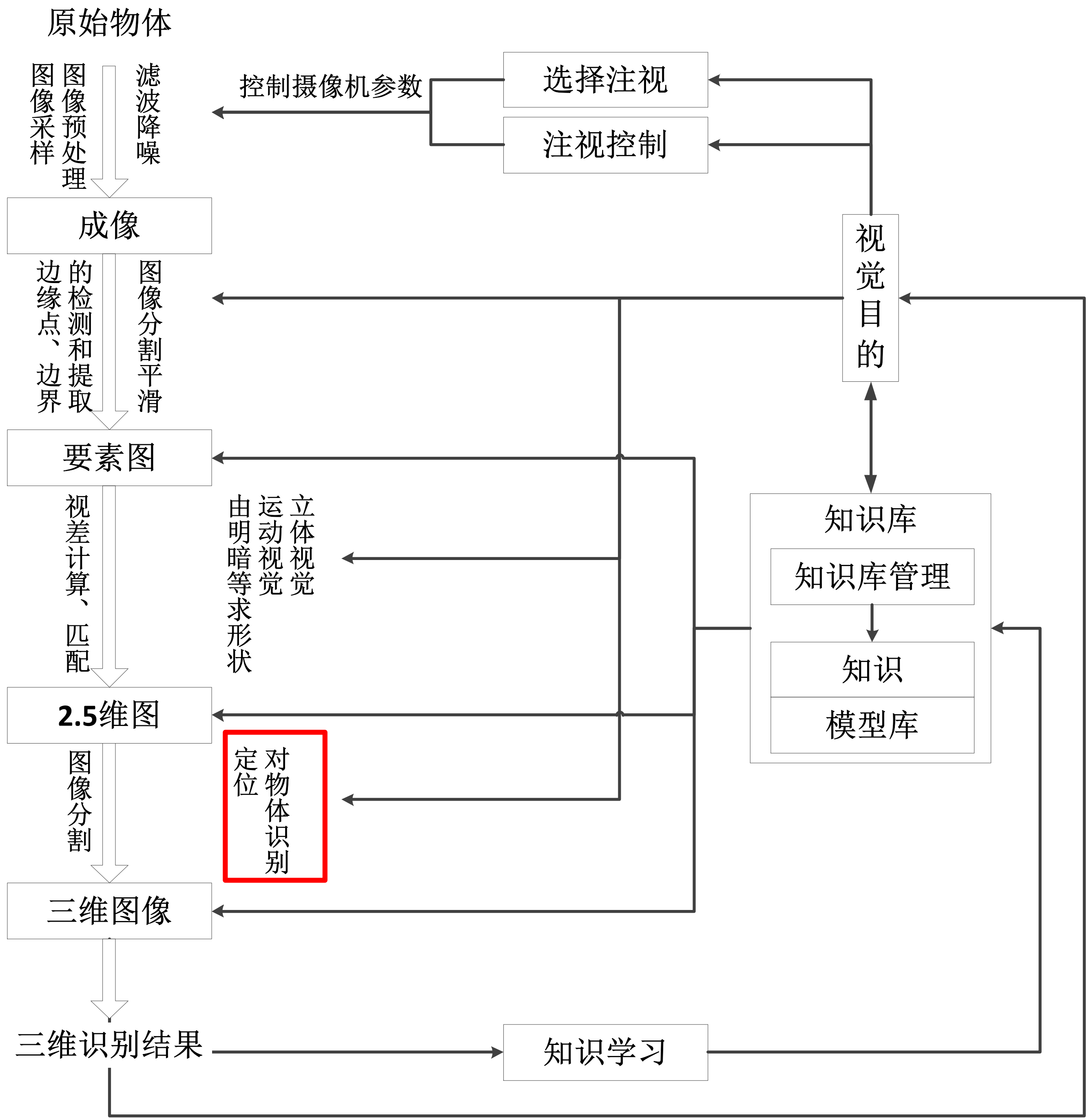
计算机视觉
——目标识别



原始物体
图像采样
图像预处理
滤波降噪
控制摄像机参数
选择注视
注视控制
成像
边缘点、边界的检测和提取
图像分割平滑
视觉目的
要素图
视差计算、匹配
由明暗等求形状
运动视觉
立体视觉
知识库
知识库管理
知识
模型库
2.5维图
图像分割
对物体识别定位
三维图像
三维识别结果
知识学习

目的

- 了解什么是目标识别？
- 熟悉两种目标识别方法
- 可在编程作业中应用

目标识别

- **目标检测**
  - 人脸检测（Haar+AdaBoost）
  - 行人检测
- **目标分类**
  - 局部特征（bag-of-words，空间金字塔）
- **目标身份确认**
  - 人脸 身份确认
  - 行人身份确认

目标识别

- 什么是目标识别？
  - What is it?

    Object classification

  - Where is it?

    Object detection

  - Who is it?
    - Object Identification

目标识别包括什么？

确认: 这是一辆公共汽车吗？

检测:图像中有小汽车吗?

身份鉴定:是毛主席的照片吗?

目标分类

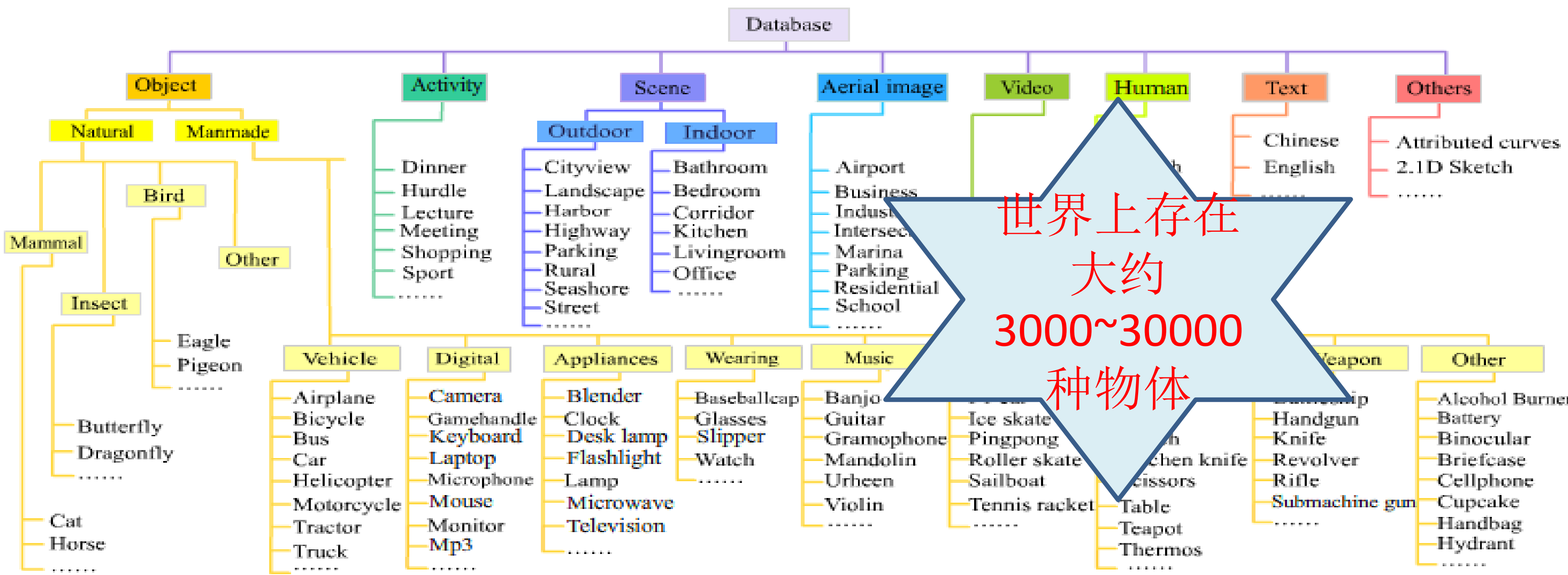
场景及上下文分类

- 室外
- 市区
- 交通
- …

目标识别

- 目标识别面临的挑战
  - 挑战 1: 目标种类多
  - 挑战 2: 视点变化
  - 挑战 3: 亮度变化
  - 挑战 4: 遮挡
  - 挑战 5: 尺度
  - 挑战 6: 形变
  - 挑战 7: 复杂背景
  - 挑战 8: 类内变化大

