
昂宝人工智能物联网 (AIOT) 科教平台

深度学习机器视觉的基本原理，应用实践与
具体设计

目录

| | | |
|-------|------------------------|----|
| 应用实践一 | 卷积神经网络（CNN）的结构与卷积特征 | 5 |
| 一、 | 实践目的 | 5 |
| 二、 | 实践原理 | 5 |
| 三、 | 设备 | 6 |
| 四、 | 实践操作 | 6 |
| 五、 | 实践数据 | 7 |
| 六、 | 数据分析 | 7 |
| 七、 | 结论与建议 | 7 |
| 应用实践二 | 使用 CNN 训练手写数字识别模型 | 8 |
| 一、 | 实践目的 | 8 |
| 二、 | 实践原理 | 8 |
| 三、 | 设备 | 9 |
| 四、 | 实践操作 | 9 |
| 五、 | 实践数据 | 10 |
| 六、 | 数据分析 | 10 |
| 七、 | 结论与建议 | 10 |
| 应用实践三 | 使用支持向量机（SVM）训练手写数字识别模型 | 11 |
| 一、 | 实践目的 | 11 |
| 二、 | 实践原理 | 11 |
| 三、 | 设备 | 14 |
| 四、 | 实践操作 | 15 |
| 五、 | 实践数据 | 16 |
| 六、 | 数据分析 | 16 |
| 七、 | 结论与建议 | 16 |
| 应用实践四 | 手写数字识别模型验证 | 17 |
| 一、 | 实验目的 | 17 |
| 二、 | 实验原理 | 17 |
| 三、 | 实验设置 | 19 |
| 四、 | 实验操作 | 19 |
| 五、 | 实验数据 | 21 |
| 六、 | 数据分析 | 21 |
| 七、 | 结论与建议 | 21 |
| 应用实践五 | 使用图传模组采集测试用图像数据 | 22 |
| 一、 | 实践目的 | 22 |
| 二、 | 实践原理 | 22 |
| 三、 | 设备 | 22 |

| | |
|---|-----------|
| 四、实践操作 | 23 |
| 五、实践数据 | 23 |
| 六、数据分析 | 24 |
| 七、结论与建议 | 24 |
| 应用实践六 人形跟踪使用的图像特征与模型 | 25 |
| 一、实践目的 | 25 |
| 二、实践原理 | 25 |
| 三、设备 | 26 |
| 四、实践操作 | 26 |
| 五、实践数据 | 28 |
| 六、数据分析 | 30 |
| 七、结论与建议 | 31 |
| 应用实践七 人形跟踪算法中的机器学习模型训练 | 32 |
| 一、实践目的 | 32 |
| 二、实践原理 | 32 |
| 三、设备 | 33 |
| 四、实践操作 | 33 |
| 六、数据分析 | 35 |
| 七、结论与建议 | 36 |
| 应用实践八 人形跟踪中的模型在人形识别中的测试 | 37 |
| 一、实践目的 | 37 |
| 二、实践原理 | 37 |
| 三、设备 | 38 |
| 四、实践操作 | 38 |
| 五、实践数据 | 40 |
| 六、数据分析 | 40 |
| 七、结论与建议 | 41 |
| 应用实践九 不同环境噪声中人形跟踪算法的测试 | 42 |
| 一、实践目的 | 42 |
| 二、实践原理 | 42 |
| 三、设备 | 42 |
| 四、实践操作 | 42 |
| 五、实践数据 | 43 |
| 六、数据分析 | 44 |
| 七、结论与建议 | 44 |
| 应用实践十 人脸识别模块的使用 | 45 |
| 一、实践目的 | 45 |
| 二、实践原理 | 45 |
| 1、MTCNN 网络：轻量级人脸检测网络，论文《JOINT FACE DETECTION AND ALIGNMENT USING MULTI-TASK CASCADED CONVOLUTIONAL NETWORK》 | 45 |

| | |
|-------------------------------|-----------|
| 三、设备 | 45 |
| 四、实践操作 | 45 |
| 五、实践数据 | 45 |
| 六、数据分析 | 45 |
| 七、结论与建议 | 45 |
| 应用实践十一 语音识别模块的使用 | 46 |
| 一、实践目的 | 46 |
| 二、实践原理 | 46 |
| 三、设备 | 46 |
| 四、实践操作 | 46 |
| 五、实践数据 | 46 |
| 六、数据分析 | 46 |
| 七、结论与建议 | 46 |

On-Bright confidential

应用实践一 卷积神经网络（CNN）的结构与卷积特征

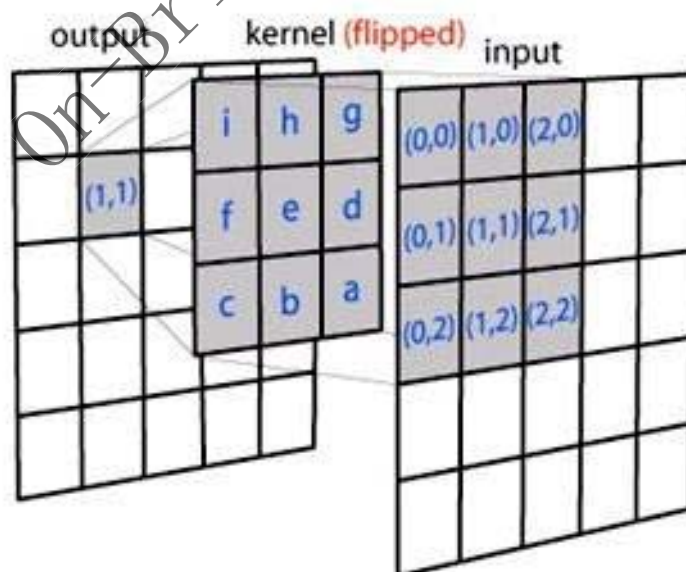
一、实践目的

- 1、了解卷积神经网络的基本概念。
- 2、了解卷积核的基本概念。
- 3、掌握不同卷积核对不同特征的提取。

二、实践原理

卷积神经网络：卷积神经网络（Convolutional Neural Networks, CNN）是一类包含卷积计算且具有深度结构的前馈神经网络（Feedforward Neural Networks），是深度学习（deep learning）的代表算法之一，它的人工神经元可以响应一部分覆盖范围内的周围单元，对于大型图像处理有出色表现。

卷积核：卷积核就是图像处理时，给定输入图像，在输出图像中每一个像素是输入图像中一个小区域中像素的加权平均，其中权值由一个函数定义，这个函数称为卷积核。CNN 中的卷积核跟传统的卷积核本质没有什么不同。仍然以图像为例，卷积核依次与输入不同位置的图像块做卷积，得到输出，如下图：



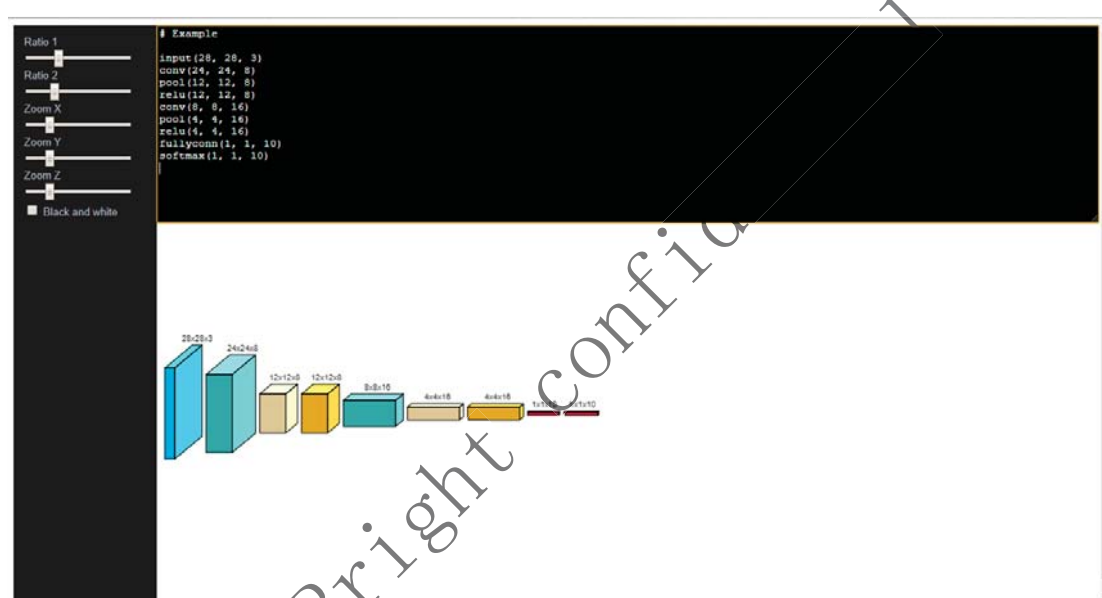
三、设备

本实验所需硬件设备如下：

- 1、PC 机一台。
- 2、0B 人工智能科教平台软件。

四、实践操作

- 1、点击教材底部进入“机器学习”按钮，选择第一行第一列的“手写数字识别相关实验”按钮，进入手写数字识别实验选择页面。点击“实验二”按钮，进入实验页面。
- 2、在软件如下界面中，有一段表示 CNN 的伪代码



input 表示输入

conv 表示卷积层

pool 表示池化层

relu 表示激活函数

fullyconn 表示全连接层

softmax 表示 softmax 函数

括号中前两个数表示输入信号维度，第三个数表示输入信号通道数

例程中输入的伪代码所生成的 CNN 如界面下方白色区域所示。请自行修改伪代码，观察输出，并查阅资料，了解不同层之间的连接顺序以及输入输出维度关系。

- 3、在界面的右边，点击按钮“请选择一张图像”，并在弹框中选择上个实验中保存的

图像，在左边输入图像黑框中会显示选择的图像。

- 4、任意修改卷积核 1、2、3 的参数，例如卷积核 1 选择 Sobel_H，数量为 1 层，卷积核 2 选择 Sobel_V，数量为 0 层，卷积核 3 选择 Laplace，数量为 0 层，点击按钮“输出特征图像”，即可在右边输出图像黑框中显示经过设置好的卷积核提取的图像特征。

