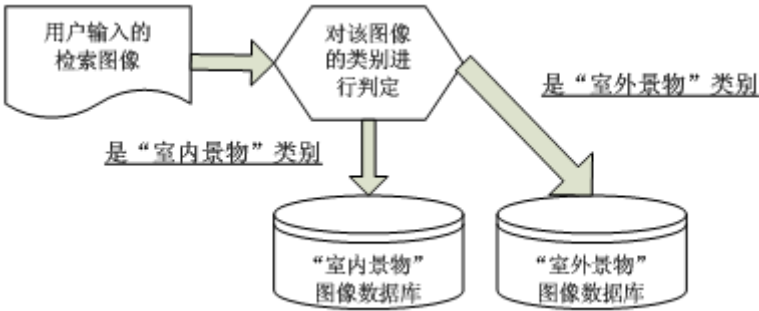


图 1 “分层检索”框架示意图

“分层检索”框架中的一个关键问题是如何判定输入的用户检索图像类别? 这属于模式识别和机器学习领域内的一个基本问题：**分类问题**，可以采用诸如神经网络等分类器来进行处理。

3. 问题和算法描述

本作业要求利用 BP 神经网络分类器对“室内景物”和“室外景物”进行分类，图 2 给出了这两类图像的示意图像。



图 2 待分类的图像类别

对任何一个分类问题而言，特征的设计总是一个基本且重要的问题。图像的基本特征有颜色、纹理和形状特征。对本问题而言，我们采用颜色和纹理特征。

3.1 颜色特征算法

可以用多种颜色空间来表达图像的色彩，*HSV* 颜色空间更接近于人对颜色的主观认知，在图像处理中得到了广泛应用。原始图像一般是基于 *RGB* 颜色空间的，需要进行色彩空间转换。*RGB* 颜色空间向 *HSV* 颜色空间转换的算法如下：

给定 RGB 颜色空间的值 $(r, g, b), r, g, b \in [0..255]$ ，则转换到 HSV 空间的 h, s, v 值计算方法如下：设 $v' = \max(r, g, b)$ ，则有：

$$\begin{cases} r' = \frac{v' - r}{v' - \min(r, g, b)} \\ g' = \frac{v' - g}{v' - \min(r, g, b)} \\ b' = \frac{v' - b}{v' - \min(r, g, b)} \end{cases} \quad (1)$$

$$v = \frac{v'}{255} \quad (2)$$

$$s = \frac{v' - \min(r, g, b)}{v'} \quad (3)$$

$$h' = \begin{cases} (5 + b'), & r = \max(r, g, b) \text{ and } g = \min(r, g, b); \\ (1 - g'), & r = \max(r, g, b) \text{ and } g \neq \min(r, g, b); \\ (1 + r'), & g = \max(r, g, b) \text{ and } b = \min(r, g, b); \\ (3 - b'), & g = \max(r, g, b) \text{ and } b \neq \min(r, g, b); \\ (3 + g'), & b = \max(r, g, b) \text{ and } r = \min(r, g, b); \\ (5 - r'), & \text{otherwise}; \end{cases} \quad (4)$$

$$h = 60 \times h' \quad (5)$$

计算得到的 $h \in [0..360], s \in [0, 1], v \in [0, 1]$ 。当 $r = g = b$ 时，即 $\max(r, g, b) = \min(r, g, b)$ 时， r', g', b' 没有意义，此时令 $h = s = 0, v = r / 255$ 。

由于一幅彩色图像的颜色非常多，如果对 HSV 空间进行适当量化后再计算颜色特征，则计算量要少很多。可以采用如下方法将 H, S, V 三个分量按照人的颜色感知进行非等间隔量化。对颜色模型进行大量分析后可以将色调 H 空间分成 8 份，饱和度 S 和亮度 V 空间分别分成 3 份，根据色彩的不同范围进行量化。量化后色调、饱和度和亮度值计算公式如下：

$$H = \begin{cases} 0 & \text{if } h \in [316, 20] \\ 1 & \text{if } h \in [21, 40] \\ 2 & \text{if } h \in [41, 75] \\ 3 & \text{if } h \in [76, 155] \\ 4 & \text{if } h \in [156, 190] \\ 5 & \text{if } h \in [191, 270] \\ 6 & \text{if } h \in [271, 295] \\ 7 & \text{if } h \in [296, 315] \end{cases} \quad (6)$$

$$S = \begin{cases} 0 & \text{if } s \in [0, 0.2) \\ 1 & \text{if } s \in [0.2, 0.7) \\ 2 & \text{if } s \in [0.7, 1] \end{cases} \quad (7)$$

$$V = \begin{cases} 0 & \text{if } v \in [0, 0.2) \\ 1 & \text{if } v \in [0.2, 0.7) \\ 2 & \text{if } v \in [0.7, 1] \end{cases} \quad (8)$$