挑战 1: 目标种类多

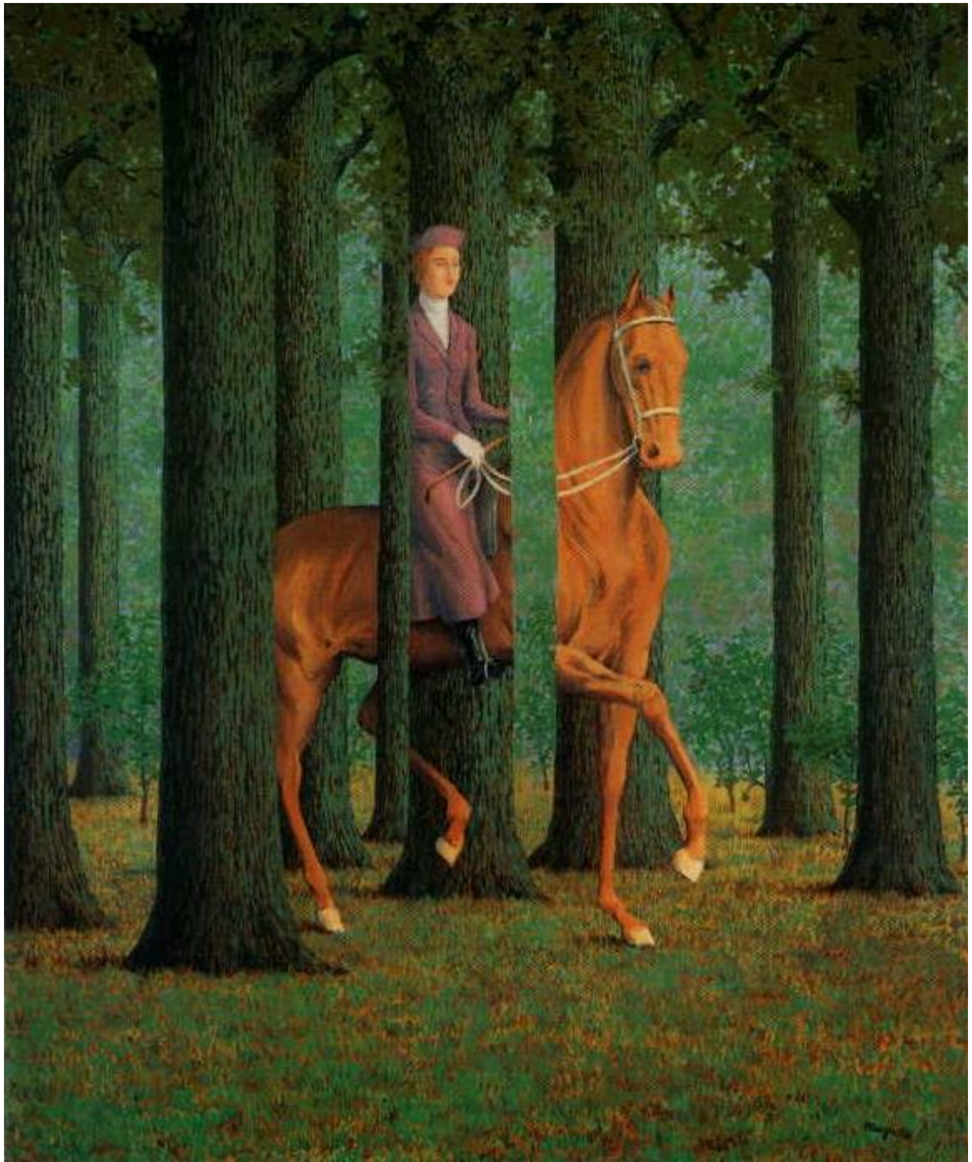

Database
Object
Activity
Scene
Aerial image
Video
Human
Text
Others
Natural
Manmade
Outdoor
Indoor
Chinese
English
Attributed curves
2.1D Sketch
Bird
Mammal
Other
Insect
Eagle
Pigeon
Butterfly
Dragonfly
Cat
Horse
Dinner
Hurdle
Lecture
Meeting
Shopping
Sport
Cityview
Landscape
Harbor
Highway
Parking
Rural
Seashore
Street
Bathroom
Bedroom
Corridor
Kitchen
Livingroom
Office
Airport
Marina
Parking
Residential
School
Vehicle
Digital
Appliances
Wearing
Music
Other
Airplane
Bicycle
Bus
Car
Helicopter
Motorcycle
Tractor
Truck
Camera
Gamehandle
Keyboard
Laptop
Microphone
Mouse
Monitor
Mp3
Blender
Clock
Desk lamp
Flashlight
Lamp
Microwave
Television
Baseballcap
Glasses
Slipper
Watch
Banjo
Guitar
Gramophone
Mandolin
Urheen
Violin
Ice skate
Pingpong
Roller skate
Sailboat
Tennis racket
Table
Teapot
Thermos
Handgun
Knife
Revolver
Rifle
Submachine gun
Alcohol Burner
Battery
Binocular
Briefcase
Cellphone
Cupcake
Handbag
Hydrant
世界上存在
大约
3000~30000
种物体