五、实践数据

- 1、按照实验步骤操作 2，输出一幅完整的 CNN 网络模型图；
- 2、输出在不同卷积核的组合下，图像的输出特征图；

六、数据分析

分析为什么不同卷积核的组合提取到的图像不一样

七、结论与建议

卷积核具有的一个属性就是局部性。即它只关注局部特征，局部的程度取决于卷积核的大小。比如用 Sobel 算子进行边缘检测，本质就是比较图像邻近像素的相似性。

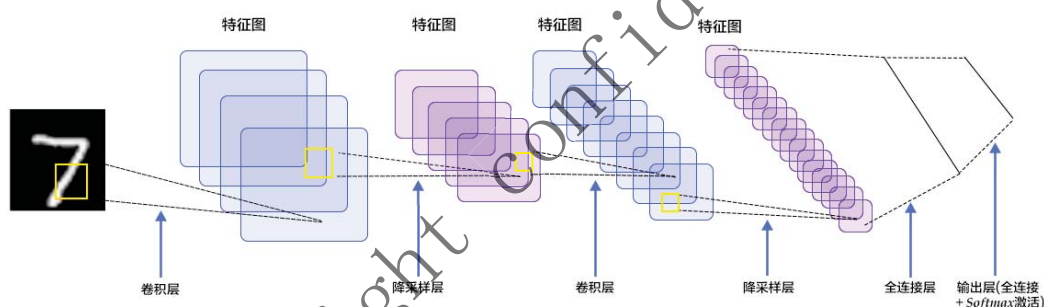
应用实践二 使用 CNN 训练手写数字识别模型

一、实践目的

- 1、了解经典卷积神经网络 Lenet-5 的结构。
- 2、了解 CNN 中基本的训练参数。
- 3、通过实验认识 CNN 的训练与参数设置对训练效果的影响。

二、实践原理

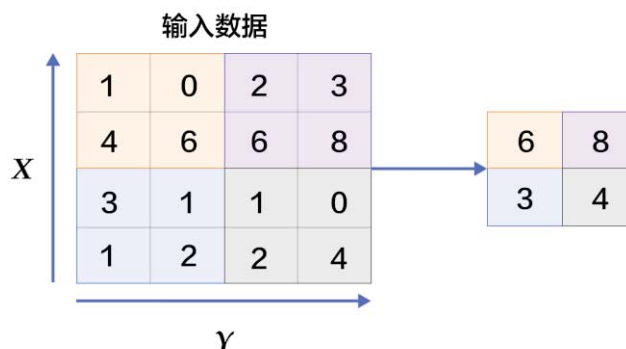
Lenet-5 卷积神经网络：出自论文 Gradient-Based Learning Applied to Document Recognition，是一种用于手写体字符识别的非常高效的卷积神经网络。网络的结构下图所示：



卷积神经网络中的卷积层、池化层、全连接层的解释如下：

卷积层：卷积层是卷积神经网络的核心基石。在图像识别里我们提到的卷积是二维卷积，即离散二维滤波器（也称作卷积核）与二维图像做卷积操作，简单的讲是二维滤波器滑动到二维图像上所有位置，并在每个位置上与该像素点及其领域像素点做内积。具体可见实验一。

池化层：池化是非线性下采样的一种形式，主要作用是通过减少网络的参数来减小计算量，并且能够在一定程度上控制过拟合。通常在卷积层的后面会加上一个池化层。池化包括最大池化、平均池化等。其中最大池化是用不重叠的矩形框将输入层分成不同的区域，对于每个矩形框的数取最大值作为输出层，如下图所示，在 2x2 范围内，选出最大的数字作为输出，最终组成新的特征图。



全连接层：在整个卷积神经网络中起到“分类器”的作用。全连接的作用是：将卷积层，池化层学到的图像特征映射到样本标签空间。其核心操作即为矩阵向量乘积：

$$y = Wx + b$$

其中， W 为全连接层的权重矩阵， b 为全连接层的偏置向量，这两个参数通过训练得到； x 为前面卷积层，池化层计算得到的卷积特征矩阵， y 为全连接层的输出，也即整个网络的预测输出。

CNN 的训练原理：由上一部分知道，一个完整的 LeNet-5 网络，其可训练的参数包括：卷积层的权重，全连接层的权重和偏置项。训练这些参数的方法主要为：

- (1) 给网络输入训练样本（图片），得到相应的输出（预测值），将输出（预测值）与样本的实际标签（正确值）进行对比，并计算损失函数 Loss，本实验使用的是交叉熵损失函数。该过程又叫前向传播。
- (2) 通过损失函数计算得到损失值，在使用链式求导法则将这一损失值回传到整个网络，并按照一定的算法根据求导得到的梯度更新网络的参数，本实验使用的参数更新方法为 Adam。该过程右键反向传播。
- (3) 通过使用大量训练样本，多次迭代训练后，最终得到训练完成的 CNN 模型。

三、设备

本实验所需硬件设备如下：

- 1、PC 机一台。
- 2、OB 人工智能科教平台软件。

四、实践操作

- 1、进入软件后，点击“手写数字识别实验”，进入实验目录，点击选择“实验二”

2、实验二界面如下：



3、点击“训练测试数据存储路径”的文件夹按钮，选择训练数据，比如放置于桌面的“语音模块及训练数据\train_data”文件夹。同理，选择模型保存位置，比如放置于桌面的“语音模块及训练数据\train_result”文件夹。

4、选择 batch，epoch，学习率，更新比率 4 个参数，参数的简单解释已在软件中说明，更详细的说明请查阅相关资料。

5、点击“训练并保存模型”按钮，稍等一会，界面右侧就会持续输出训练的状态数值。训练结束时，会显示“Model has been saved!”字样，训练完成，模型保存于指定的文件夹。

6、训练可以中断，点击“取消训练”按钮，即可中断训练，注意：中断训练不会保存训练结果。

五、实践数据

统计在不同的训练参数下，Loss 值的变化趋势，以及测试集的正确率

六、数据分析

通过查询资料，分析不同参数设置下，对 Lenet-5 的测试准确率有何影响。

七、结论与建议

应用实践三 使用支持向量机（SVM）训练 手写数字识别模型

一、实践目的

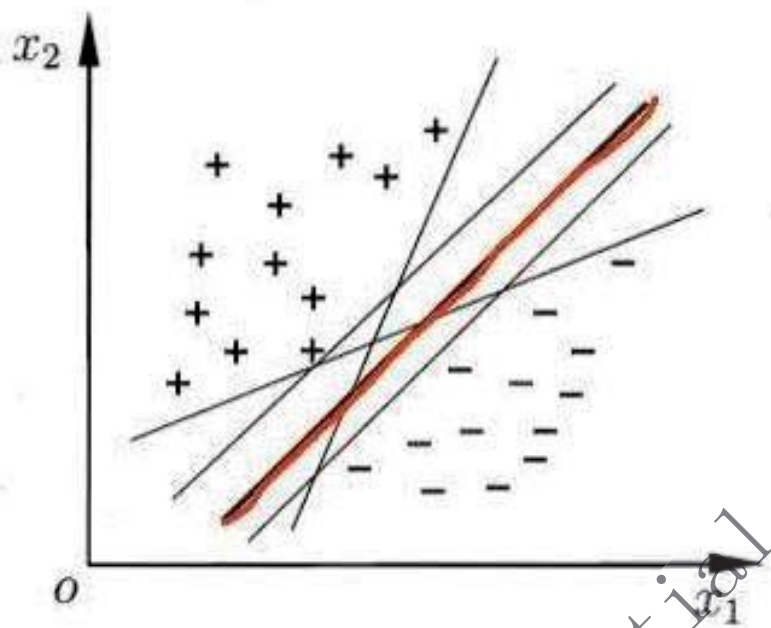
- 1、了解数据集制作的一种方法与数据格式。
- 2、了解训练分类器的学习方法 SVM（支持向量机）。
- 3、了解 SVM 中几个参数的意义和用法。

二、实践原理

支持向量机（SVM）：一种二分类模型，它的目的是寻找一个超平面来对样本进行分割，分割的原则是间隔最大化，最终转化为一个凸二次规划问题来求解。由简至繁的模型包括：

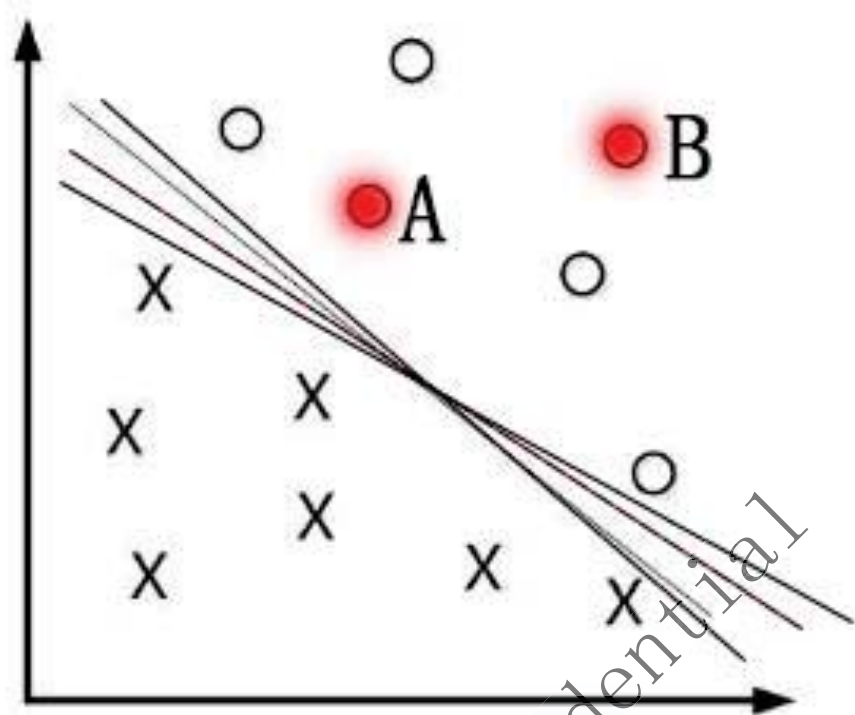
- 当训练样本线性可分时，通过硬间隔最大化，学习一个线性可分支持向量机
- 当训练样本近似线性可分时，通过软间隔最大化，学习一个线性支持向量机
- 当训练样本线性不可分时，通过核技巧和软间隔最大化，学习一个非线性支持向量机

