

# 第11章 目标识别

计算机科学与技术学院



# 本次课程内容

1. 目标识别概念

2. 目标识别的问题

3. 目标识别应用领域

4. 目标识别的难点

5. 目标识别主要流程

6. 目标识别的步骤

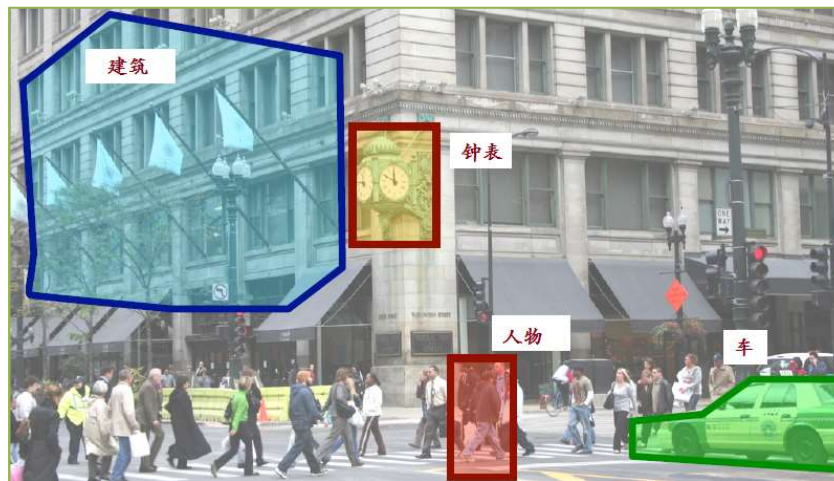
7. 目标识别的案例

8. 传统学习方法特点



# 1. 目标识别概念

- ◆ **目标识别**:是指从一幅图像中, 识别出一个特殊目标或一类目标。
- ◆ 例如: 物体识别



## 目标识别概念

- ◆ 怎样识别图像中物体，如汽车、牛等？



## 目标识别类型

- ◆ 物体识别算法大体有两类：
  - 基于模型方法（或者基于上下文识别的方法）
  - 二维物体识别（三维物体识别）方法



## 2. 目标识别的问题

- ◆ 一个给定的图片进行目标识别，首先要判断有没有目标（检测 **detection**）。
  - 如果没有目标，则检测过程结束。
  - 如果有目标，就要进一步判断有几个目标，目标分别所在的位置，然后对目标进行分割，即判断哪些像素点属于该目标（识别 **Recognition**）。



## 目标的检测与识别



### 3. 目标识别应用领域



计算摄影技术



人机交互



视频监控



安全检查



辅助驾驶





视频监控



智能汽车



场景搜索



基于内容的图像\视频检索



医学图像分析

## 4. 目标识别的难点

◆ 目标识别问题面临**几大难点**:

- 外观变形
- 光照变化
- 快速运动
- 运动模糊
- 背景相似干扰



## 5. 目标识别主要流程

- ◆ 输入图像
- ◆ 特征提取
- ◆ 分类器的训练



## 关于训练样本

- ◆ 训练样本包括正样本和负样本两类：
  - 正样本：指待检目标样本(例如人脸或汽车等)
  - 负样本：指其它不包含目标的样例（如背景等）



## 6. 目标识别的步骤

◆ 目标识别包含以下几个阶段：

- 预处理
- 特征提取
- 建模
- 匹配
- 定位



## (1) 预处理

- ◆ 目标：使得每张图像的表现特性（如颜色分布，整体明暗，尺寸大小等）尽可能的一致
- ◆ 增强：完成模式的采集、模数转换、滤波、消除模糊、减少噪声、纠正几何失真等预处理操作
- ◆ 有时预处理和特征提取同时进行



◆ 典型的预处理方法：

- 直方图均衡及滤波等
- 高斯模糊



## (2) 特征提取

- ◆ 特征提取是物体重要步骤
- ◆ 为了有效地实现分类识别，要对原始数据进行变换，得到合适的特征，即特征的选择和提取过程。





## 特征提取

- ◆ 特征提取是识别方法的一个重要组成部分，经常对图像进行分割后，提取出特征，对这些特征建立索引。
- ◆ 特征类别：颜色、纹理、形状、空间特征、边缘、梯度、深度、运动等。



## 常用局部特征描述子

- ◆ LBP特征
- ◆ HOG特征
- ◆ Shif特征



### (3) 建模

- ◆ 规则：提取同一类物体相同或者相近特点，用于区分
- ◆ 本质：特征与特征之间的空间结构关系
- ◆ 原则：
  - 模型是否适用于当前问题；
  - 模型所需的计算复杂度能够接受
  - 是否有更高效精确或者近似的算法。
- ◆ 模型：可能由物体的特征以及物体特征之间的相互关系定义的，比如位置关系



## (4) 分类器

- ◆ 分类器按特征类型分为数值型分类器和符号型两大类:
  - 数值型分类器:
    - 统计分类器 (统计理论为基础)
    - 模糊分类器 (模糊集理论为基础)
    - 人工神经网络
    - 人工智能分类器 (基于逻辑推理或专家系统结构)
  - 符号型分类器: 目前这类分类器研究还不成熟



# 常用分类器

- ◆ SVM支持向量机
- ◆ K近邻 (KNN)
- ◆ 神经网络 (NN)
- ◆ 支持向量机 (SVM)
- ◆ Boosting(Adaboost等)
- ◆ 隐马尔科夫模型 (HMM)

例如:

- HOG特征+SVM
- Haar+AdaBoost
- LBP+ AdaBoost



## 学习的手段：生成学习与判别学习

- ◆ 生成学习 (Generative learning)

- 目标是学习到符合训练数据的类别模型

- ◆ 判别学习 (discriminative learning)

- 在训练阶段即考虑类别之间的判别信息
- 包括SVMs, Boosting等

判别学习算法比生成学习算法表现出更好的分类性能



## 常用目标识别的策略—词袋方法（Bag of Words, BoW）方法

- ◆ **词袋方法（Bag of Words, BoW）方法：**采用分类方法来识别物体
  - 从图像中提取的特征比作词语，即一幅图像是有一袋子特征组成的。
  - 建立一个特征库，特征库中的特征之间是相互独立的。



## 常用目标识别的策略—结构与部件方法

### ◆ 结构与部件方法：

- 例如弹簧模型将物体部件之间的关系用伸缩的弹簧表示。
- 除了弹簧模型，还有星型结构、层次结构、树状结构等。





## 常用目标识别的策略—生成与鉴别方法

### ◆ 生成与鉴别方法：

- 给定物体类别的条件下，将出现物体的可能性进行打分，利用得分获取检测结果



## 7. 目标识别的案例

- ◆ 物体检测（利用词袋方法物体识别）
- ◆ 人脸检测
- ◆ 人脸识别



## 物体检测问题

### ◆ 基于二分类器



Not a car.