挑战 2: 视点变化

Michelangelo 1475-1564

挑战 3: 亮度变化

挑战 4: 遮挡

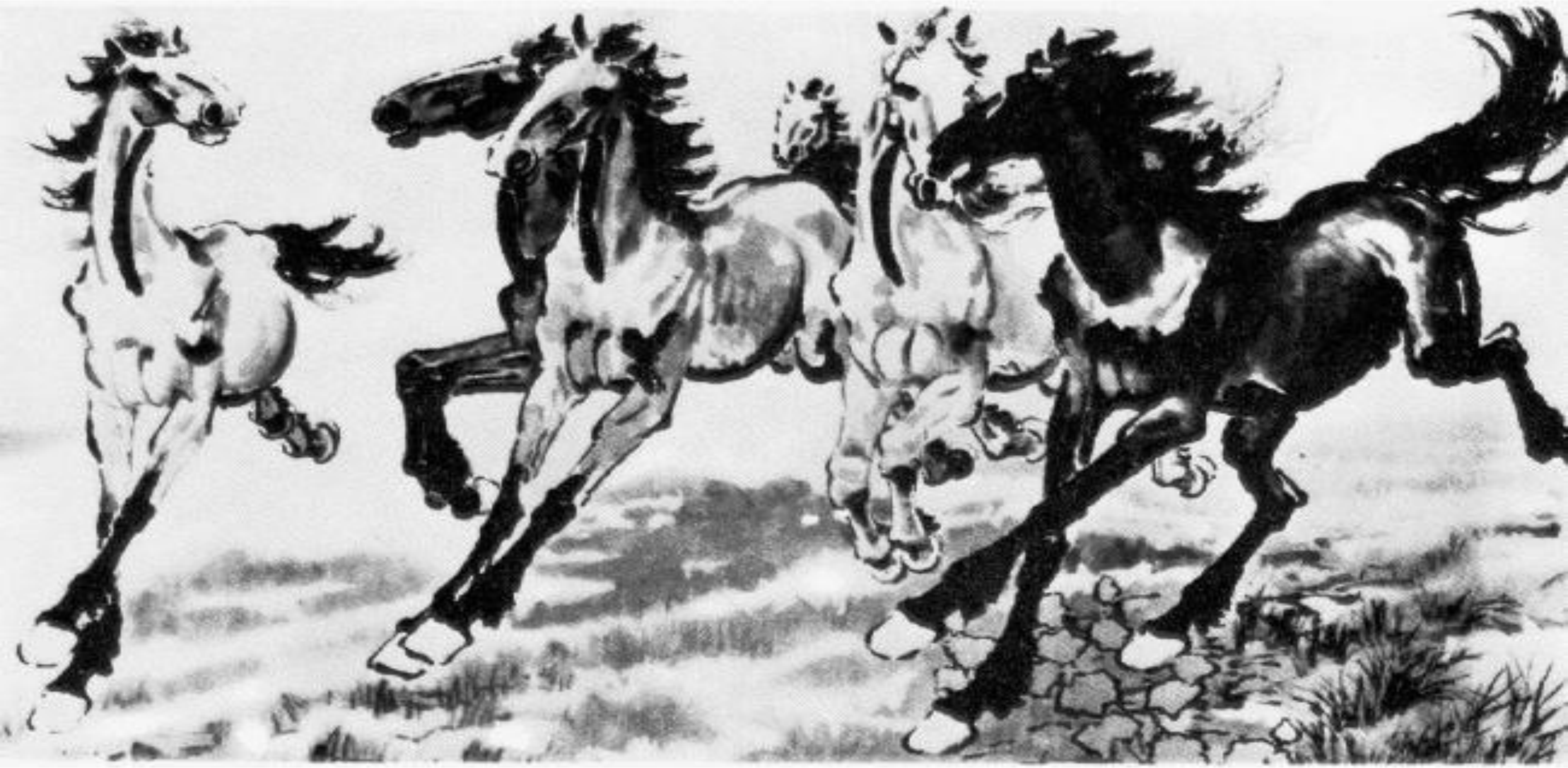


Magritte, 1957

挑战 4: 遮挡

挑战 5: 尺度

挑战6: 形变



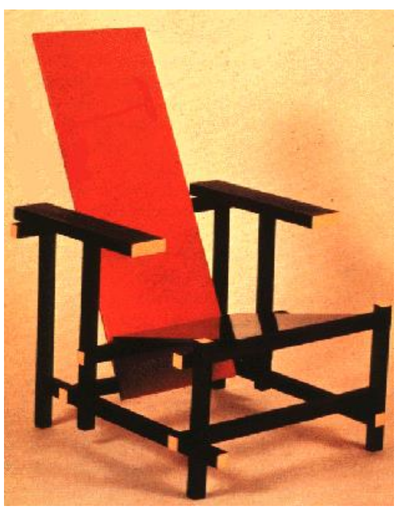
Xu, Beihong 1943

挑战 7: 复杂背景

Klimt, 1913

挑战 8: 类内变化大



目标识别方法

- 特征提取
  - Haar
  - SIFT
  - HOG
  - 直方图
  - 形状特征
  - LBP

\+

- 分类器学习
  - SVM
  - 线性判别器
  - 决策树

局部 or 全局？

PCA是不是一个分类器？

最近邻是不是一个分类器？

目标识别框架

- 利用一个预测函数预测类别，输入为图像的特征
-

f() = “apple”

f() = “tomato”

f() = “cow”

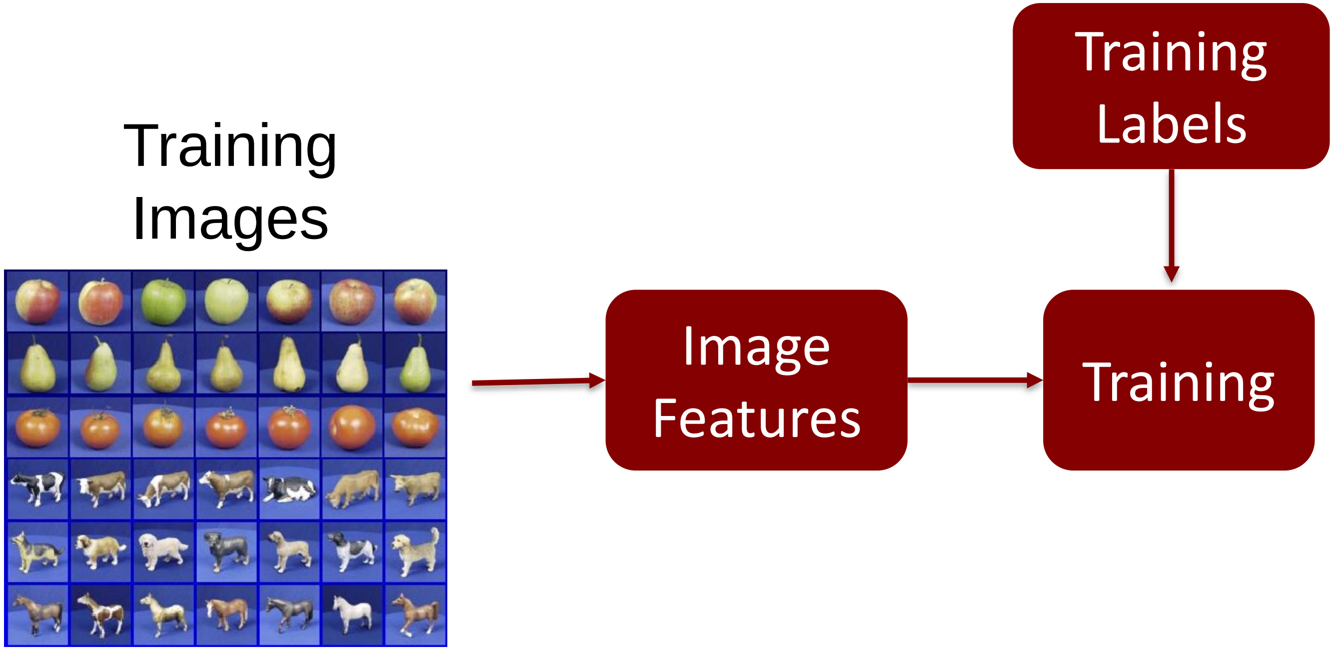
Dataset: ETH-80, by B. Leibe　　Slide credit: L. Lazebnik