给定训练样本集，分类学习最基本的想法就是基于训练集 D 在样本空间中找到一个划分超平面，将不同类别的样本分开：



存在多个划分超平面将两类训练样本分开

线性可分：如果一个线性函数能够将样本分开，称这些数据样本是线性可分的。线性函数在二维空间中就是一条直线，在三维空间中就是一个平面，以此类推，如果不考虑空间维数，这样的线性函数统称为超平面。看一个简单的二维空间的例子，0 代表正类，X 代表负类，样本是线性可分的，但是很显然不只有这一条直线可以将样本分开，而是有无数条，所说的线性可分支持向量机就对应着能将数据正确划分并且间隔最大的直线：



一般来说，一个点距离分离超平面的远近可以表示分类预测的确信度，如图中的 A B 两个样本点，B 点被预测为正类的确信度要大于 A 点，所以 SVM 的目标是寻找一个超平面，使得离超平面较近的异类点之间能有更大的间隔，即不必考虑所有样本点，只需让求得的超平面使得离它近的点间隔最大。

间隔计算：在样本空间中，划分超平面可通过如下线性方程来描述：

$$W^T x + b = 0$$

其中 w 为法向量，决定了超平面的方向， b 为位移量，决定了超平面与原点的距离。假设超平面能将训练样本正确地分类，即对于训练样本 (x_i, y_i) ，满足以下公式：

$$\begin{cases} W^T x_i + b \geq +1 & y_i = +1 \\ W^T x_i + b \leq -1 & y_i = -1 \end{cases}$$

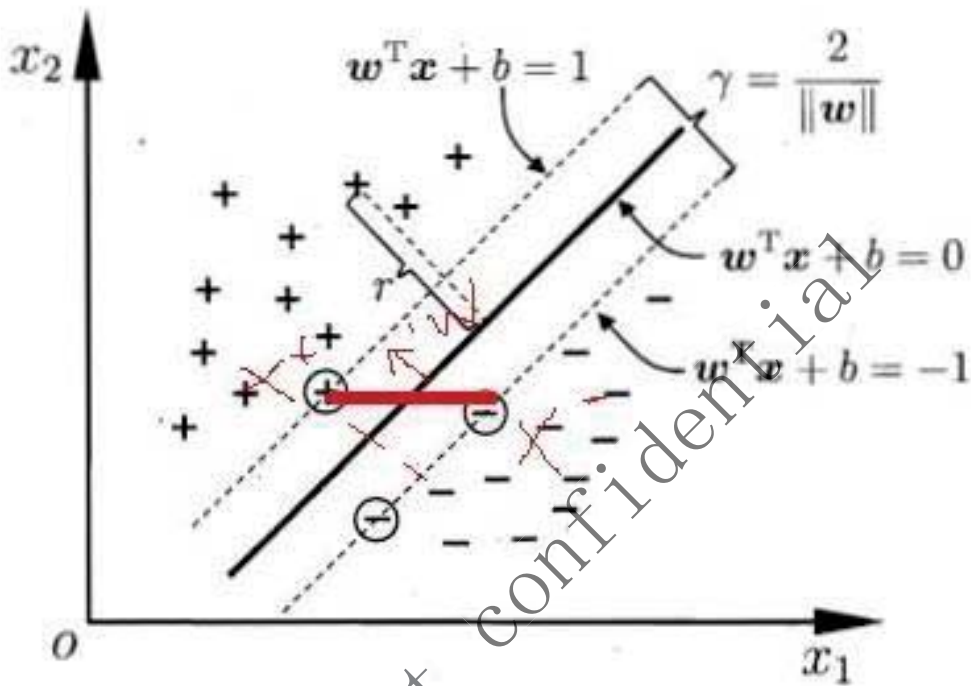
该公式被称为最大间隔假设， $y_i = +1$ 表示样本为正样本， $y_i = -1$ 表示样本为负样本，式子前面选择大于等于+1，小于等于-1 只是为了计算方便，原则上可以是任意常数，但无论是多少，都可以通过对 w 的变换使其为 +1 和 -1 。实际上该公式等价于：

$$y_i(W^T x_i + b) \geq +1$$

如下图所示，距离超平面最近的这几个样本点满足：

$$y_i(W^T x_i + b) = 1$$

它们被称为“支持向量”。虚线称为边界，两条虚线间的距离称为间隔（margin）。



核函数：一种非线性变换，将输入空间映射到高维特征空间。核函数包括线性核函数、多项式核函数、高斯核函数等，其中高斯核函数最常用，可以将数据映射到无穷维，也叫做径向基函数（Radial Basis Function 简称 RBF）。

SVM 模型有两个非常重要的参数 C 与 Gamma。其中 C 是惩罚系数，即对误差的宽容度。C 越高，说明越不能容忍出现误差，容易过拟合。C 越小，容易欠拟合。C 过大或过小，泛化能力变差。Gamma 是选择 RBF 函数作为 kernel 后，该函数自带的一个参数。隐含地决定了数据映射到新的特征空间后的分布，Gamma 越大，支持向量越少，Gamma 值越小，支持向量越多。支持向量的个数影响训练与预测的速度。

三、设备

本实验所需硬件设备如下：

- 1、PC 机一台。

2、0B 人工智能科教平台软件。

四、实践操作

1、点击教材底部进入“机器学习”按钮，选择第一行第一列的“图像识别相关实验”按钮，进入图像识别实验选择页面。点击“实验三”按钮，进入实验页面。

2、软件界面如下：



界面左边是几个选择项，用以选择图像数据，制作数据集和相应的标签。

界面右边是训练所需要调整的参数，可以执行选择设置。

3、使用步骤：

- 1) 选择图像文件，即选择之前采集好的图像数据所在文件夹；
- 2) 数据保存路径，即选择所制作数据集将要保存的路径；
- 3) 开始制作，点击该按钮，即开始利用选择好的图像制作数据集和标签文件；
- 4) 选择训练数据，即选择上一步骤制作好的数据集
- 5) 选择数据标签，即选择上一步骤制作好的数据集对应标签文件
- 6) 选择保存路径，即选择即将训练的模型保存路径
- 7) 自行修改“选择核函数，惩罚系数，Gamma 系数，迭代次数”等参数
- 8) 开始训练，点击即可开始以上一步选择的参数进行模型训练，训练好的模型将以参数名及其数值进行命名，保存在第 6 步选择的文件夹中

4、修改“选择核函数，惩罚系数，Gamma 系数，迭代次数”的参数，得到不同的模型

并保存，以便接下来实验的使用

五、实践数据

- 1、按照实验步骤操作 3，得到一个.xml 格式的模型文件；
- 2、选择不同参数，得到不同模型；

六、数据分析

选择不同参数，得到不同模型，以便接下来实验使用

七、结论与建议

On-Bright confidential

应用实践四 手写数字识别模型验证

一、实验目的

- 1、了解模型验证的流程。
- 2、了解不同数据集和参数下训练模型的差异。
- 3、了解模型在实际使用中的泛化性能。

二、实验原理

1、数据集内容：

数据集可以自己采集制作，也可以从 MNIST 官网下载如下文件：

| | |
|-----------------------------|-----------------------|
| train-images-idx3-ubyte.gz: | 训练集图像 (9912422 bytes) |
| train-labels-idx1-ubyte.gz: | 训练集标签 (28881 bytes) |
| test-images-idx3-ubyte.gz: | 测试集图像 (1648877 bytes) |
| test-labels-idx1-ubyte.gz: | 测试集标签 (4542 bytes) |

gz 是压缩文件，首先要解压这些文件

idx3 的数字表示数据维度，图像数据为 3 维，标签数据为 1 维。

图像的 3 个维度分别为：

训练集图像 28pix×28pix×60000 张

测试集图像 28pix×28pix×10000 张

2、数据集解析：

1) 标签数据：

TRAINING SET LABEL FILE (train-labels-idx1-ubyte):

| [offset] | [type] | [value] | [description] |
|----------|----------------|------------------|--------------------------|
| 0000 | 32 bit integer | 0x00000801(2049) | magic number (MSB first) |
| 0004 | 32 bit integer | 60000 | number of items |
| 0008 | unsigned byte | ?? | label |
| 0009 | unsigned byte | ?? | label |
| | | | |
| xxxx | unsigned byte | ?? | label |

The labels values are 0 to 9.