## 物体检测问题

- ◆ 在复杂背景下，通过滑动窗口（sliding windows）搜索感兴趣的物体。

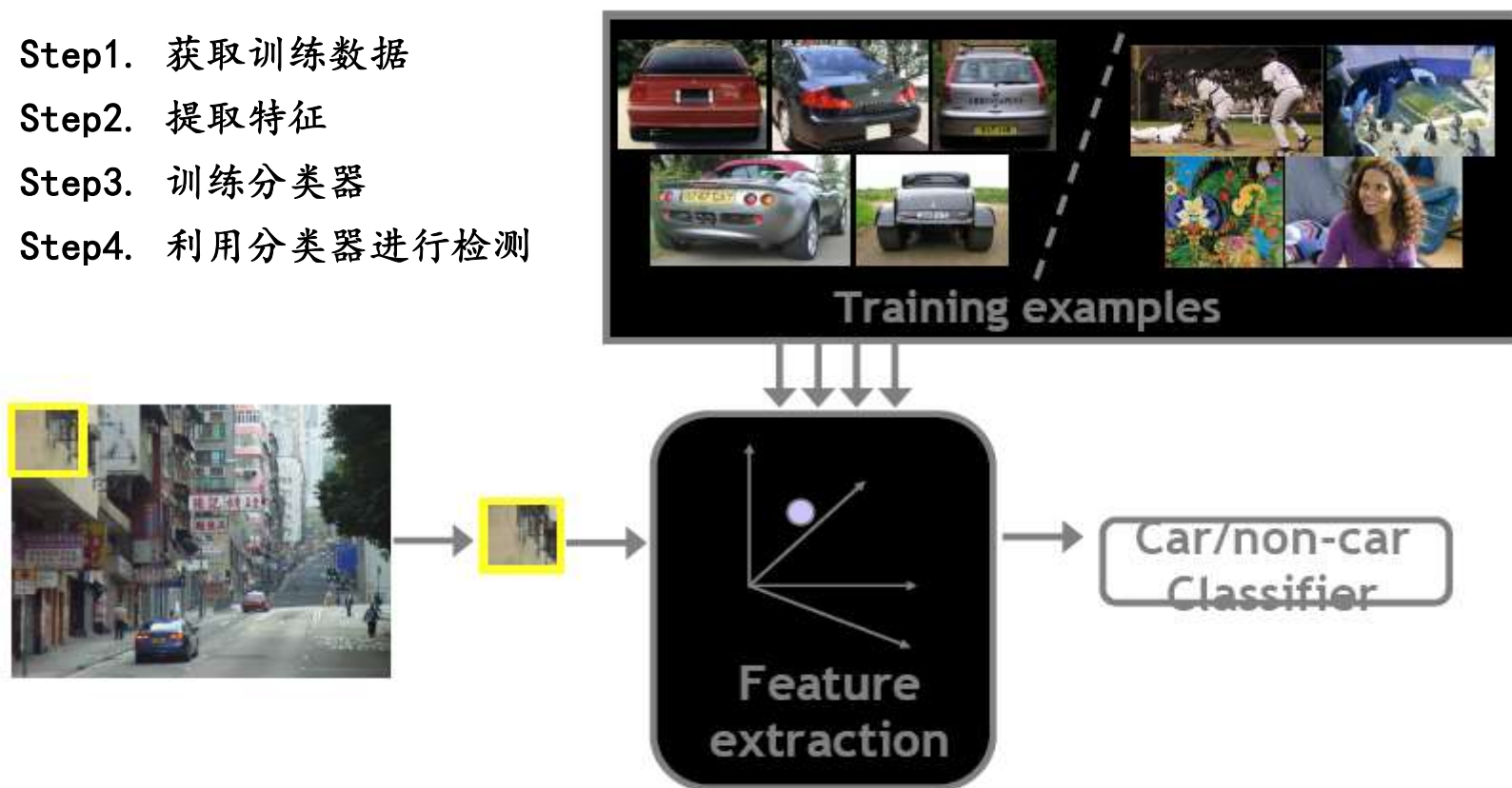


Car/non-car  
Classifier



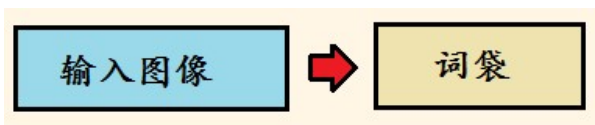
## 物体检测问题

- ◆ Step1. 获取训练数据
- ◆ Step2. 提取特征
- ◆ Step3. 训练分类器
- ◆ Step4. 利用分类器进行检测



## 实例1—利用词袋方法物体识别实例

- ◆ 为了表示一幅图像，我们可以将图像看作文档，即若干个“视觉词汇”的集合，同样的，视觉词汇相互之间没有顺序。



- ◆ 要从图像中提取出相互独立的视觉词汇，需要三个步骤：

- 特征检测
- 特征表示
- 单词本的生成



- ◆ 这里用**SIFT**算法提取不变特征点，作为视觉词汇，并构造单词表，用单词表中单词表示一幅图像



## 步骤1：特征检测

- ◆ 利用**SIFT**算子，从每类图像中提取可视词汇，将所有的词汇集合在一起
- ◆ **SIFT**特征作为词汇，形式：



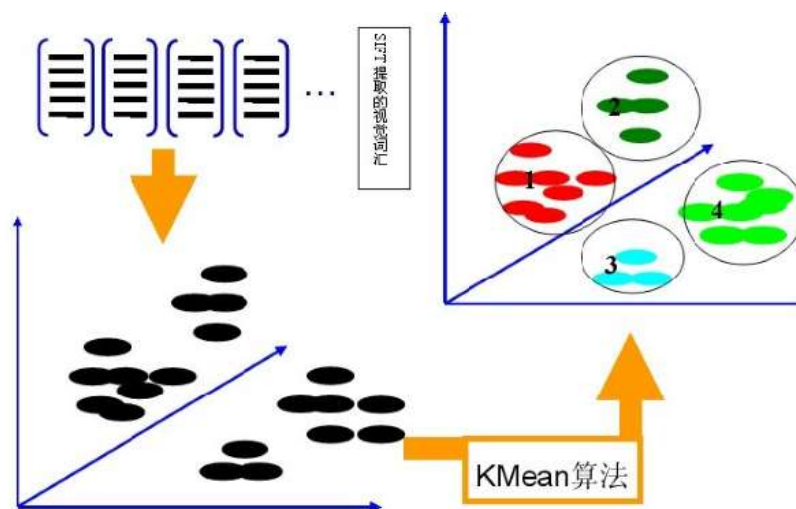
128维向量

- ◆ 从图像中提取视觉词汇，如图：



## 步骤2：构造单词表

- ◆ 利用K均值聚类（ K-Means ）算法构造单词表
- ◆ 对视觉词汇向量根据距离的远近，可以利用K-Means算法将词义相近的词汇合并，  
作为单词表中的基础词汇