按照以上的量化级，把三个颜色分量合成为一维特征矢量：

$$l = Q_s Q_v H + Q_v S + V \quad (9)$$

其中 Q_s 和 Q_v 分别是分量 S 和 V 的量化级数，取 $Q_s = 3, Q_v = 3$ 。上式可表示为：

$$l = 9H + 3S + V \quad (10)$$

这样将 H, S, V 这 3 个分量在一维矢量上分布开来， l 的取值范围为 $[0..71]$ 。综上所述，利用公式 (1) ~ (5) 可以将原始图像的 R 、 G 、 B 颜色通道转换为 H 、 S 、 V 颜色通道；再利用公式 (6) ~ (8) 可以将上述的 H 、 S 、 V 颜色通道分别进行量化；最后利用公式 (10) 则可以将量化后的 H 、 S 、 V 颜色通道合并为一个值。图 3 给出了示意图。

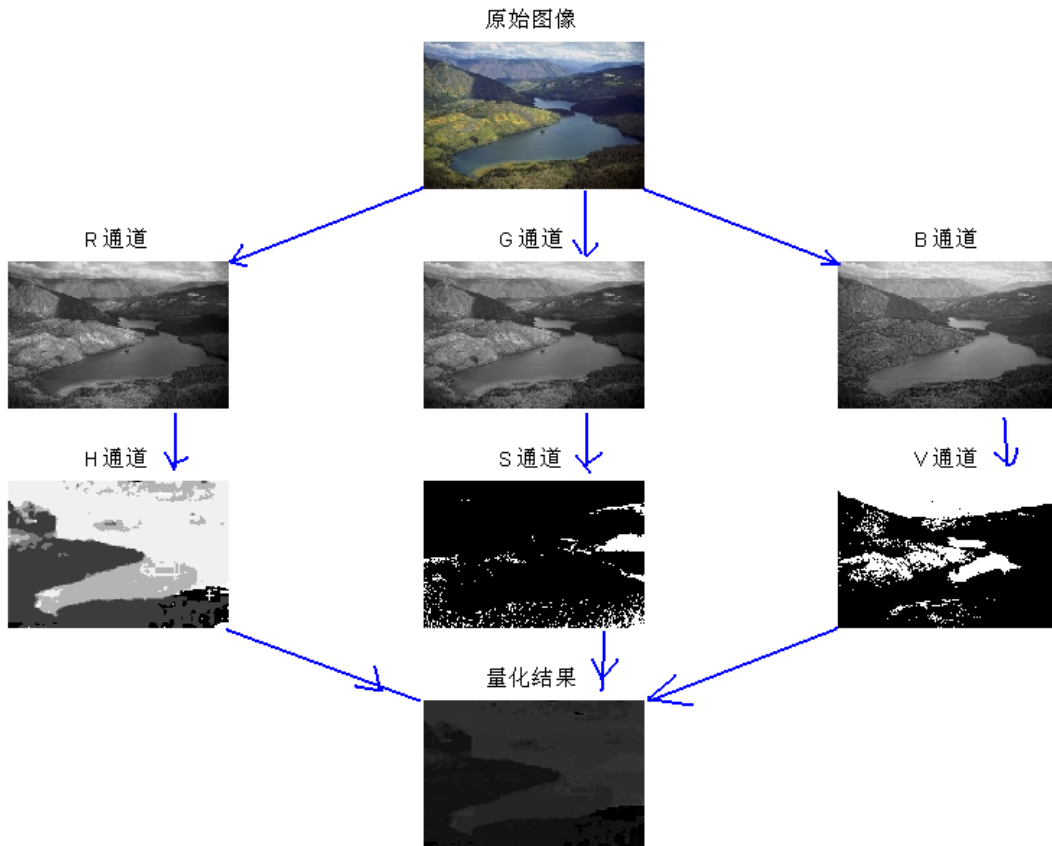


图 3 颜色空间转换

由图 3 可以看出，原始的彩色图像在 RGB 颜色空间中可以分解为三个颜色通道 R 、 G 、 B （每个颜色通道为一个矩阵，其大小和原始图像的大小相同），这三个颜色通道可以分别转换为 HSV 颜色空间的 H 、 S 、 V 通道（每个颜色通道为一个矩阵），最后量化合成为一个矩阵。颜色特征就是在这个矩阵的基础上进行计算的，如图 4 所示。

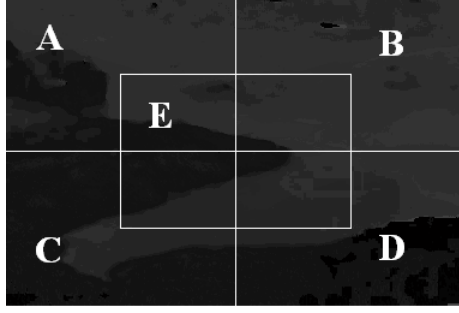


图 4 矩阵分块和颜色特征计算

得到 HSV 颜色空间的量化矩阵后，将该矩阵均匀分为 4 块（如图 4 中的 A 、 B 、 C 、 D 块），同时以该矩阵中心为中心，大小设置为和上述块一样，可以得到一个如图 4 中的 E 块（这样的方法不但可以提取原始图像中的色彩信息，还可以提取空间位置信息）。分别计算 $A \sim E$ 块的前三阶矩，这样可以得到 15 维的向量作为颜色特征。矩的计算公式如下：