TEST SET LABEL FILE (t10k-labels-idx1-ubyte):

| [offset] | [type] | [value] | [description] |
|----------|----------------|------------------|--------------------------|
| 0000 | 32 bit integer | 0x00000801(2049) | magic number (MSB first) |
| 0004 | 32 bit integer | 10000 | number of items |
| 0008 | unsigned byte | ?? | label |
| 0009 | unsigned byte | ?? | label |
| | | | |
| xxxx | unsigned byte | ?? | label |

The labels values are 0 to 9.

offset 是一个标号，记录这个字段起始位置在这串数据的第几个字节。

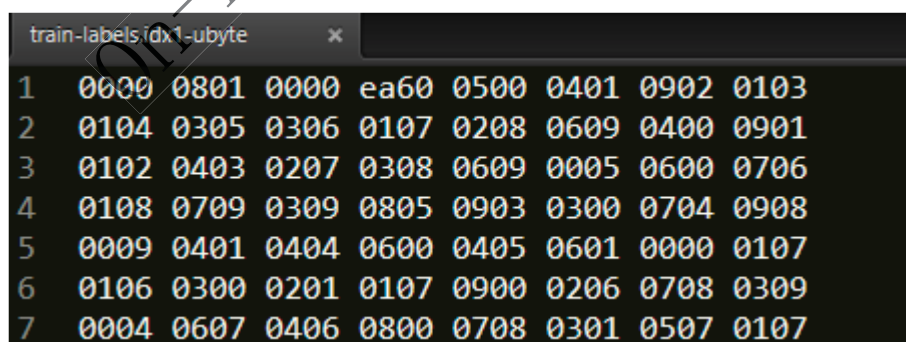
例如第一行，从 0000 的位置开始，(这个 0000 是 16 进制，每个 16 进制位表示是 4bit)，从 0000 开始到 0003 结束，实际上是 4 字节 = 32bit，所以可以看到这个 offset 字段的数据类型 type 是 32 位整数。

value 指的是这个字段数据为 0000 0801，用十进制表示是 2049。

description 是这个字段数据的含义，这个字段存放的 801（也就是十进制 2049），意思是魔数 (magic number)，魔数概念如下：

魔数：很多类型的文件，其起始的几个字节的内容是固定的（或是有意填充，或是本就如此）。根据这几个字节的内容就可以确定文件类型，因此这几个字节的内容被称为魔数 (magic number)。此外在一些程序代码中，程序员常常将在代码中出现但没有解释的数字常量或字符串称为魔数 (magic number) 或魔字符串。

用 sublime 编辑器打开这个训练标签的 ubyte 文件：



对于第一行：

第一个数据为 32 位，读取前 32bit，00000801（即 $8 \times 4\text{bit} = 32\text{bit}$ ）

第二个数据也是 32 位，继续读取 32bit，0000ea60（十进制 60000）

从第三个数据开始，就是真正要用的标签了。类型是无符号数，是 1 个字节，往后读两

个数字，即 05，可以看出这一行数据数据存放的标签以此为 5，0，4，1，9，2，1，3。

对于图像数据：

TRAINING SET IMAGE FILE (train-images-idx3-ubyte):

| [offset] | [type] | [value] | [description] |
|----------|----------------|------------------|-------------------|
| 0000 | 32 bit integer | 0x00000803(2051) | magic number |
| 0004 | 32 bit integer | 60000 | number of images |
| 0008 | 32 bit integer | 28 | number of rows |
| 0012 | 32 bit integer | 28 | number of columns |
| 0016 | unsigned byte | ?? | pixel |
| 0017 | unsigned byte | ?? | pixel |
| | | | |
| xxxx | unsigned byte | ?? | pixel |

TEST SET IMAGE FILE (t10k-images-idx3-ubyte):

| [offset] | [type] | [value] | [description] |
|----------|----------------|------------------|-------------------|
| 0000 | 32 bit integer | 0x00000803(2051) | magic number |
| 0004 | 32 bit integer | 10000 | number of images |
| 0008 | 32 bit integer | 28 | number of rows |
| 0012 | 32 bit integer | 28 | number of columns |
| 0016 | unsigned byte | ?? | pixel |
| 0017 | unsigned byte | ?? | pixel |
| | | | |
| xxxx | unsigned byte | ?? | pixel |

与标签数据的区别在于，每个无符号数表示一个像素，可以根据 description 看出，图像是 28×28 的，往后每 28×28 个像素位为一幅图像。读取出来是一行数据，之后恢复成方形矩阵就可以了。

三、 实验设置

本实验所需硬件设备如下：

- 1、PC 机一台。
- 2、OB 人工智能科教平台软件。

四、 实验操作

- 1、点击教材底部进入“机器学习”按钮，选择第一行第一列的“图像识别相关实验”按钮，进入图像识别实验选择页面。点击“实验四”按钮，进入实验页面。

2、软件界面如下：



界面左边是几个选择项，用以选择图像数据，制作数据集和相应的标签。

界面右边是实际测试输入输出。

3、使用步骤：

- 9) 选择图像文件，即选择之前采集好的图像数据所在文件夹；
 - 10) 数据保存路径，即选择所制作数据集将要保存的路径；
 - 11) 开始制作，点击该按钮，即开始利用选择好的图像制作验证集和标签文件；
 - 12) 选择验证数据，即选择上一步骤制作好的验证集
 - 13) 选择验证标签，即选择上一步骤制作好的验证集对应标签文件
 - 14) 选择模型，即选择上一个实验中训练好的模型
 - 15) 开始验证，点击即可开始验证上一个实验训练好的模型，测试训练精度
 - 16) 在黑色区域下方调整好画笔大小，然后在黑色区域中任意书写 0~9 中的一个数字，点击识别当前数字，可在下方显示识别结果
- 4、选择不同参数下训练好的模型，验证模型精度，尽可能找出精度最高的那个模型，精度要求不低于 96%
 - 5、选择不同画笔大小，在黑色区域写下相同数字，测试识别结果准确性。
 - 6、选择实验二中得到 CNN 模型，在黑色区域写下相同数字，测试识别结果准确性，并和 SVM 算法训练得到的结果进行对比分析。

五、 实验数据

- 1、按照实验步骤操作 3，得到一个.xml 格式模型文件的识别精度；
- 2、选择不同模型，得到相应的识别精度；
- 3、在黑色区域测试模型的实际使用精度；

六、 数据分析

- 1、分析不同模型精度不同的原因
- 2、分析不同画笔大小，模型实际测试精度差异产生的原因

七、 结论与建议

On-Bright confidential

应用实践五 使用图传模组采集测试用图像数据

一、实践目的

- 1、了解图传模组的使用方法。
- 2、掌握使用图传模组拍摄视频的方法。
- 3、使用图传模组采集后续实验需要的训练、测试数据。

二、实践原理

摄像图传模组：主要部分由绿色的图传板和黑色的摄像头模块组成，其中摄像头模块包括镜头和 CCD 传感器。

图传模组通过 WiFi 和 PC 相连，进而将视频码流数据传输到 PC 中。OB 人工智能科教平台通过对视频码流数据进行解码，构建完整的图片，通过定时器按照设置好的帧率将图片写入本地硬盘中。



三、设备

本实验所需硬件设备如下：

- 1、OB 摄像图传模块



- 2、可接收无线 WiFi 信号的 PC 机一台。
- 3、OB 人工智能科教平台软件。

使用 OB 图传模组连接 PC:

- 1、打开图传模组电源。
- 2、在 PC 中的“网络与 Internet 设置”中找到“OB_720P-XXX”的 WiFi 信号，并点击连接。
- 3、连接成功后，即可打开 OB 人工智能科教平台软件。

注意：需将“OB_720P-XXX”信号的网络连接属性设置为“专用”。

四、实践操作

(本实验推荐两人以上协同操作)

- 1、点击教材底部进入“机器学习”按钮，选择第一行第二列的“人形跟踪”按钮，进入人形跟踪实验选择页面。
- 2、点击“实验一”按钮，进入实验页面。点击“连接图传”按钮，等待图像画面出现在教材页面中。
- 3、将 OB 图传模块设置在合适的位置，确保能够拍摄到完整的人形。输入图像数据存储的路径，点击“录制图像数据”按钮，开始使用 OB 图传录制测试用图像数据。点击“开始”按钮，系统会默认录制 5 秒钟、每秒 10 帧的视频，以上两个参数可自行修改。
- 4、需要录制的视频包括以下内容：
用于人形跟踪的测试视频：
 - ①视频中的所有帧都应包含有待跟踪的人形。
 - ②至少应有三种以上不同的场景，如：光照变化，部分遮挡，尺度变化。（即跟踪目标与摄像图传模块的距离）
 - ③应包含短视频和长视频两种情况。
- 5、录制结束则会在第三步中输入的存储路径中找到以录制时间为命名的数据文件夹，可自行前往查看。

五、实践数据

统计采集的视频数据分割结果。

人形跟踪的测试视频：

| 场景 | 采集的视频时间长度、帧率 | 分割出的图片数量 |
|------|--------------|----------|
| 光照变化 | 5s, 30fps | 143 张 |
| 部分遮挡 | 5s, 30fps | 140 张 |
| 尺度变化 | 10s, 30fps | 283 张 |
| 短时跟踪 | 5s, 10fps | 45 张 |
| 长时跟踪 | 10s, 10fps | 100 张 |

六、数据分析

由于视频录制采用的是软件定时，因此大概会有 $0.5s \sim 1s$ 左右的定时误差，实际反映到相应的帧率上，则会有 $10 \sim 20$ 张左右的图片数量误差，且帧率越小，误差越小。

七、结论与建议

跟踪算法测试数据的采集上建议使用短时（ $5 \sim 6s$ ）低帧率、高帧率（10fps、30fps）的设置，这样可以测试跟踪算法的时空鲁棒性和对运动模糊的处理能力。

手势姿态训练数据的采集上建议使用长时（10s）中等帧率（20fps）的设置，保证采集的数据足够多的同时能够保证成像质量，不会出现运动模糊的情况。

应用实践六 人形跟踪使用的图像特征与模型

一、实践目的

- 1、了解图像处理中的 RGB 颜色特征和 HOG 特征原理及特点。
- 2、了解使用 RGB + HOG 特征训练线性回归模型的原理。

二、实践原理

- 1、本实验的基本流程如下图所示：



- (一) 使用 OB 摄像图传模块拍摄指定图片（实验一）。
- (二) 使用 OB 人工智能科教平台进行提取拍摄图像特征及构建模型的实验。

- 2、RGB 颜色通道原理：

RGB 即所谓的三原色：红、绿、蓝。任何彩色数字图像都是由红、绿、蓝三种颜色构成，我们称之为 RGB 三通道，如图 1。在计算机中，颜色的亮度是用数字来定义的，常见的亮度范围介于 $[0\sim 255]$ ，之间，其中，0 为最暗，255 最亮。因此，彩色数字图像的每一个点（像素），都是由一个三维向量来表示（如 $[255, 255, 255]$ ，即该像素点为白色），表示这个像素点是由亮度均为 255 的红绿蓝三种颜色通道组成的。