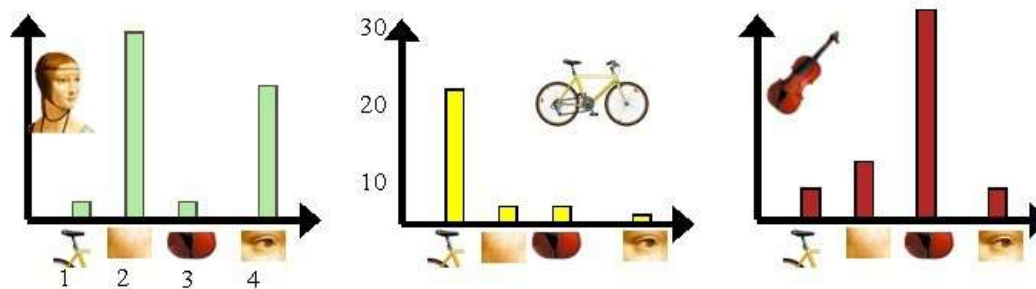




## 步骤3：利用单词表的中词汇表示图像

- (1) 首先，利用SIFT算法，从每幅图像中提取多个特征点
- (2) 然后，提取的特征点用单词表中的单词近似代替
- (3) 最后，通过统计单词表中每个单词在图像中出现的次数，将图像表示成为一个K=4维数值向量。

◆ 如图：



## (2) 实例2—人脸识别 (Face detection)

### ◆ Viola-Jones人脸检测算法 (基于AdaBoost)



**Viola & Jones, Robust Real-Time Face Detection. IJCV, 2004**



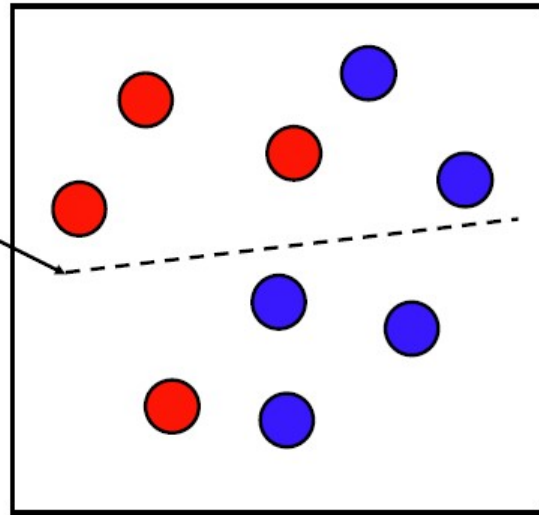
## (a) Adaboost算法原理

- ▶ **Adaboost**是一种迭代算法，核心思想是：针对同一个训练集训练不同的分类器(弱分类器)，然后把这些弱分类器集合起来，构成一个更强的最终分类器(强分类器)。
- ▶ **Adaboost**通过改变数据分布来实现，根据每次训练集之中每个样本的分类是否正确，以及上次的总体分类的准确率，来确定每个样本的权值。
- ▶ 将每次训练得到的分类器最后融合起来，作为最后的决策分类器。
- ▶ **adaboost**分类器可以将关键步骤放在关键的训练数据上面。