令 $K = m \times n$ 为矩阵中元素的总个数（矩阵的行数和列数分别为 m ， n ），若位于矩阵第 i 行，第 j 列的元素值为 p_{ij} ，则前 3 阶矩特征的计算公式为：

$$\mu = \frac{1}{K} \sum_{i=1}^m \sum_{j=1}^n p_{ij} \quad (11)$$

$$\sigma = \left(\frac{1}{K} \sum_{i=1}^m \sum_{j=1}^n (p_{ij} - \mu)^2 \right)^{\frac{1}{2}} \quad (12)$$

$$s = \left(\frac{1}{K} \sum_{i=1}^m \sum_{j=1}^n (p_{ij} - \mu)^3 \right)^{\frac{1}{3}} \quad (13)$$

3.2 纹理特征算法

纹理是像素灰度值在空间上的变化函数，纹理特征计算是图像处理中一个很大的问题。在本作业中采用和颜色特征类似的计算方法，计算一个较为简单的图像纹理特征。具体方法是：先对原始彩色图像进行灰度化，然后按照图 4 中矩阵分块的方法，用公式 (11) ~ (13) 计算图像的纹理特征，如图 5 所示（子图 (a) 是原始彩色图像，子图 (b) 是对应的灰度化图像）。显然，计算得到的纹理特征也是 15 维的向量。



图 5 纹理特征计算

3.3 特征的融合与归一化

根据上述 3.1 和 3.2 节的方法，分别计算出 15 维的颜色特征和 15 维的纹理特征，将它们融合形成 30 维的特征向量，则最终一幅彩色图像的特征向量为 $V = [c_1, c_2, \dots, c_{15}, t_1, t_2, \dots, t_{15}]'$ 。其中 $c_i (i=1 \sim 15)$ 表示 15 个颜色特征分量， $t_i (i=1 \sim 15)$ 则表示 15 个纹理特征分量。

由于不同类型数据（如颜色特征和纹理特征）的不一致，对其进行归一化后会取得较好的结果。假定数据的分布是满足高斯分布的，那么可以采用如下的归一化方法：

令 μ 和 σ 分别表示融合后特征向量 V 的均值和标准差（ $\sigma \geq 0$ ），则有：

$$\begin{cases} V_i = (V_i - \mu) / (3 \times \sigma), & \text{if } (\sigma > 0) \\ V_i = (V_i - \mu), & \text{if } (\sigma = 0) \end{cases} \quad i = 1 \sim 30 \quad (14)$$

用公式（14）计算完后，再进行如下的处理：

$$\begin{cases} V_i = 1, & \text{if } (V_i > 1) \\ V_i = -1, & \text{if } (V_i < -1) \end{cases} \quad i = 1 \sim 30 \quad (15)$$

最后再利用下面的公式（16）将所有的数据转换到 $[0.0, 1.0]$ 的范围内：

$$V_i = (V_i + 1) / 2, \quad i = 1 \sim 30 \quad (16)$$

3.4 BP 网络的训练与分类

提供给同学们的图像数据总共有 200 幅，存放在名为“**ImgDB**”的目录中。其中 100 幅为室内景物图像（文件以“**Indoor_XXX.jpg**”命名），100 幅为室外景物图像（文件以“**Outdoor_XXX.jpg**”命名）。

具体处理步骤如下：

- （1）按照 3.1~3.3 节描述的算法计算所有 200 幅图像的特征并进行归一化；之后将这些特征数据储存到外部磁盘文件或数据库中（自行设计决定）；
- （2）从室内和室外景物图像中各随机挑选 30 幅作为训练样本训练 BP 神经网络；
- （3）训练好的 BP 神经网络对剩下的图像进行分类测试并计算分类正确率。

4. 要求

本大作业对锻炼同学们系统分析问题和解决问题的能力很有帮助，希望同学们认真查阅资料，上机实践，独立完成大作业。具体要求如下：

- （1）自行查阅相关文献和资料；
- （2）必须采用一种高级语言及其开发平台（如 C++、C#、Java 等，具体没有限制）完成上述作业中的所有程序设计，特别注意不能采用 Matlab 之类的语言和开发平台；

(3) 必须采用**面向对象的程序设计方法 (OOP 方法)**。设计的类中至少应该包括：矩阵（向量）类、特征类、BP 神经网络类（至少包括神经元类、神经网络层类、多层感知器网络类）；

(4) 提示：给定的图像文件是 JPEG 格式，可以自行寻找一些 JPEG 格式图像读取的类或开发库，可以把 JPEG 格式的图像数据读入内存矩阵中；

(5) 程序源代码具有良好的编程风格；

(6) 完成程序设计和实验后，**必须撰写一个正式的书面报告**。报告按照给定的模板文件“《人工智能》课程大作业模板.doc”中的排版格式和内容要求撰写。

5. 上交材料

期末课程结束后上交的材料包括：**正式报告、所有的程序源码及说明文档**。这些材料是本门课程成绩判定的依据。