流程示例——训练

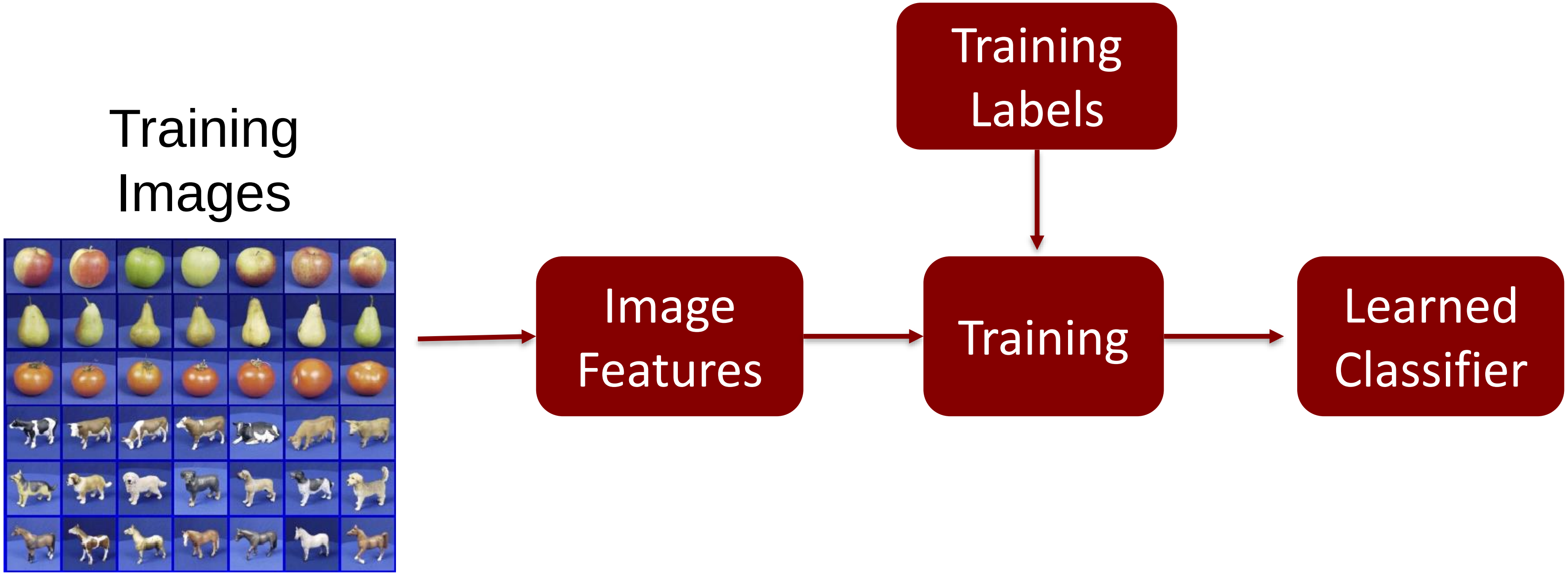
Training Images

Image Features

流程示例——训练



流程示例——训练


Training Images
Image Features
Training Labels
Training
Learned Classifier

流程示例——训练

Training Images

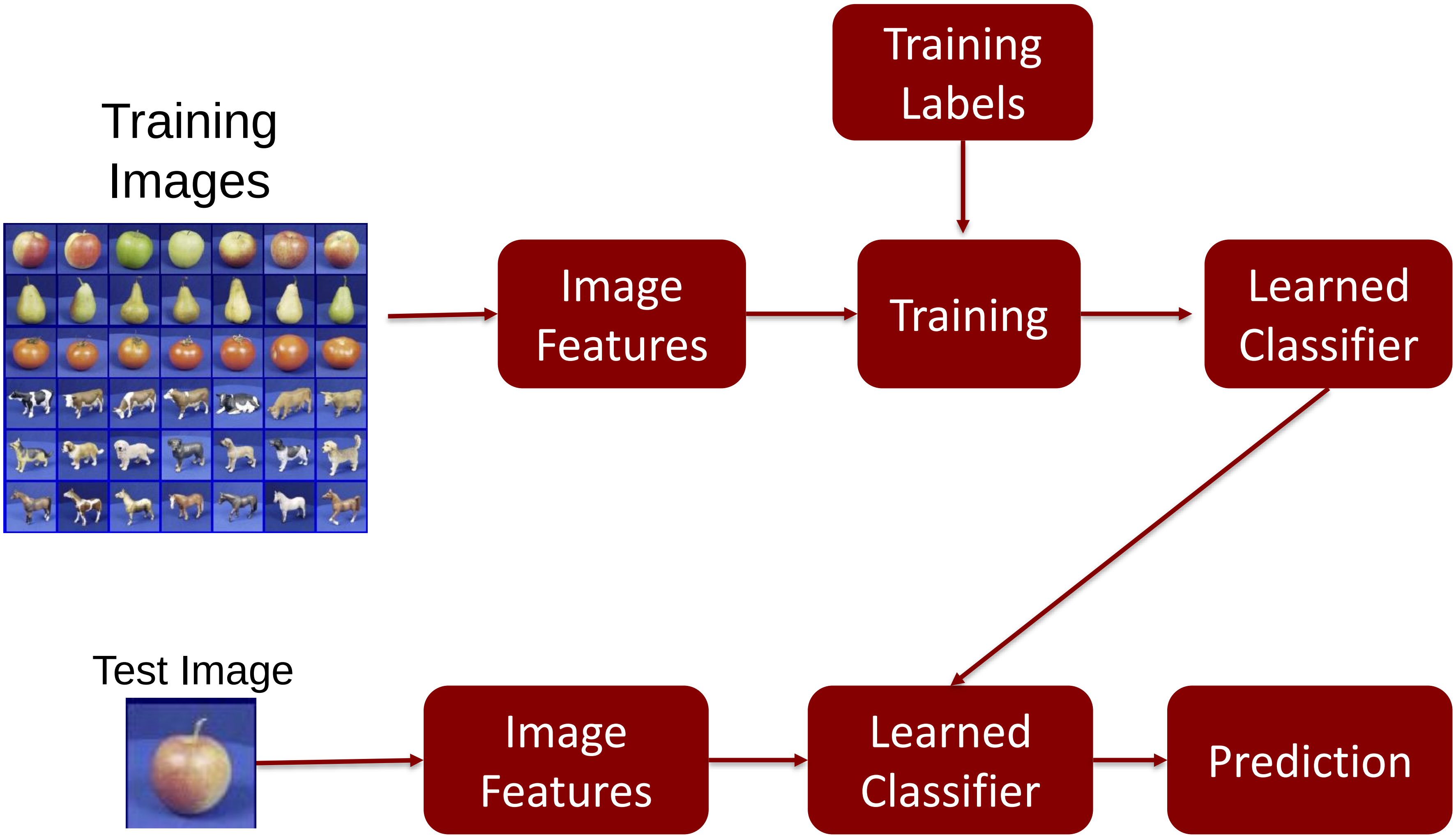
Image Features → Training → Learned Classifier

Training Labels → Training

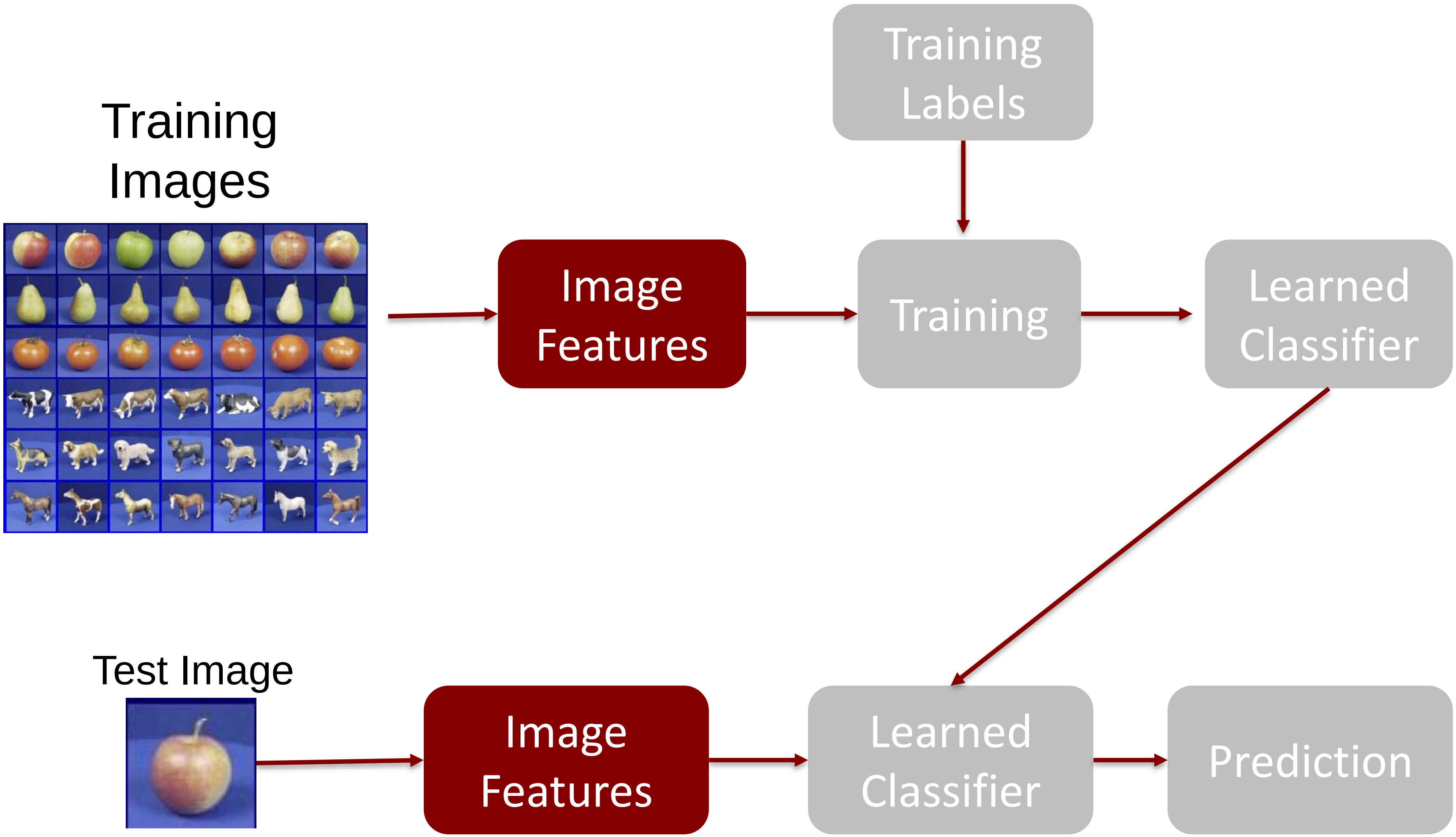
Test Image

Image Features

流程示例——训练


Training Images
Training Labels
Image Features
Training
Learned Classifier
Test Image
Image Features
Learned Classifier
Prediction

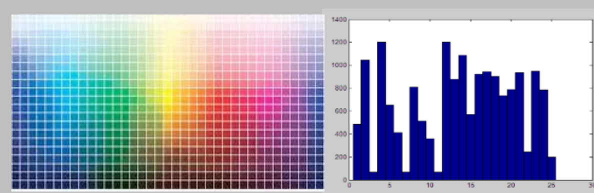
流程示例——训练


Training Images
Image Features
Training Labels
Training
Learned Classifier
Test Image
Image Features
Learned Classifier
Prediction

图像特征

Input image

Color: Quantize RGB values



Invariance?