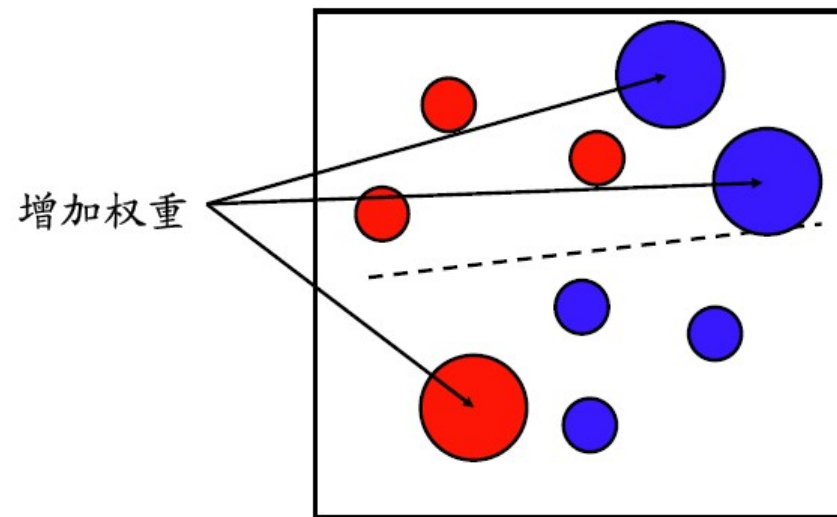


## Boosting实例

弱分类器 1

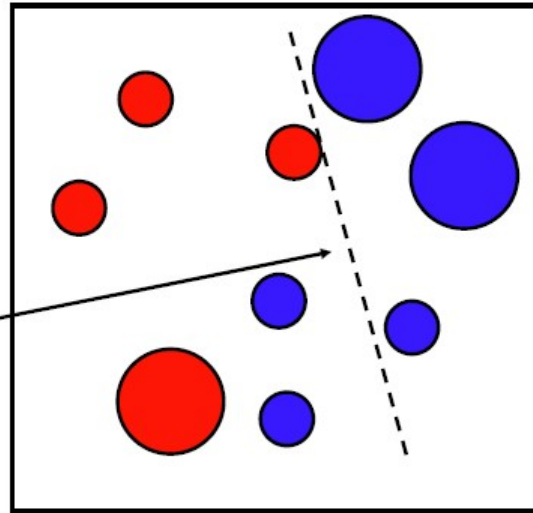


## Boosting实例

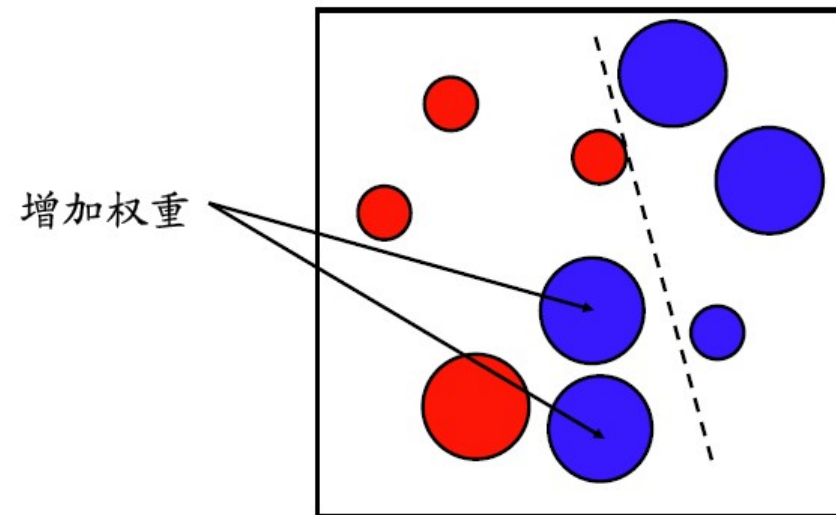


## Boosting实例

弱分类器2

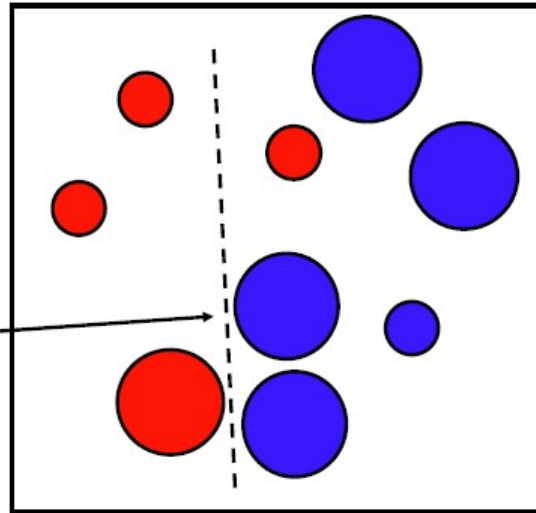


## Boosting实例



## Boosting实例

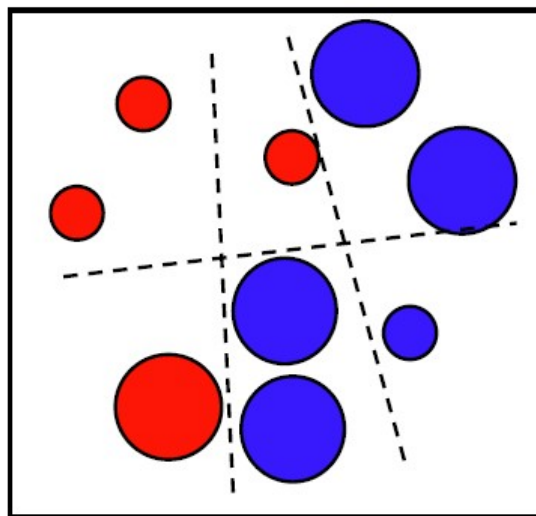
弱分类器3





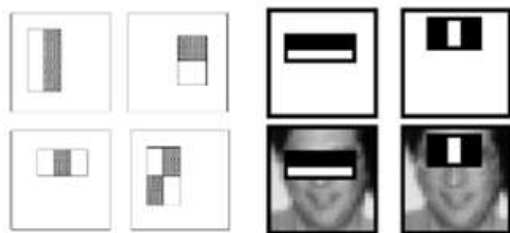
## Boosting实例

最终分类器 = 弱  
分类器线性组合



## (b) 人脸识别

### ◆ 矩形滤波器

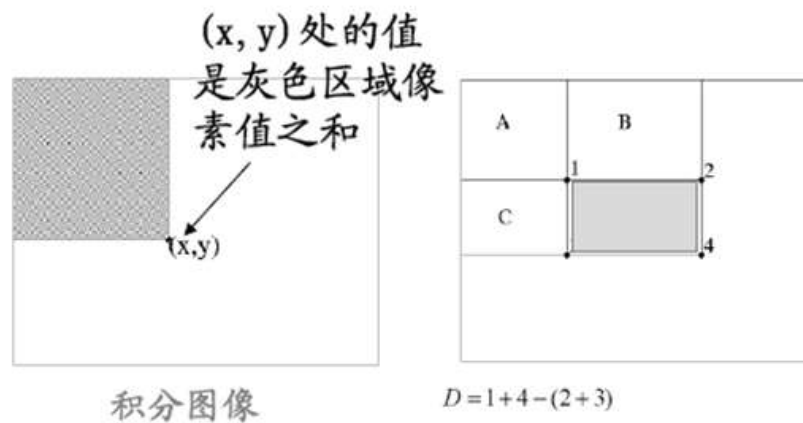


- 计算矩形区域之间的差异
- 反映了邻近空间区域之间的对比
- 近似于Haar小波，但计算效率高

Value =

$$\sum (\text{pixels in white area}) - \sum (\text{pixels in black area})$$

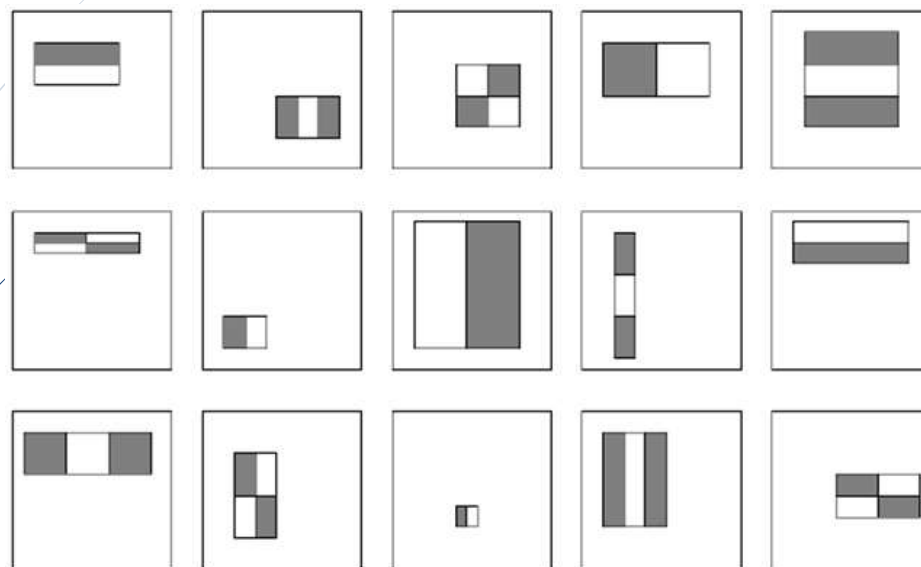
1. 用积分图像实现高效计算
2. 避免了形成不同尺度上的图像



$$\begin{aligned} D &= 1 + 4 - (2 + 3) \\ &= A + (A + B + C + D) - (A + C + A + B) \\ &= D \end{aligned}$$