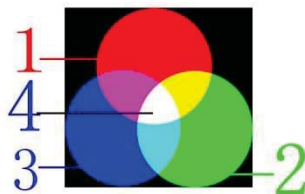


图 1 RGB 三通道

- 3、HOG 特征原理：

HOG 特征，全称方向梯度直方图（Histogram of Oriented Gradient, HOG），它是通过计算和统计图像局部区域的梯度方向直方图来构成特征向量，如图 2 所示。HOG 特征结合各种机器学习算法已经被广泛应用于图像识别中，尤其在行人检测中获得了极大的成功。

更详细的数学原理可以参考论文：

《Histograms of oriented gradients for human detection》

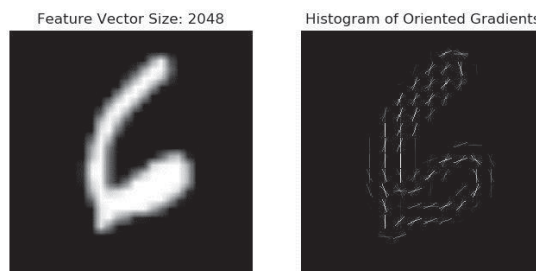


图 2 HOG 特征示例

4、线性回归模型——岭回归的原理：

岭回归（ridge regression）实质上是改良的最小二乘法，它通过舍弃最小二乘法中的无偏性来获得更可靠，跟符合实际的数学模型，它对病态数据（如传感器的错误测量值等）的拟合能力要强于最小二乘法。岭回归的解析解如下：

$$w = (X^T X + \alpha I)^{-1} X^T y$$

其中 w 即为回归模型， X 为用于计算回归模型的数据， I 为单位矩阵， y 为数据的标签。

更详细的数学原理推导可查看：<https://baike.baidu.com/item/岭回归/554917?fr=aladdin>

三、设备

本次实验所需要的设备：

- 1、安装了 OB 智能科教平台的 PC。

四、实践操作

前置实验：由实验一，已经获得大量带有完整人形的图片，请选择这些图片数据中的包含完整人形的图片进行本实验（不少于两张）。

实验一、RGB 颜色特征的特点

1、打开 OB 人工智能科教平台，选择下方“机器学习”按钮，点击第一个图标，进入机器视觉相关实验，点击第二列的人形图标，进入实验选择页面，点击选择“实验二”，进入实验二页面。

2、点击 ，选择待测试图片，点击 选择数据存储路径，注意，若未选择，系统不会保存实验数据！

3、等待测试图片出现后，点击“统计 RGB 三通道像素值分布”按钮，画面右侧将会显示粗

略的统计结果，同时相应数据将会自动保存到 XXX_RGB_hist.txt 文件中。

4、点击“旋转图像”按钮，图像旋转后，再次点击“统计 RGB 三通道像素值分布”按钮，记录观察右侧粗略统计结果有何变化。

5、再次点击“旋转图像”按钮，使图像返回正常显示，点击“调整图像亮度”按钮，亮和暗至少各执行一次，每执行一次“亮”或“暗”操作，都要点击一次“统计 RGB 三通道像素值分布”，记录观察右侧粗略统计结果有何变化。

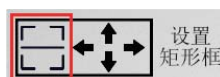


6、选取后续图片，重复执行 2~5 实验步骤。

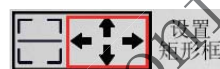
实验二、HOG 颜色特征的特点

1、执行实验一的步骤 2。

2、等待测试图片出现后，点击“设置矩形框”按钮的左侧矩形框按钮，



通过点击右侧“上下左右”按钮调整绿色矩形框的位置，使绿色框能够覆盖到图像中的人体。



3、矩形框设置完成后，点击“计算 HOG 特征”按钮，画面右侧会出现计算得到的 HOG 特征，同时 HOG 特征数据将会自动保存到 XXX_hog_feature.txt 文件中。

4、点击“旋转图像”按钮，图像旋转后，再次点击“计算 HOG 特征”按钮，记录观察右侧特征图有何变化。

5、再次点击“旋转图像”按钮，使图像返回正常显示，点击“调整图像亮度”按钮，亮和暗至少各执行一次，每执行一次“亮”或“暗”操作，都要点击一次“计算 HOG 特征”，记录观察右侧特征图有何变化。

6、选取后续图片，重复执行 2~5 实验步骤。注意，重复执行的时候请确保框选同一个人。

实验三、使用 RGB + HOG 特征构建单张图片的岭回归模型

1、执行实验一的步骤 2。

2、执行实验二的步骤 2。

3、矩形框设置完成后，点击“构建线性模型”按钮，画面右侧会出现计算得到的基于 RGB + HOG 特征的岭回归模型图像，同时岭回归模型数据将会自动保存到 XXX_model.txt 文件中。

4、点击“旋转图像”按钮，图像旋转后，再次点击“构建线性模型”按钮，记录观察右侧

岭回归模型有何变化。

5、再次点击“旋转图像”按钮，使图像返回正常显示，点击“调整图像亮度”按钮，亮和暗至少各执行一次，每执行一次“亮”或“暗”操作，都要点击一次“构建线性模型”，记录观察岭回归模型有何变化。

6、选取后续图片，重复执行 2~5 实验步骤。注意，重复执行的时候请确保框选同一个人。

五、实践数据

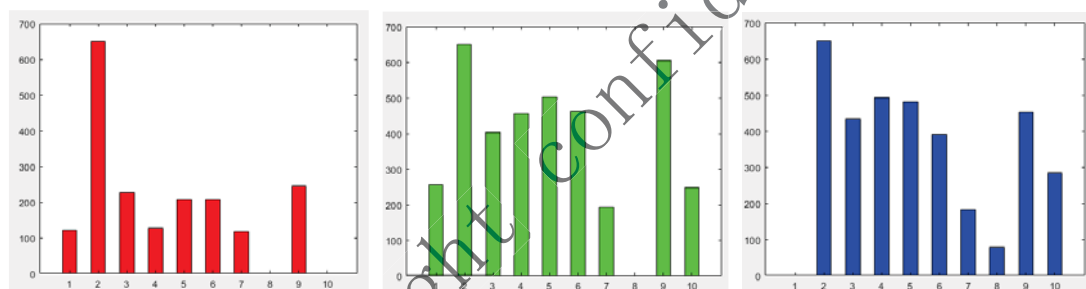
实验一：

使用 Matlab、Excel 等工具，以直方图的形式绘制所有 XXX_RGB_hist.txt 文件中的记录的对应图像的 R、G、B 三个通道的像素值分布。

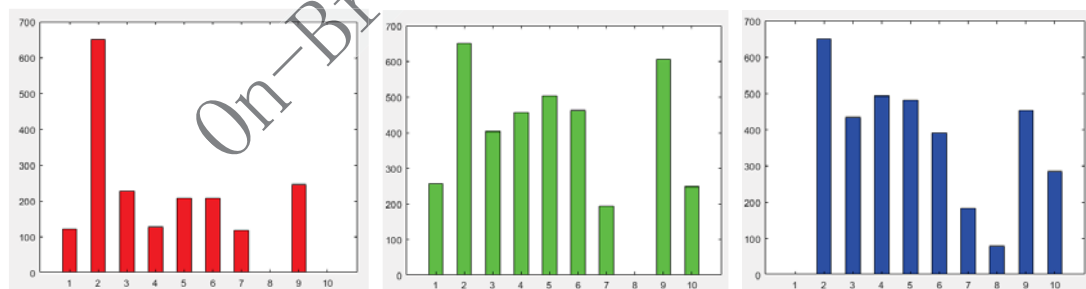
测试图片一：

从左到右分别为 R 分量，G 分量，B 分量：

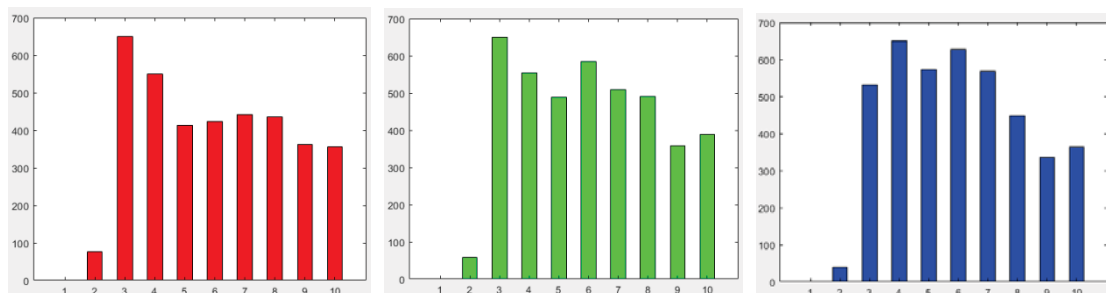
无操作：



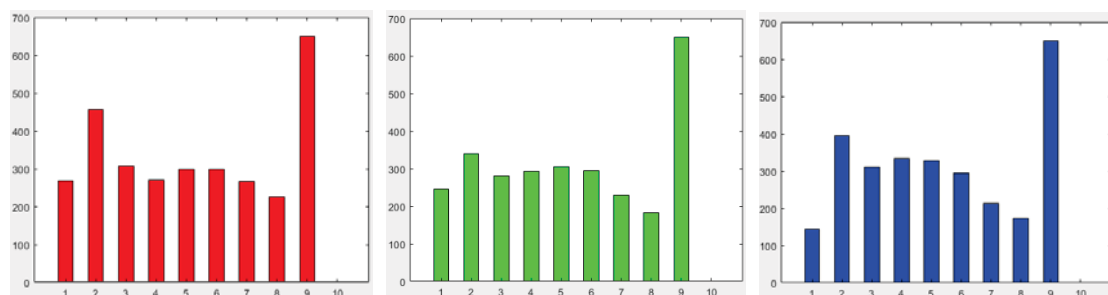
翻转操作：



亮度提升操作：



亮度降低操作：



实验二：

- 1、使用 Matlab、Excel 等工具，选取同一个 XXX_hog_feature.txt 文件中的一到两个 hog 特征向量，计算选取的特征向量之间的欧式距离。
- 2、使用 Matlab、Excel 等工具，选取不同 XXX_hog_feature.txt 文件中一到两个 hog 特征向量，计算选取的特征向量之间的欧式距离。

| 图片名称 | 所做操作（亮/暗/旋转） | 对比操作（亮/暗/旋转） | Hog 特征向量欧式距离 |
|-------------|--------------|--------------|--------------|
| Test3. JPG | 无操作 | 亮 | 1. 1657 |
| Test3. JPG | 无操作 | 暗 | 4. 5230 |
| Test3. JPG | 无操作 | 翻转 | 7. 4305 |
| Test63. JPG | 无操作 | 亮 | 3. 5241 |
| Test63. JPG | 无操作 | 暗 | 2. 9350 |
| Test63. JPG | 无操作 | 翻转 | 6. 6497 |