? Translation

? Scale

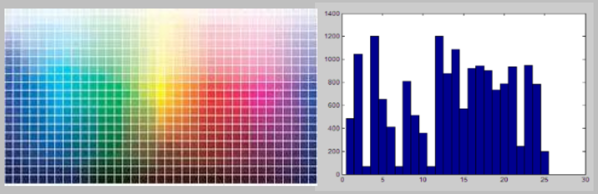
? Rotation

? Occlusion

图像特征

Input image

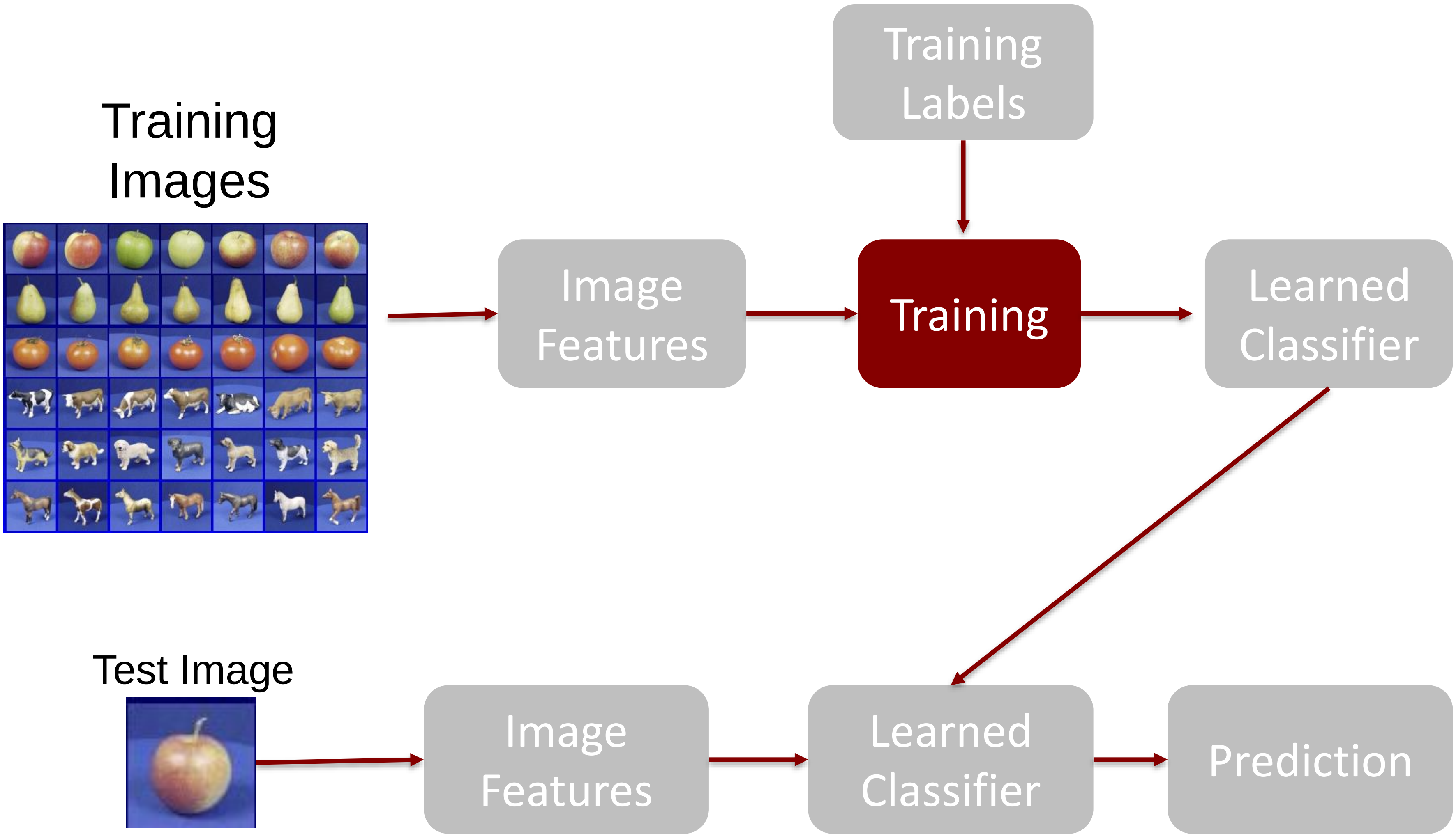
Color: Quantize RGB values



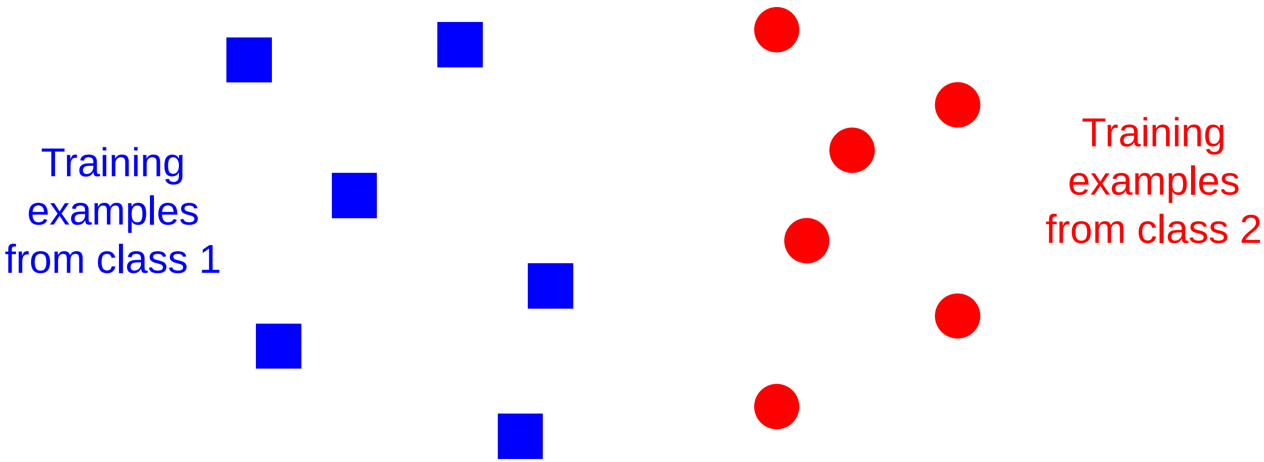
Invariance?