## 滤波器设计



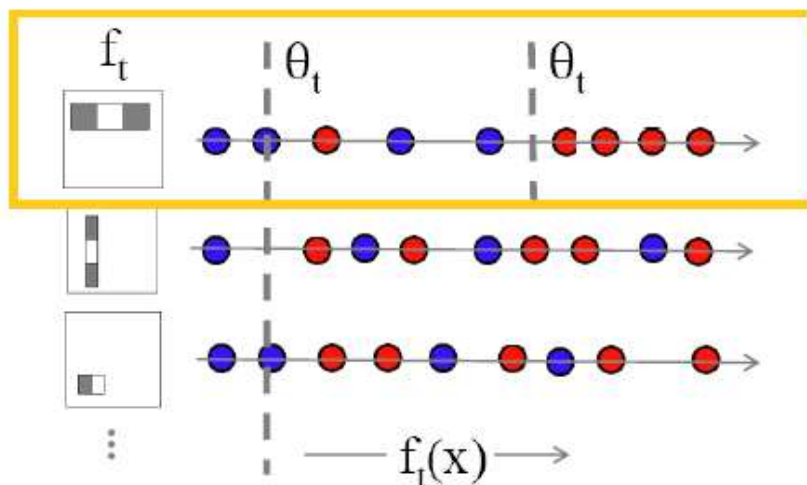
- 考虑所有可能的滤波器  
参数：位置，尺度，类型：  
在24 x 24窗口上形成  
160,000个特征

使用AdaBoost来选择最有效的特征并形成分类器



# Adaboost

- ◆ 学习目标：选择能够最有效地区分人脸与非人脸的矩形特征及其阈值



1. 矩形特征在人脸与非人脸图像上的输出

2. 形成弱分类器:

$$h_t(x) = \begin{cases} +1 & \text{if } f_t(x) > \theta_t \\ -1 & \text{otherwise} \end{cases}$$

3. 根据误差计算样本权重，重复以上选择.



# Adaboost

- ◆ 组合弱分类器 (weak learners), 得到更为精确的集成分类器 (ensemble classifier)。
- ◆ 弱分类器: 性能仅比随机分类稍好
- ◆ 根据矩形特征定义弱分类器:

$$h_t(x) = \begin{cases} 1 & \text{if } p_t f_t(x) > p_t \theta_t \\ 0 & \text{otherwise} \end{cases}$$

Diagram annotations:

- 矩形特征值 (points to  $p_t f_t(x)$ )
- 窗口 (points to  $h_t(x)$ )
- 符号 (points to  $p_t$ )
- 阈值 (points to  $\theta_t$ )

Y. Freund and R. Schapire, A short introduction to boosting, 1999.



## Adaboost算法步骤

- ◆ 初始给每个训练样本以同等权重
- ◆ 循环执行以下步骤：
  - 根据当前加权训练集，选择最佳弱分类器
  - 提升被当前弱分类器错分的训练样本的权重
- ◆ 按照各弱分类器分类精度对其加权，然后将各个弱分类器形成线性组合，得到最终分类器。



## Viola-Jones算法中的AdaBoost

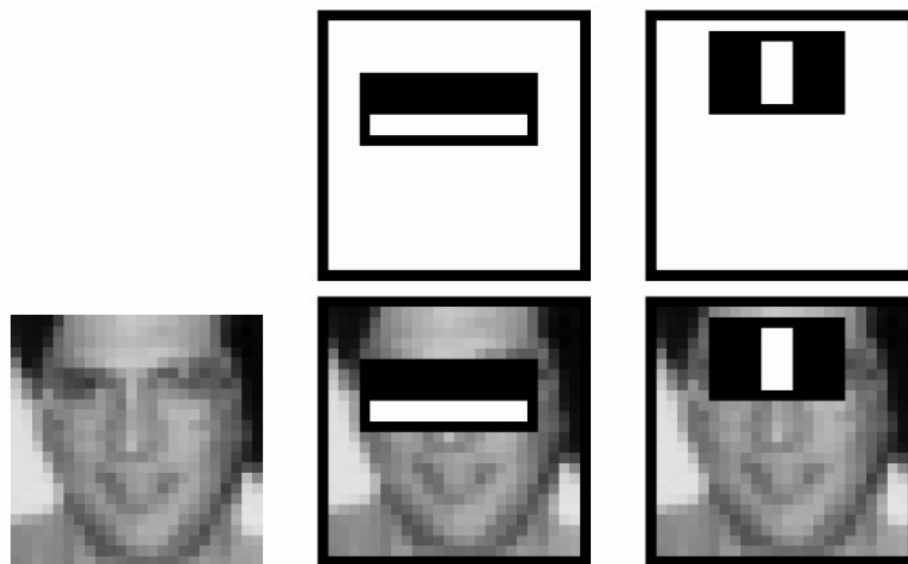
- ◆ 每一次boosting迭代如下:

- 评价每一个样本上的每一种矩形特征
- 为每一种矩形特征选择最佳分类阈值
- 选择最优的矩形特征及其阈值组合
- 改变样本权重

- ◆ 计算复杂度:  $O(MNT)$

- $M$ : 特征数,  $N$ : 样本数,  $T$ : 阈值数



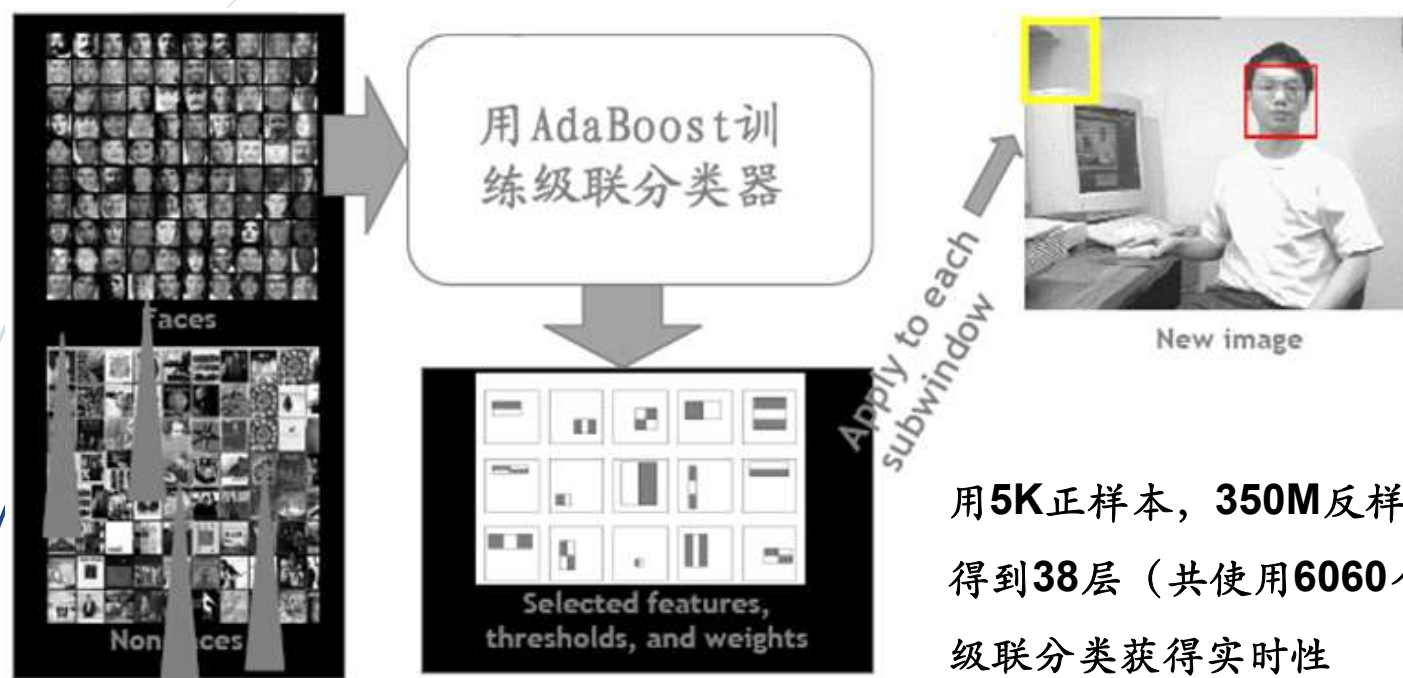


用AdaBoost选出的头两个矩形特征





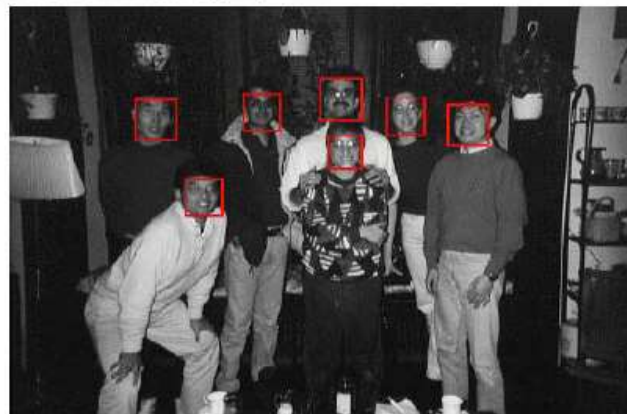
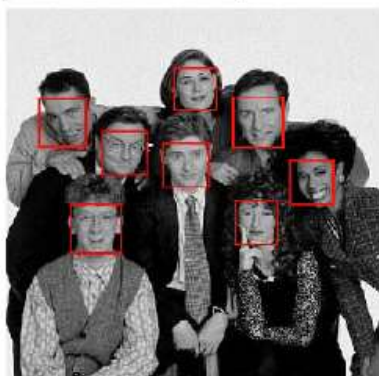
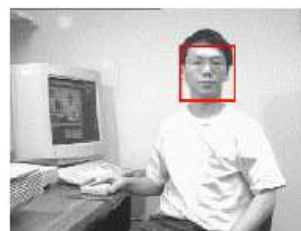
## Viola-Jones检测算法-总体流程



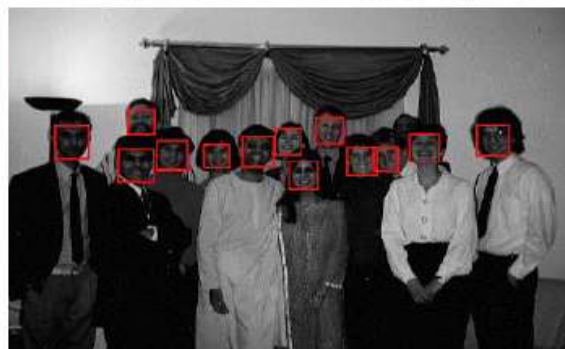
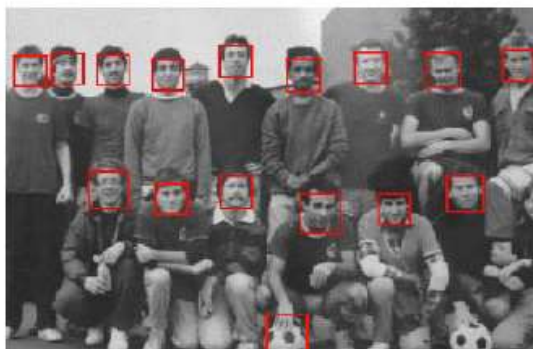
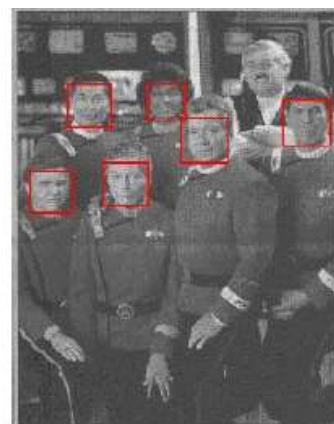
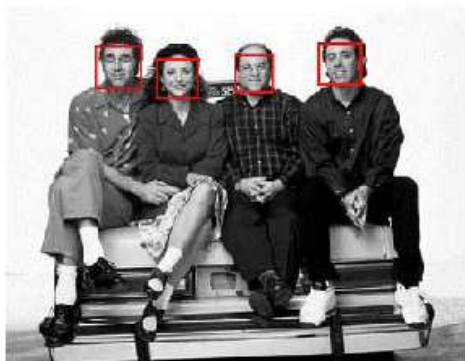
用**5K**正样本，**350M**反样本学习  
得到**38**层（共使用**6060**个特征）  
级联分类获得实时性



## Viola-Jones人脸检测结果



## Viola-Jones人脸检测结果



## Viola-Jones人脸检测结果



### (3) 实例3—人脸识别主要原理

- 人脸认证 (Authentication/Verification) (二分类)