表 1 同一个 XXX_hog_feature.txt 文件中的 hog 特征向量对比

| 图片名称 | 所做操作（亮/暗/旋转） | 对比图像名称 | 对比图像所做操作（亮/暗/旋转） | Hog 特征向量欧式距离 |
|------------|--------------|-------------|------------------|--------------|
| Test3. JPG | 无操作 | Test63. JPG | 无操作 | 6. 9273 |
| Test3. JPG | 亮 | Test63. JPG | 暗 | 7. 8686 |
| Test3. JPG | 暗 | Test63. JPG | 暗 | 7. 8377 |

表 2 不同 XXX_hog_feature.txt 文件中的 hog 特征向量对比

实验三：

- 1、使用 Matlab、Excel 等工具，选取同一个 XXX_model.txt 文件中的一到两个模型向量，计算选取的模型向量之间的余弦距离。

2、使用 Matlab、Excel 等工具，选取不同 XXX_model.txt 文件中一到两个模型向量，计算选取的模型向量之间的余弦距离。

| 图片名称 | 所做操作（亮/暗/旋转） | 对比操作（亮/暗/旋转） | 模型向量余弦距离 |
|-------------|--------------|--------------|----------|
| Test26. JPG | 无操作 | 亮 | 1 |
| Test26. JPG | 无操作 | 暗 | 1 |
| Test26. JPG | 无操作 | 翻转 | 0.9742 |
| Test61. JPG | 无操作 | 亮 | 1 |
| Test61. JPG | 无操作 | 暗 | 0.9653 |
| Test61. JPG | 无操作 | 翻转 | 1 |

表 3 同一个 XXX_model.txt 文件中的模型特征向量对比

| 图片名称 | 所做操作（亮/暗/旋转） | 对比图像名称 | 对比图像所做操作（亮/暗/旋转） | 模型向量余弦距离 |
|-------------|--------------|-------------|------------------|----------|
| Test26. JPG | 无操作 | Test61. JPG | 无操作 | 1 |
| Test26. JPG | 亮 | Test61. JPG | 暗 | 1 |
| Test26. JPG | 暗 | Test61. JPG | 暗 | 0.6154 |
| Test26. JPG | 无操作 | Test61. JPG | 翻转 | 1 |

表 4 不同 XXX_model.txt 文件中的模型特征向量对比

六、数据分析

实验一：

分析 RGB 通道像素分布直方图，思考 RGB 特征具有什么特点。

由实验一的数据可以知道：

- 1、当图像翻转时，图像的 RGB 三个分量的像素值分布没有改变；
- 2、提高亮度时，RGB 三个分量的像素值分布整体向高亮度区域聚集。
- 3、降低亮度时，RGB 三个分量的像素值分布整体向低亮度区域聚集。
- 4、综合以上现象，可以得到的得到：RGB 三通道特征对旋转具有不变性，易受光照影响。

实验二：

分析 hog 特征向量的相应数据，思考 hog 特征具有什么特点。

由实验二的数据可以知道：

- 1、图像翻转后，其 HOG 特征向量与原图像 HOG 特征向量差异较大；
- 2、提高亮度时，其 HOG 特征向量与原图像 HOG 特征向量差异比翻转的情况小；
- 3、降低亮度时，其 HOG 特征向量与原图像 HOG 特征向量差异比翻转的情况小；
- 4、综上可以得到的：HOG 特征易受旋转等方向变化的操作影响，但对光照不敏感。

实验三：

分析岭回归模型向量的相应数据，思考基于 RGB+HOG 特征建立的岭回归模型有什么特点。

由实验三的数据可以知道：

虽然采用不同的图片，增加了不同的噪声，但由于使用了 RGB + HOG 两种特征来对同一个目标建立线性模型，从表格数据中可以明显看到 RGB 特征（实验一），HOG 特征（实验二）的优点得到了整合，对同一目标的建立的线性模型在旋转、亮度变化的情况下基本没有差异。

七、结论与建议

RGB + HOG 两种特征的整合可以使图像中的目标对象具有旋转不敏感性，光照不敏感性，适合于基本的图像物体识别等任务。

应用实践七 人形跟踪算法中的机器学习

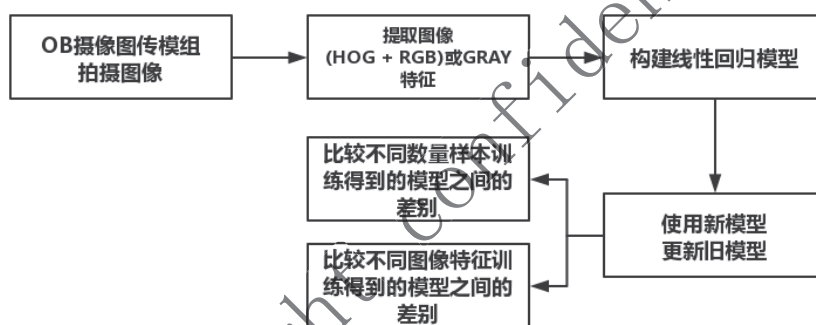
模型训练

一、实践目的

- 1、了解基本的线性回归模型的训练方式。
- 2、了解在不同条件下，训练得到的机器学习模型之间的差异。
- 3、测试使用不同图像特征训练得到的线性回归模型的差异。

二、实践原理

- 1、本实验的基本流程如下图所示：



- (一) 使用指定的图像特征对同一目标构建线性模型。
- (二) 根据不同的图片迭代更新、训练模型。
- (三) 比较不同设置条件下训练所得模型的差异性。

- 2、线性模型的训练方式：

本实验使用线性加权的方式训练、更新模型，公式如下：

$$w_{final} = w_{old} + \beta w_{new}$$

其中， w_{final} 即为最终训练得到的模型， w_{new} 为当前训练样本构建得到的模型， β 为学习率，指的是使用新模型更新到旧模型所占的权重。

- 3、GRAY 特征

GRAY 特征即为灰度特征，与实验二中介绍的 RGB 特征不同，灰度图像只有一个颜色通道，图像上的每个像素点只有一个值，而非 RGB 图像的三个值。灰度特征可由 RGB 特征转

换而来，具体公式有以下几种：

- 1.浮点算法: $\text{Gray} = R \cdot 0.3 + G \cdot 0.59 + B \cdot 0.11$
- 2.整数方法: $\text{Gray} = (R \cdot 30 + G \cdot 59 + B \cdot 11) / 100$
- 3.移位方法: $\text{Gray} = (R \cdot 76 + G \cdot 151 + B \cdot 28) \gg 8;$
- 4.平均值法: $\text{Gray} = (R + G + B) / 3;$

三、设备

本次实验所需要的设备：

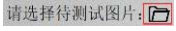

- 1、安装了 OB 智能科教平台的 PC。

四、实践操作

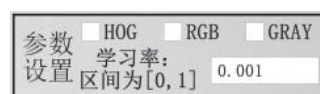
前置实验：由实验一与二，已经了解了图像采集和部分人形跟踪中使用的图像特征、以及如何构建相应的线性模型等知识。

实验一、使用 HOG+RGB 特征进行模型训练

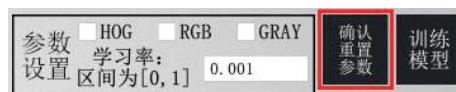
1、打开 OB 人工智能科教平台，选择下方“机器学习”按钮，点击第一个图标，进入机器视觉相关实验，点击第二列的人形图标，进入实验选择页面，点击选择“实验二”，进入实验二页面。

2、点击 ，选择待测试图片，点击 选择数据存储路径，**注意，若未选择，系统不会保存实验数据！**

3、等待测试图片出现后，进行参数设置：



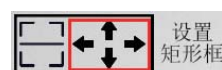
勾选 HOG、RGB 特征，设置学习率（0~1）之间。然后点击“确认重置参数”按钮。



4、点击“设置矩形框”按钮的左侧矩形框按钮，



通过点击右侧“上下左右”按钮调整绿色矩形框的位置，**使绿色框能够覆盖到图像中的人体。**



- 5、点击“训练模型”按钮，界面中右侧会出现构建得到而线性模型。
- 6、**更换新的图片**，重复执行步骤 4~5 共 10 次，得到 10 个样本的训练得到的模型数据，保存于“存储路径 /XXX_model.txt”中。
- 7、改变参数设置中的学习率参数，或者对图片进行“亮”、“暗”、“旋转”操作（改变参数和处理图像操作二选一，至少做一个），重复 4~6 的操作。
- 8.使用 MATLAB、Excel 等工具，计算**最终输出的模型**与十张训练图片**各自构建的线性模型**的余弦距离。