- Translation
- ? Scale
- Rotation (in-planar)
- Occlusion

流程示例——训练

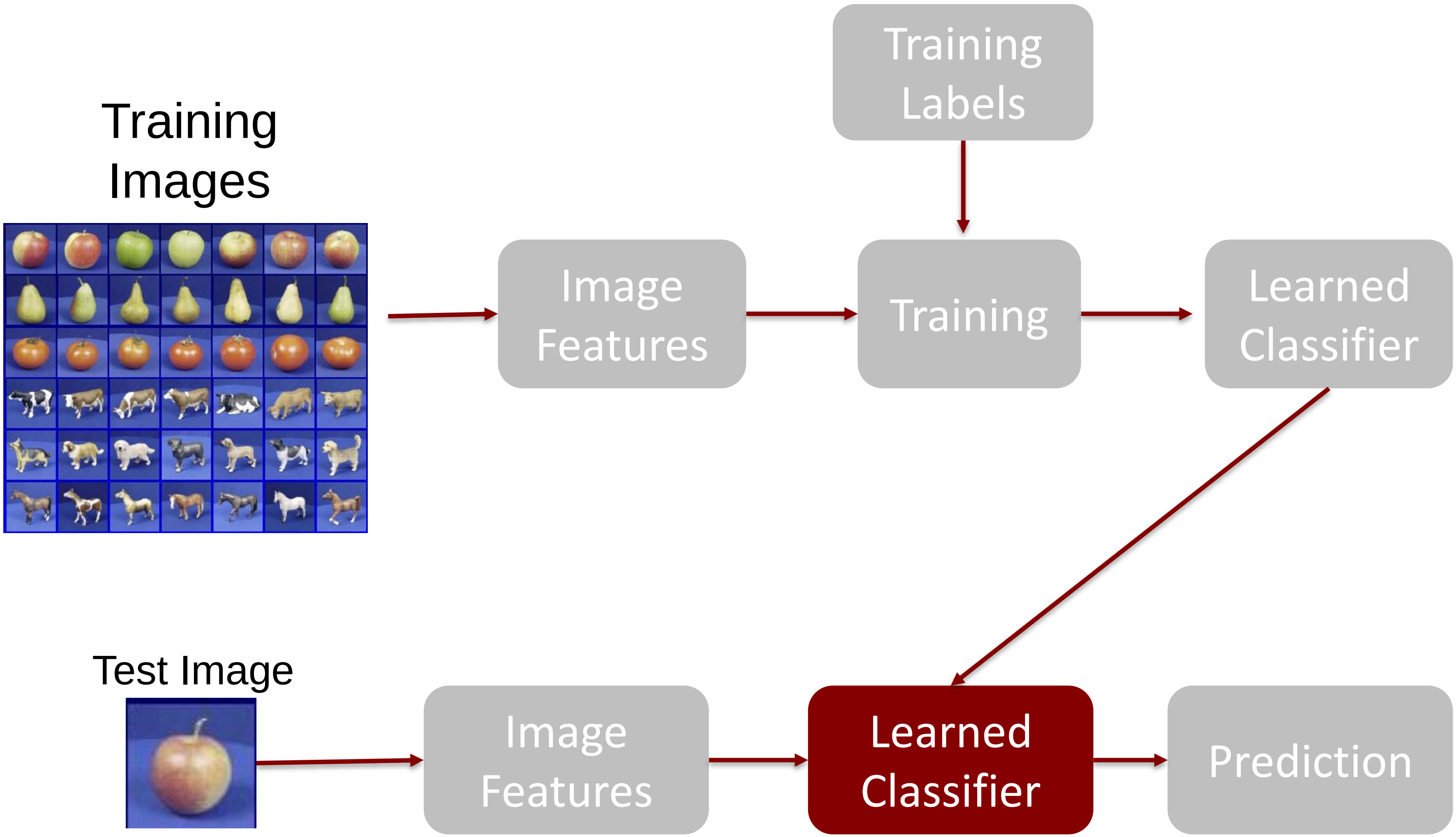

Training Images
Training Labels
Image Features
Training
Learned Classifier
Test Image
Image Features
Learned Classifier
Prediction

分类：最近邻分类器

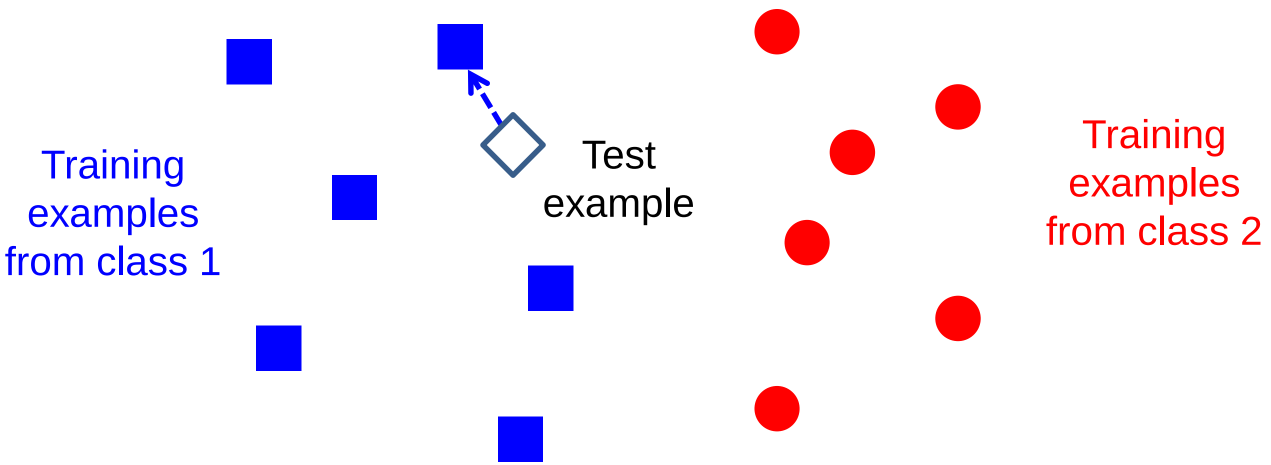


Slide credit: L. Lazebnik

流程示例——训练


Training Labels
Training Images
Image Features
Training
Learned Classifier
Test Image
Image Features
Learned Classifier
Prediction