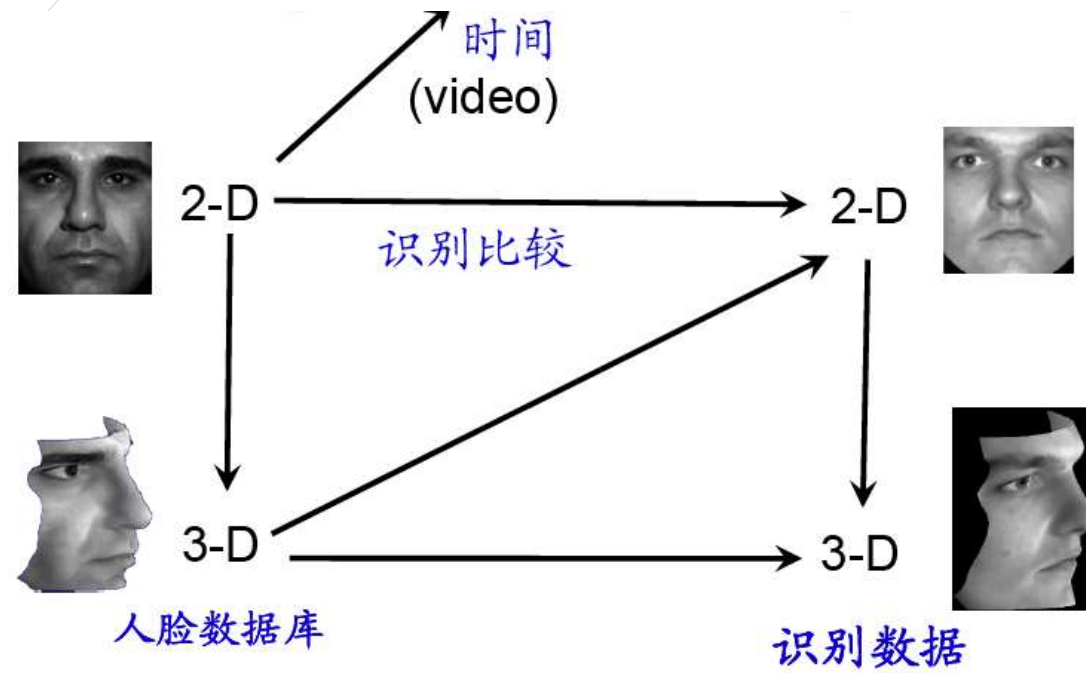
- 人脸识别 (Identification/Recognition) (多分类)