注意：选定了 10 张训练图片后不要在后续实验中更换，以保证实验的准确性。

实验二、使用 GRAY 特征进行模型训练

- 1、执行实验一的步骤 1~2。
- 2、在选取参数阶段，选择 GRAY 特征和学习率，点击“确认重置参数”按钮。
- 3、执行实验一的步骤 4~8。

注意：选定了 10 张训练图片后不要在后续实验中更换，以保证实验的准确性。

五、实践数据

实验一：使用 HOG + RGB 特征进行模型训练

| 图片名称 | 参数设置 (亮/暗/旋转/ 学习率)/具体 参数 | 原模型与 最终模型的 余弦距离 | 图片名称 | 参数设置 (亮/暗/旋转/ 学习率)/具体 参数 | 原模型与 最终模型的 余弦距离 |
|--------|-----------------------------------|-----------------------|--------|-----------------------------------|-----------------------|
| 12.jpg | 学习率/0.001 | 0.0003 | 17.jpg | 学习率/0.001 | 1 |
| 13.jpg | 学习率/0.001 | 0.9276 | 18.jpg | 学习率/0.001 | 1.0028 |
| 14.jpg | 学习率/0.001 | 1 | 19.jpg | 学习率/0.001 | 0.0167 |
| 15.jpg | 学习率/0.001 | 1 | 20.jpg | 学习率/0.001 | 1 |
| 16.jpg | 学习率/0.001 | 1 | 21.jpg | 学习率/0.001 | 1.003 |

| 图片名称 | 参数设置 (亮/暗/旋转/ 学习率)/具体 参数 | 原模型与 最终模型的 余弦距离 | 图片名称 | 参数设置 (亮/暗/旋转/ 学习率)/具体 参数 | 原模型与 最终模型的 余弦距离 |
|------|-----------------------------------|-----------------------|------|-----------------------------------|-----------------------|
| | | | | | |

| | | | | | |
|--------|---------|--------|--------|---------|--------|
| 12.jpg | 学习率/0.1 | 0.8438 | 17.jpg | 学习率/0.1 | 1.1022 |
| 13.jpg | 学习率/0.1 | 1.1050 | 18.jpg | 学习率/0.1 | 1.1114 |
| 14.jpg | 学习率/0.1 | 0.4939 | 19.jpg | 学习率/0.1 | 0.8414 |
| 15.jpg | 学习率/0.1 | 1 | 20.jpg | 学习率/0.1 | 0.9732 |
| 16.jpg | 学习率/0.1 | 0.4939 | 21.jpg | 学习率/0.1 | 0.4940 |

实验二：使用 GRAY 特征进行模型训练

| 图片名称 | 参数设置 (亮/暗/旋转/ 学习率)/具体 参数 | 原模型与 最终模型的 余弦距离 | 图片名称 | 参数设置 (亮/暗/旋转/ 学习率)/具体 参数 | 原模型与 最终模型的 余弦距离 |
|--------|-----------------------------------|-----------------------|--------|-----------------------------------|-----------------------|
| 12.jpg | 学习率/0.001 | 0.0003 | 17.jpg | 学习率/0.001 | 0.9980 |
| 13.jpg | 学习率/0.001 | 0.9981 | 18.jpg | 学习率/0.001 | 0.9979 |
| 14.jpg | 学习率/0.001 | 0.7435 | 19.jpg | 学习率/0.001 | 0.9981 |
| 15.jpg | 学习率/0.001 | 0.9981 | 20.jpg | 学习率/0.001 | 1 |
| 16.jpg | 学习率/0.001 | 0.7435 | 21.jpg | 学习率/0.001 | 0.7452 |

| 图片名称 | 参数设置 (亮/暗/旋转/ 学习率)/具体 参数 | 原模型与 最终模型的 余弦距离 | 图片名称 | 参数设置 (亮/暗/旋转/ 学习率)/具体 参数 | 原模型与 最终模型的 余弦距离 |
|--------|-----------------------------------|-----------------------|--------|-----------------------------------|-----------------------|
| 12.jpg | 学习率/0.1 | 0.9999 | 17.jpg | 学习率/0.1 | 1.0001 |
| 13.jpg | 学习率/0.1 | 0.9995 | 18.jpg | 学习率/0.1 | 1.0001 |
| 14.jpg | 学习率/0.1 | 0 | 19.jpg | 学习率/0.1 | 0.9999 |
| 15.jpg | 学习率/0.1 | 0.9995 | 20.jpg | 学习率/0.1 | 1 |
| 16.jpg | 学习率/0.1 | 0.0070 | 21.jpg | 学习率/0.1 | 0 |

六、数据分析

实验一：

当使用 HOG+RGB 特征，且学习率较小时（0.001），训练得到的最终模型能够和绝大部分训练样本匹配（余弦距离 ≥ 0.9 ）；当学习率提高时（0.1），训练得到的最终模型与训练样本匹配度明显下降。

实验二：

当使用 GRAY 特征，且学习率较小时（0.001），训练得到的最终模型和绝大部分训练样本匹配程度（余弦距离 ≥ 0.9 ）；当学习率提高时（0.1），训练得到的最终模型与训练样本匹配度明显上升，但仍不及 HOG+RGB 特征。

七、结论与建议

（1）如果使用 HOG+RGB 特征，则训练的学习率不能设置太高。因为 HOG 特征主要是描述边缘信息，过高的学习率会导致边缘信息丢失。

（2）如果使用 GRAY 特征，则训练的学习率可以设置高一些，因为 GRAY 特征本身属于颜色空间中的一种，较高的学习率能够让模型快速适应颜色的变化。

On-Bright confidential

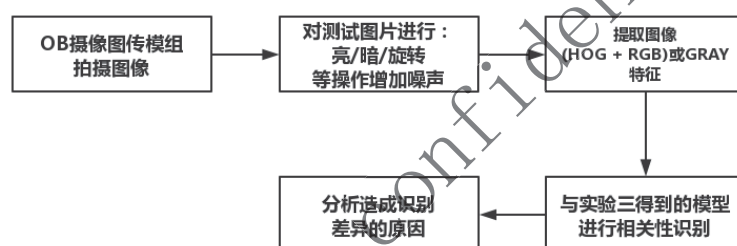
应用实践八 人形跟踪中的模型在人形识别中的测试

一、实践目的

- 1、了解线性回归模型的测试方式。
- 2、了解模型和测试样本的相关性原理
- 3、测试并分析实验三得到的模型对不同噪声下的测试样本的在识别结果上的差异。

二、实践原理

- 1、本实验的基本流程如下图所示：



- (四) 对实验三未使用的图片增加噪声（或者使用原图）。
- (五) 提取不同的图像特征，并和实验三训练得到的模型进行相关性识别。
- (六) 比较不同设置条件下训练所得模型在不同噪声下识别结果的差异性。

- 2、使用相关性滤波的跟踪算法：

实验三中得到了通过线性加权的方式训练得到的模型 w_{final} ，存储于 XXX_latest_model.txt 文件中，同时生成的文件还包括：

XXX_latest_alpha_c1.txt，XXX_latest_alpha_c2.txt，XXX_hann.txt，XXX_tmpl_params.txt，一共有 5 个与模型相关的文件，具体解释如下文所述。

（一）相关滤波

在信号处理领域，使用相关性（correlation）来描述两个信号之间的联系，简而言之，相关性越高，两个信号就越相似。将这种理论引入到基于机器学习的视觉跟踪领域，就逐渐演变为基于相关滤波（correlation filter）的视觉跟踪算法。

算法的思路很明确：对特定目标训练一个模型 w ，对于一张新的、未训练过的图片 X ，计算两者之间的相关性，如果相关性高，说明新图片中的内容包含有算法需要的特定目标。基于这种思路，产生了以 KCF 算法为开端的一类相关滤波视觉跟踪算法。

（二）KCF 算法

KCF 算法是一种典型的基于机器学习的视觉跟踪算法，其图像特征使用了实验二介绍的 HOG+RGB，训练方法则为实验二与实验三介绍的岭回归和线性加权。

本实验介绍 KCF 算法的如何验证新的、未训练的图片。现对实验二和三的内容进一步补充：

- ① 由于视觉跟踪任务每次只有新的一张图片，训练样本不足，故对图片采取循环位移的方式增加训练的样本。该操作将会引入边界问题，因而需要加入汉明窗函数抑制边界效应，汉明窗的参数保存于 XXX_hann.txt 中。同时，使用循环卷积的操作也能将模型的训练转换到傅里叶域，进而加速模型的运算速度。
- ② 为增强 HOG+RGB 特征的表达能力，对其进行核化操作，即从将计算得到的图像特征通过核函数从低维度映射到高维度，映射函数的参数保存于 XXX_latest_alpha_c1.txt，XXX_latest_alpha_c2.txt 文件中。一些其它的训练信息如：模型维度、尺度变换因子等参数则保存于 XXX_tmpl_params.txt 中。

KCF 算法对于新的图片 z 的计算相关性的公式为：

$$f(z) = w^T z = \sum_i \alpha \varphi^T(x_i) \varphi(z_i)$$

其中 φ 为提取图像特征的运算， α 为核化运算的参数。

更具体的数学推导可见论文：

《High-Speed Tracking with Kernelized Correlation Filters》

三、设备

本次实验所需要的设备：



- 1、安装了 OB 智能科教平台的 PC。

四、实践操作

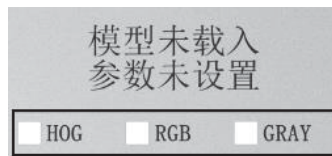
前置实验：由前面的实验，已经了解了图像采集和部分人形跟踪中使用的图像特征、以及如何构建相应的线性模型、训练等知识。

实验一、测试使用 HOG+RGB 特征训练的模型


1、打开 OB 人工智能科教平台，选择下方“机器学习”按钮，点击第一个图标，进入机器视觉相关实验，点击第二列的人形图标，进入实验选择页面，点击选择“实验四”，进入实验四页面。