分类：最近邻分类器



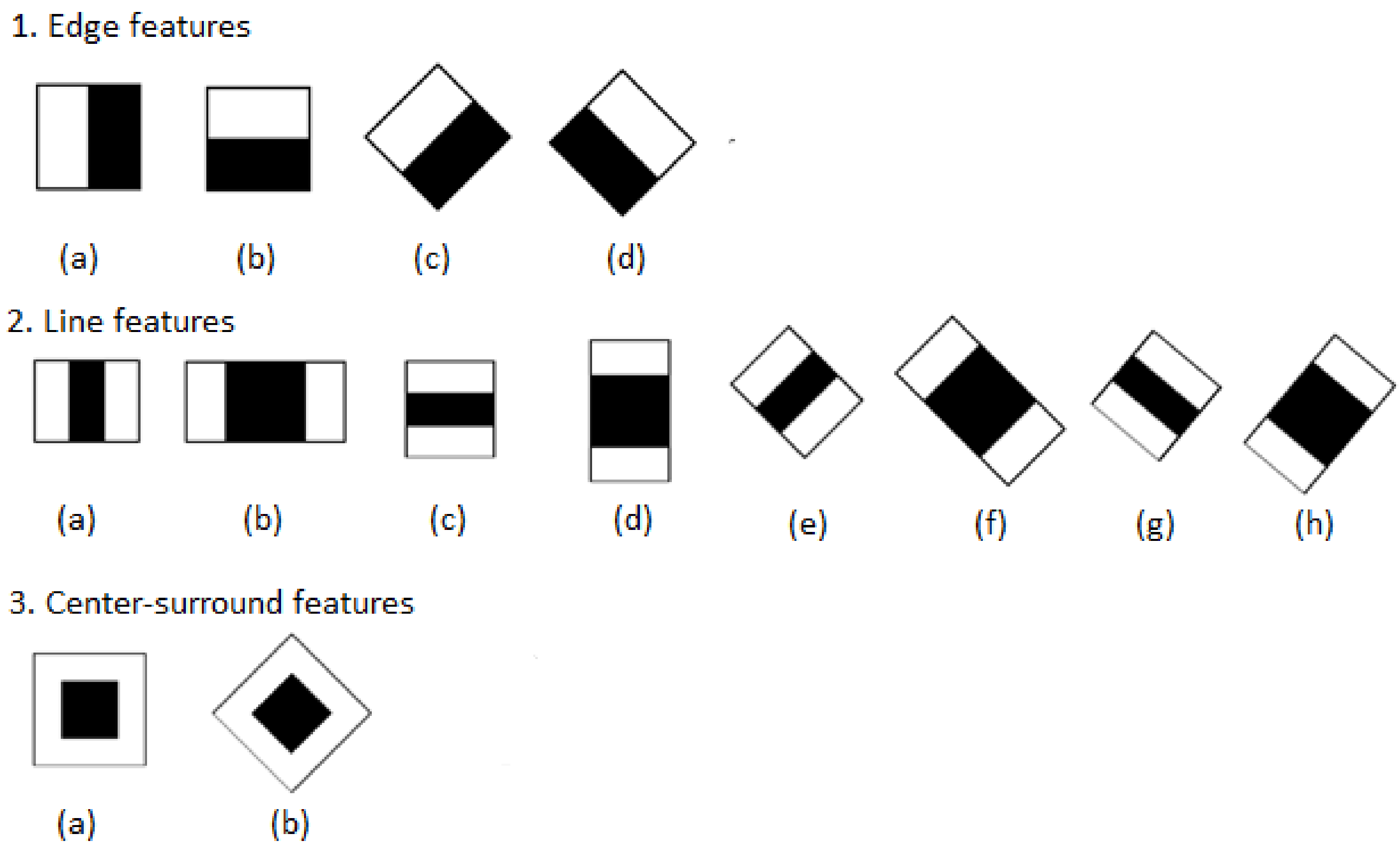
Slide credit: L. Lazebnik

人脸检测、人脸识别

- 人脸检测是目标检测中比较成功的例子

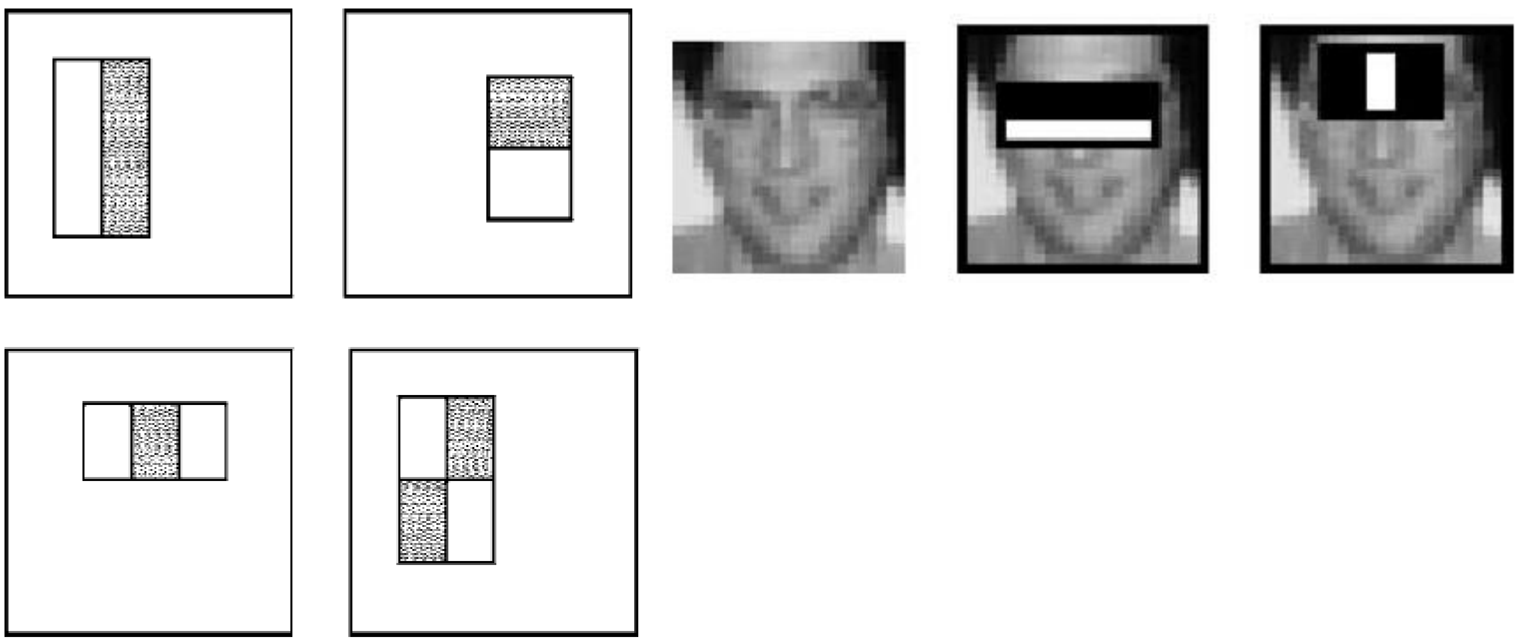
人脸检测

- Haar特征


1. Edge features
(a)
(b)
(c)
(d)
2. Line features
(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
3. Center-surround features
(a)
(b)

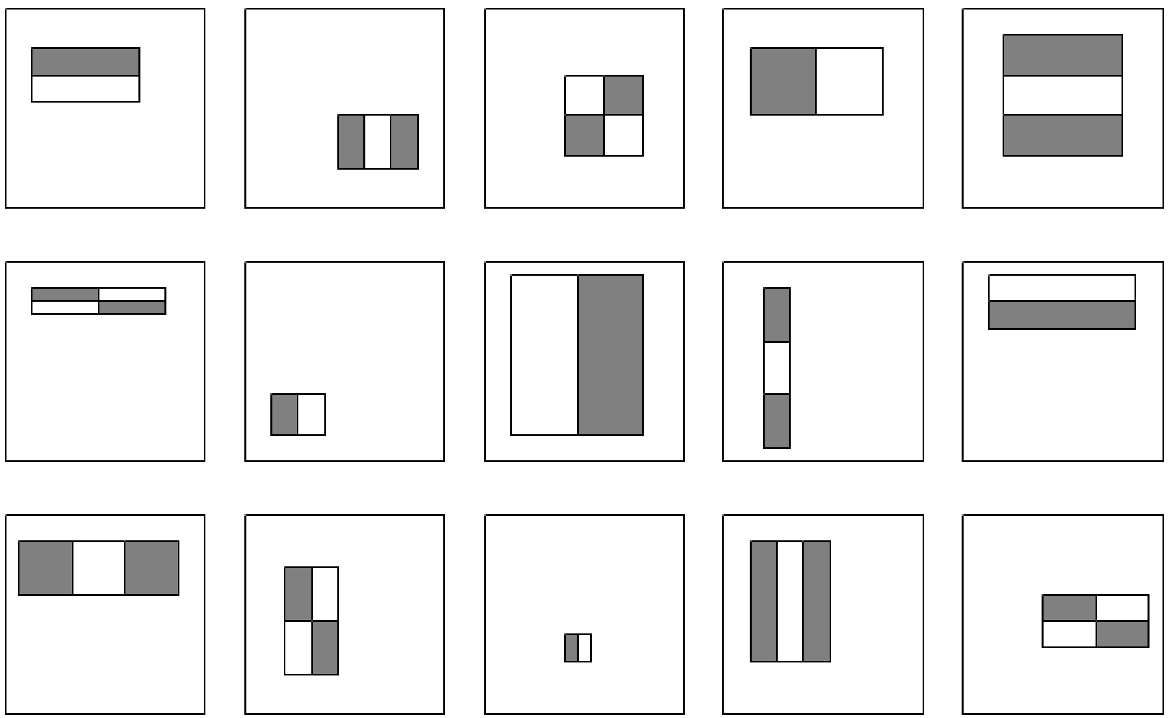
人脸检测

- 特征： Haar-like Features



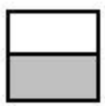
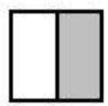
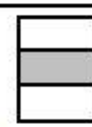
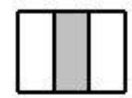
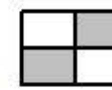
人脸检测

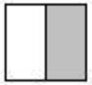
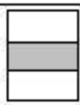
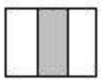
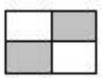
- Haar-like特征原型



人脸检测

- 过完备特征
  - 从24*24像素的图像上可以提取多少个Haar-like features？

特征模板	1、	2、	3、	4、	5、
(s, t)条件	(1, 2)	(2, 1)	(1, 3)	(3, 1)	(2, 2)

特征模板	数量
1、 2、	$2\times\Omega_{(1,2)}^{24}=2\times\left(\left[\frac{24}{1}\right]+\left[\frac{23}{1}\right]+\cdots+\left[\frac{2}{1}\right]+1\right)\times\left(\left[\frac{24}{2}\right]+\left[\frac{23}{2}\right]+\cdots+\left[\frac{3}{2}\right]+1\right)$ $=2\times(24+23+\cdots+2+1)\times(12+11+11+\cdots+2+1+1)$ $=2\times300\times144=2\times43{,}200$ $=86{,}400$
3、 4、	$2\times\Omega_{(1,3)}^{24}=2\times\left(\left[\frac{24}{1}\right]+\left[\frac{23}{1}\right]+\cdots+\left[\frac{2}{1}\right]+1\right)\times\left(\left[\frac{24}{3}\right]+\left[\frac{23}{3}\right]+\cdots+\left[\frac{4}{3}\right]+1\right)$ $=2\times(24+23+\cdots+2+1)\times(8+7+7+\cdots+1+1+1)$ $=2\times300\times92=2\times27{,}600$ $=55{,}200$
5、	$\Omega_{(2,2)}^{24}=\left(\left[\frac{24}{2}\right]+\left[\frac{23}{2}\right]+\cdots+\left[\frac{3}{2}\right]+1\right)\times\left(\left[\frac{24}{2}\right]+\left[\frac{23}{2}\right]+\cdots+\left[\frac{3}{2}\right]+1\right)$ $=(12+11+11+\cdots2+1+1)\times(12+11+11+\cdots+2+1+1)$ $=144\times144$ $=20{,}736$
总数	$\Omega^{24}=162{,}336$

人脸检测

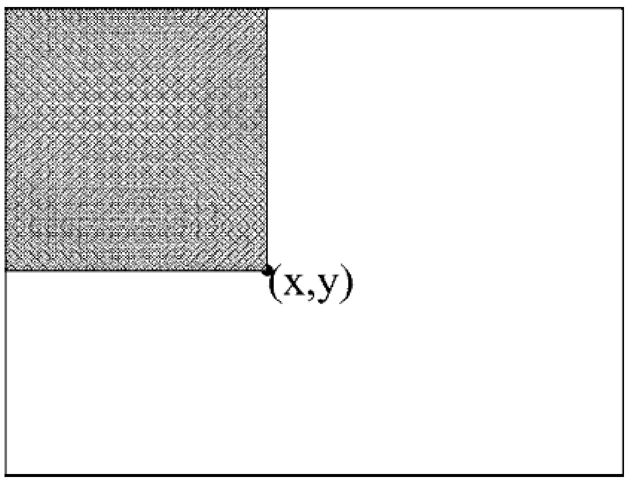
- 过完备特征
  - 从24*24像素的图像上可以提取多少个Haar-like features？
  - 160,000!