Zhao et al., Face Recognition: a literature survey. ACM Computing survey, 2003

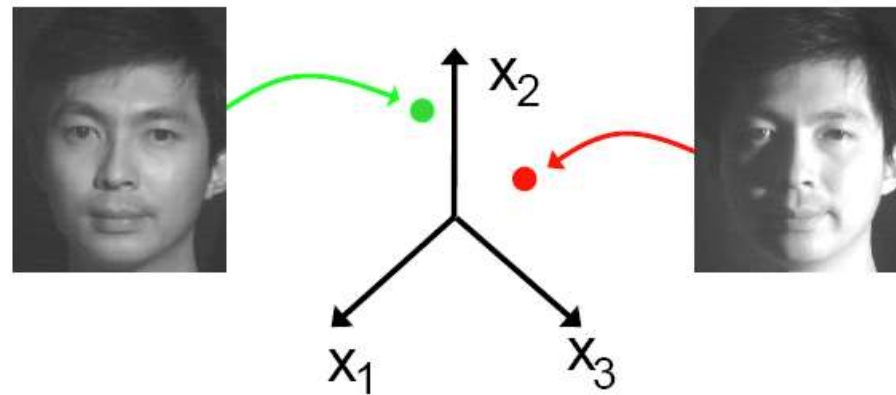


## 人脸识别：2-D and 3-D





## 图像看作是像素的集合

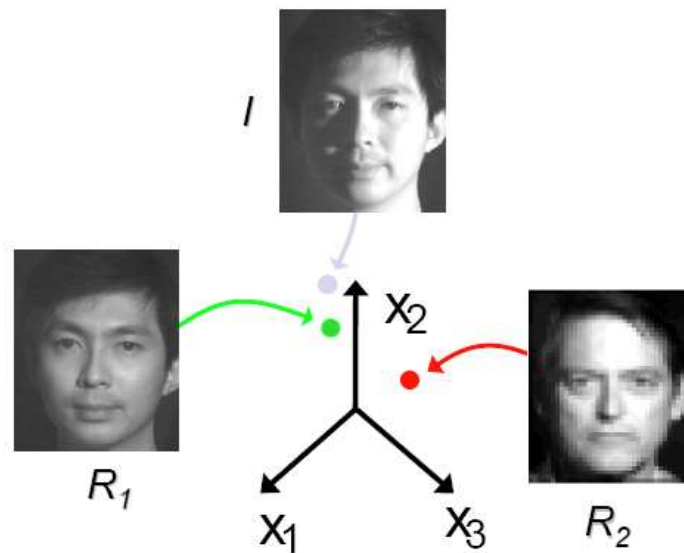


将由 $n$ 个像素构成的图像视为 $n$ 维空间中的点



## 最近邻分类器

- ◆ 在特征图像上执行分类



$$ID = \arg \min_j dist(R_j, I)$$

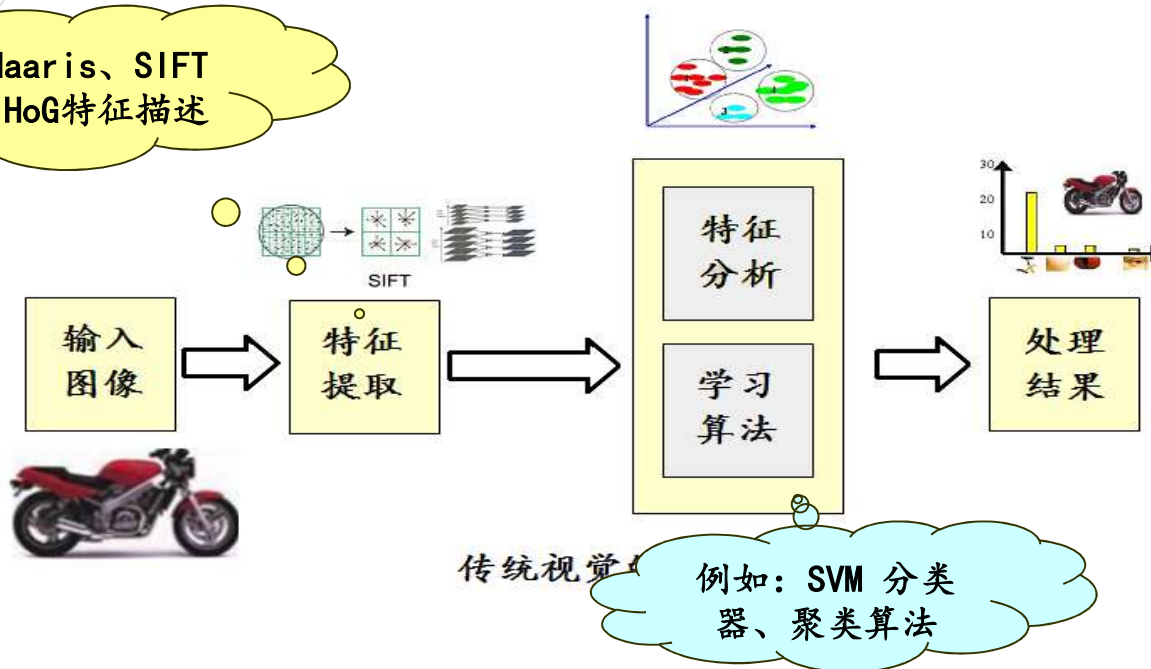




## 8. 传统识别算法的特点

- ◆ 传统视觉与智能视觉有着不同的理解框架

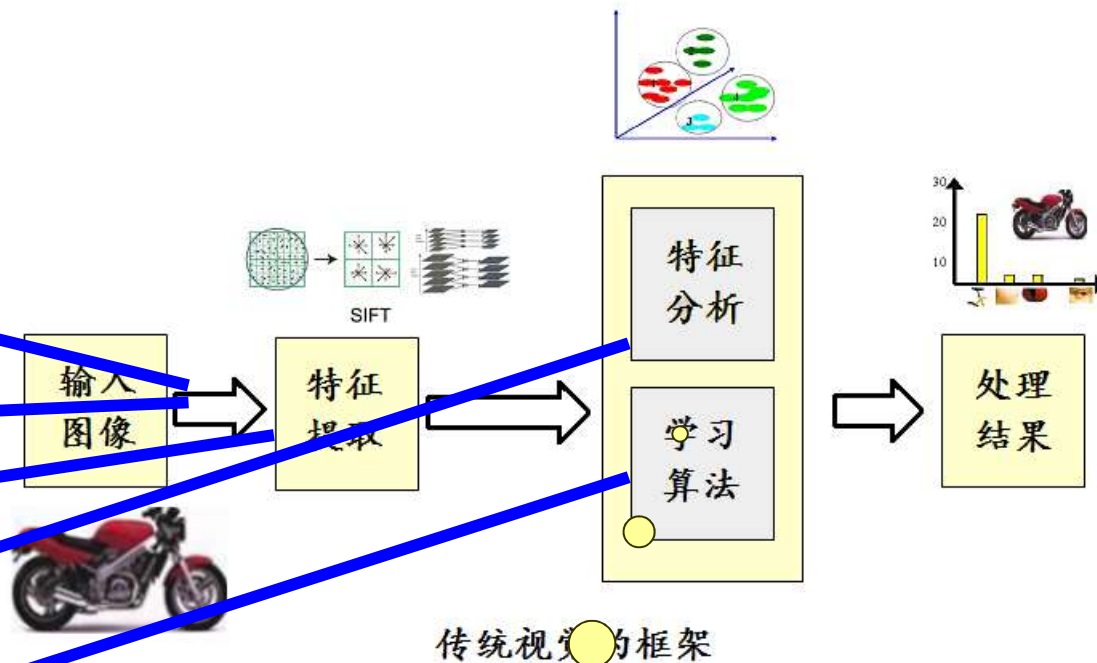
例如：Haaris、SIFT  
算子、HoG特征描述



## 传统视觉处理主要步骤

### ◆ 主要步骤

- 图像预处理
- 特征感知
- 特征提取
- 特征筛选
- 推理预测与识别。

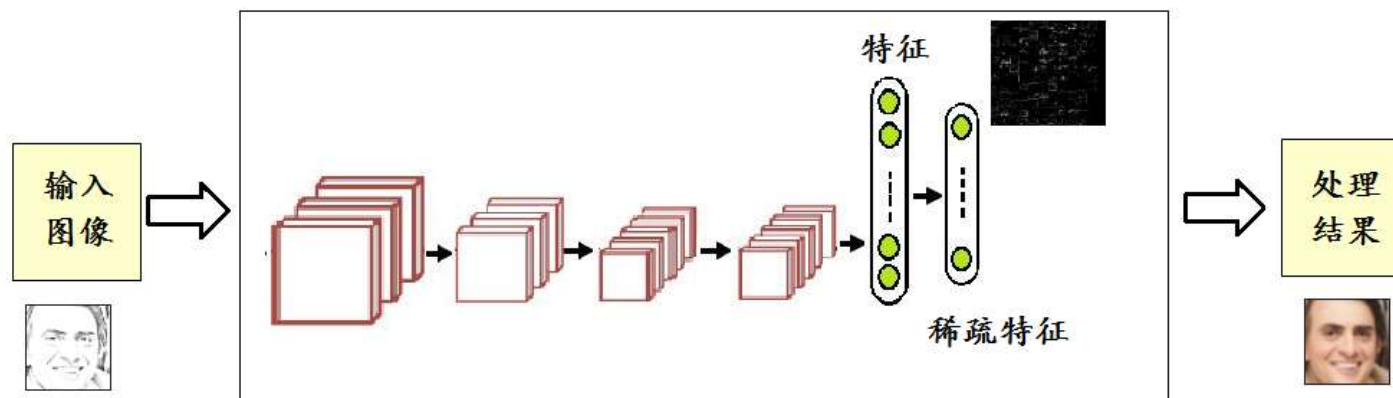


需要选取合适的特征和学习算法，才能得到最优的处理效果



## 智能视觉的深度学习框架

- ◆ 深度学习模型的多隐藏层多感知器
- ◆ 深度学习模型能完成信息的降维，剔除冗余信息，又能抽取到丰富的特征向量，优化特征向量



## 传统视觉与智能视觉区别

- ◆ **传统视觉**：把特征提取和学习过程分开设计，即先提取特征，然后在视觉学习时，结合预先提取的特征进行分析决策

- ◆ 例如：进行人脸识别，其主要步骤：

输入人脸图像—特征提取与表达—学习处理—识别结果

## 传统视觉与智能视觉区别

- ◆ **智能视觉**：不需要人工提取特征，直接建立输入和输出之间的映射关系，从而得到处理结果（例如，数据标签结果、合成的人脸结果、人脸识别结果）
- ◆ 无需手动提取特征
- ◆ 在多隐层中自动生成特征
- ◆ 深度学习符合人类脑思考机理