2、点击 ，选择**实验三中未使用的待测试图片**，点击 选择数据存储路径，**注意，若未选择，系统不会保存实验数据！**

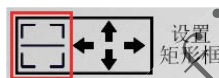
3、等待测试图片出现后，进行参数设置：



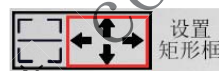
勾选 HOG、RGB 特征。

4、点击 按钮，选择实验三种保存的模型的路径，**路径正确则会显示“模型载入成功”。**

5、点击“设置矩形框”按钮的左侧矩形框按钮，



通过点击右侧“上下左右”按钮调整绿色矩形框的位置，**使绿色框能够覆盖到图像中的人体。**



6、点击“测试模型”按钮：



画面右边会出现相关性运算得到的结果，**白色的亮点所处的位置**即为矩形框所围住的画面中，**与保存的模型相关性最高的目标的位置。**

7、**更换新的图片**，重复执行步骤 5~6，至少使用**同一目标在不同背景下**的图片 5 张，设置的矩形框可有所偏移，或者对图片进行“亮”、“暗”、“旋转”操作，进而能够观察到白色亮点的位置变化。相关性运算得到的结果保存于 XXX_res_map.txt 中。

8.使用 MATLAB 计算 5 个相关性运算结果中的各自的最大值，分析不同的背景，操作对模型的相关性识别运算所带来的影响。

实验二、测试使用 GRAY 特征训练的模型

4、执行实验一的步骤 1 ~ 2。

5、在选取参数阶段，选择 GRAY 特征，点击“确认重置参数”按钮。

6、执行实验一的步骤 4~8。

五、实践数据

数据存储格式：

在 XXX_res_map.txt 文件中保存的数据实际上是一个矩阵，全选，复制到 MATLAB 中即可。

实验一：测试使用 HOG+RGB 特征训练的模型

| 测试图片名称 | 所做操作 | 相关性（计算结果的最大值） |
|--------|------|---------------|
| 11.jpg | 无 | 0.7941 |
| 22.jpg | 不同背景 | 0.2674 |
| 26.jpg | 亮 | 0.2652 |
| 32.jpg | 暗 | 0.2957 |
| 64.jpg | 不同背景 | 0.2078 |

实验二：测试使用 GRAY 特征训练的模型

| 测试图片名称 | 所做操作 | 相关性（计算结果的最大值） |
|--------|------|---------------|
| 11.jpg | 无 | 0 |
| 22.jpg | 不同背景 | 0 |
| 26.jpg | 亮 | 0 |
| 32.jpg | 暗 | 0 |
| 64.jpg | 不同背景 | 0 |

六、数据分析

实验一：测试使用 HOG+RGB 特征训练的模型

测试图片与训练图片相似时，模型与测试样本的相关性最高。其他情况下相关性有所削弱，但仍然能将矩形框中的特定目标与背景区别出来。这从侧面验证了即使在训练样本较少的情况下 HOG+RGB 特征的组合仍然拥有对特定目标较强的描述能力，能够在一定程度上区

分目标和背景。

实验二：测试使用 GRAY 特征训练的模型

所有测试图片与模型的相关性都很差。这从侧面验证了 GRAY 特征的对图像的表征能力非常弱，根据实验三的测试结果，学习率较低，训练样本不足的原因更凸显了这一不足。

七、结论与建议

（一）使用 HOG+RGB 特征来描述图像中的特定物体具有一定的鲁棒性，对光照，背景变化等情况具有一定的不敏感性。使用这种特征组合训练得到的模型有一定的抗干扰能力，不仅在视觉跟踪的场景中，在某些特定的目标识别场景中也可以应用。

（二）GRAY 特征本身属于颜色特征的一种，容易受到环境变化的影响。在环境变化不大的等特定场景中可以使用，但需要比 HOG+RGB 特征更多的训练数据。

On-Bright confidential

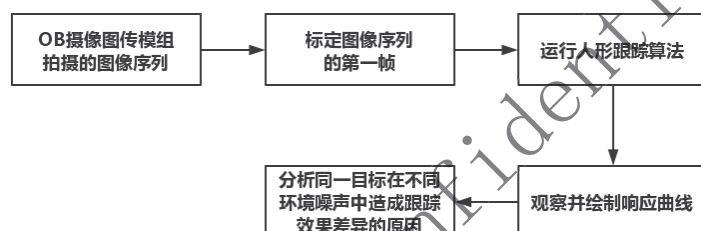
应用实践九 不同环境噪声中人形跟踪算法的测试

一、实践目的

- 1、使用 KCF 算法测试实验一拍摄的人形跟踪视频序列。
- 2、通过实验二、三、四，综合分析不同的环境噪声对人形跟踪效果的影响。

二、实践原理

- 1、本实验的基本流程如下图所示：



- (一) 使用 OB 人工智能科教平台自动标定图像序列的第一帧。
- (二) 运行跟踪算法，观察算法的响应输出。
- (三) 比较并分析不同设置条件，不同噪声下人形跟踪效果的差异性。

- 2、使用相关性滤波的跟踪算法：

本实验的输出为实验四中介绍的 KCF 跟踪算法的相关性输出，其本质是一张三维的峰值图，其峰值的大小，以及有无旁瓣都可以作为跟踪视觉跟踪效果差异性的分析依据。

更具体的数学推导可见论文：

《High-Speed Tracking with Kernelized Correlation Filters》

三、设备

本次实验所需要的设备：

- 1、安装了 OB 智能科教平台的 PC。

四、实践操作

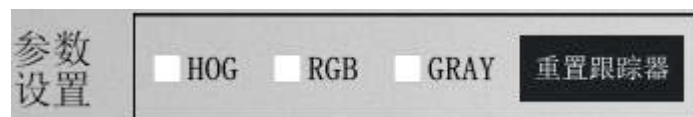
前置实验：由前面的实验，已经了解了图像采集和部分人形跟踪中使用的图像特征、以

及如何构建相应的线性模型、训练、测试等知识。

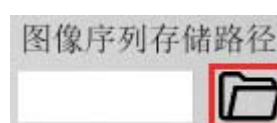
实验一、使用 KCF 算法测试实验一中拍摄的不同背景、噪声下的跟踪序列

1、打开 OB 人工智能科教平台，选择下方“机器学习”按钮，点击第一个图标，进入机器视觉相关实验，点击第二列的人形图标，进入实验选择页面，点击选择“实验五”，进入实验五页面。

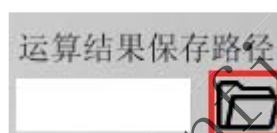
2、点击“参数设置”一栏，选择要使用的图像特征（如 HOG + RGB 或者 GRAY）。



3、点击以下图标选择测试的图像序列：



点击以下图标选择数据存储路径，注意，若未选择，系统不会保存实验数据！



此时，若成功则画面中会出现“图像序列载入成功”，否则则显示“图像序列载入失败”。情检查序列载入路径是否正确，文件必须以（1~n.jpg）格式命名。

4、等待第一帧的测试图片出现后，点击“标定第一帧目标位置”按钮，系统会启动人形检测算法进行标定。如未检测到人体则需要进行手动标定，标定和前置实验中“设置矩形框”操作一致，请参考前置实验。

5、点击“运行人形跟踪算法”按钮，开始进行图像序列测试的左侧矩形框按钮，画面中的右侧会出现当前跟踪目标和实时训练模型的相关响应程度，请注意观察。

6、更换新的图片序列，重复执行步骤 3~5，然后使用同样的图像序列，改变预先设置的图像特征，点击“重置跟踪器”按钮，重复执行以上步骤。相关性运算得到的结果保存于 XXX_exp5_test.txt 中。

7、使用 MATLAB 绘制保存的数据的曲线图，分析不同的背景，操作对模型的相关性识别运算所带来的影响。

五、实践数据

数据存储格式：

在 XXX_exp5_test.txt 文件中保存的数据实际上是一个一维向量，保存的是该测试序列中每一张图片的最高相关响应值，全选，复制到 MATLAB 中，使用响应的绘制函数即可。

六、数据分析

七、结论与建议

On-Bright confidential

应用实践十 人脸识别模块的使用

一、实践目的

- 1、了解如何使用 Python 编程使用 MTCNN 人脸检测神经网络。
- 2、了解如何使用 Python 编程使用 MobileFaceNet 人脸识别神经网络。

二、实践原理

- 1、MTCNN 网络：轻量级人脸检测网络，论文《Joint Face Detection and Alignment using Multi-task Cascaded Convolutional Network》
- 2、MobileFaceNet 网络：轻量级人脸识别网络，论文《MobileFaceNets: Efficient CNNs for Accurate Real-time Face Verification on Mobile Devices》

三、设备

本次实验所需要的设备：

- 1、安装了 OB 智能科教平台使用环境的 PC。

四、实践操作

本实验为程序编写实验，桌面文件夹“语音模块与训练数据”中的“人脸识别”文件夹已包含了 MTCNN、MobileFaceNet 的模型和调用代码，请仔细阅读研究，并基于 opencv、tensorflow（PC 机已配备该环境），使用 Python 编写一个建议的人脸考勤系统。

Python 的编写可使用：PC 中已配置的 IDLE 或者自行安装 Pycharm 社区版。

五、实践数据

六、数据分析

七、结论与建议

应用实践十一 语音识别模块的使用

一、实践目的

- 1、了解如何使用 Python 编程调用 Pyaudio 模块进行音频录制、播报。
- 2、了解如何使用 Python 编程调用科大讯飞的语音识别、语音合成 SDK。

二、实践原理

本实验为程序编写实验，桌面文件夹“语音模块与训练数据”中的“语音识别”文件夹已包含了音频录制、音频播放、讯飞语音识别、语音合成的调用代码，请仔细阅读研究，并基于 Pyaudio（PC 机已配备该环境），使用 Python 编写一个建议的语音交互系统。

三、设备

本次实验所需要的设备：

- 1、安装了 OB 智能科教平台使用环境的 PC。
- 2、可连接互联网的 WiFi、有线环境。

四、实践操作

使用讯飞语音识别模块需要在讯飞官网上进行注册，并获取相应的 ID 和 Key，并在代码相应的位置填入该 ID 和 Key，（具体可见代码注释）讯飞开发者官网为：

<https://www.xfyun.cn/>

音频录制、播放功能则不需要。

五、实践数据

六、数据分析

七、结论与建议