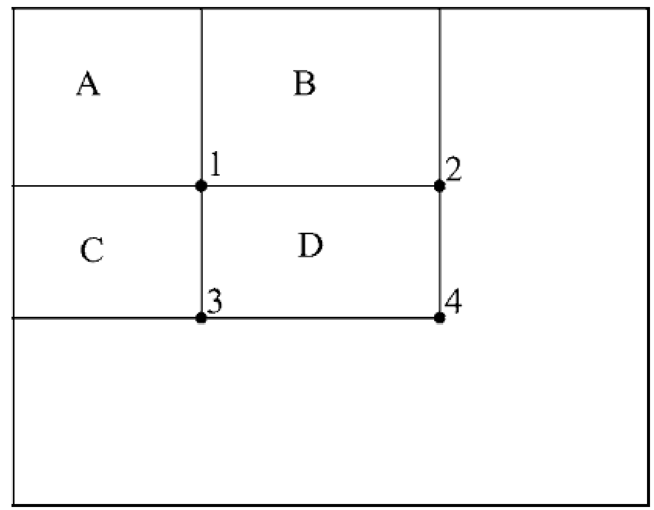
窗口大小	36×36	30×30	24×24	20×20	16×16
特征数量	816,264	394,725	162,336	78,460	32,384

  - 运算量过大

人脸检测

- 积分图像





$$ii(x,y) = \sum_{x' \le x, y' \le y} i(x', y')$$

$$s(x,y) = s(x, y-1) + i(x,y)$$

$$ii(x,y) = ii(x-1, y) + s(x,y)$$

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

人脸检测

- 分类器
  - **AdaBoost（Adaptive Boosting）分类器**
  - Boosting：是一种提高任意给定学习算法准确度的方法。这种方法依次顺序地加入多个分量分类器(弱分类器即可)，其中每个分量分类器的训练集都选自已有的其他各个分类器所给出的“最富信息(most informative)”的样本点组成，而最终的判决结果则是根据这些分量分类器的结果共同决定。
  - “三个臭皮匠，赛过诸葛亮”

人脸检测

- 分类器
  - **AdaBoost（Adaptive Boosting）分类器**
  - 允许设计者不断地加入新的弱分类器，直到达到某个预定的足够小的错误率。
  - 每个训练样本都被赋予一个权重，表明它被某个分量分类器选中的概率。如果每个样本点已经被准确地分类，那么在构造下一个训练集的时候，它被选中的概率就降低；相反，如果某个样本点没有被正确地分类，那么它的权重得到提高。

人脸检测

- 分类器
  - **AdaBoost（Adaptive Boosting）分类器**

Given: $(x_1, y_1), \dots, (x_m, y_m)$ where $x_i \in X$, $y_i \in Y = \{-1, +1\}$
Initialize $D_1(i) = 1/m$.
For $t = 1, \dots, T$:

- Train weak learner using distribution D_t.
- Get weak hypothesis $h_t : X \to \{-1, +1\}$ with error

$$\epsilon_t = \Pr_{i \sim D_t}\left[h_t(x_i) \neq y_i\right].$$

- Choose $\alpha_t = \frac{1}{2} \ln\left(\frac{1-\epsilon_t}{\epsilon_t}\right)$.
- Update:

$$\begin{aligned} D_{t+1}(i) &= \frac{D_t(i)}{Z_t} \times \begin{cases} e^{-\alpha_t} & \text{if } h_t(x_i) = y_i \\ e^{\alpha_t} & \text{if } h_t(x_i) \neq y_i \end{cases} \\ &= \frac{D_t(i)\exp(-\alpha_t y_i h_t(x_i))}{Z_t} \end{aligned}$$

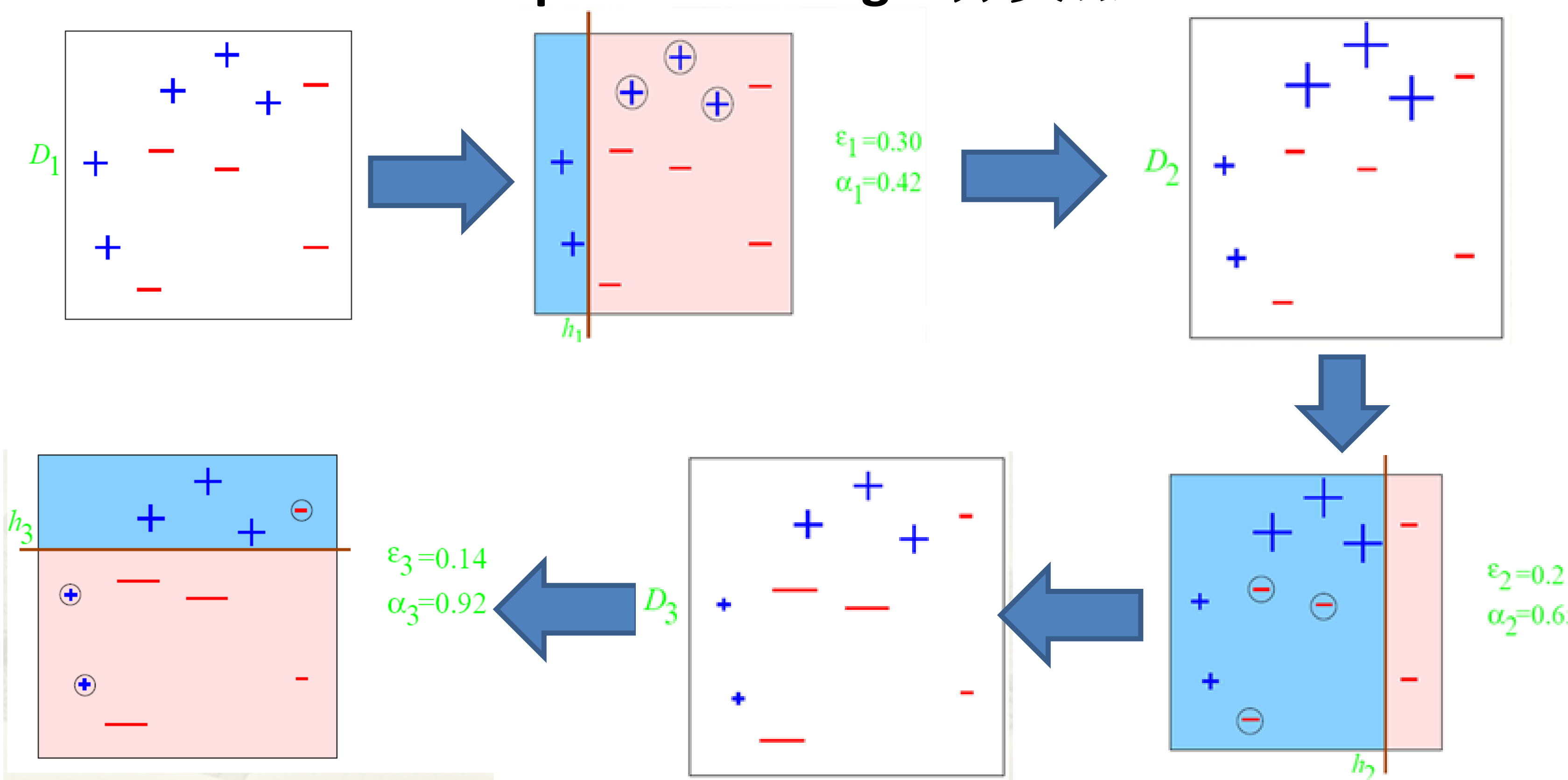
where Z_t is a normalization factor (chosen so that D_{t+1} will be a distribution).

Output the final hypothesis:

$$H(x) = \operatorname{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right).$$

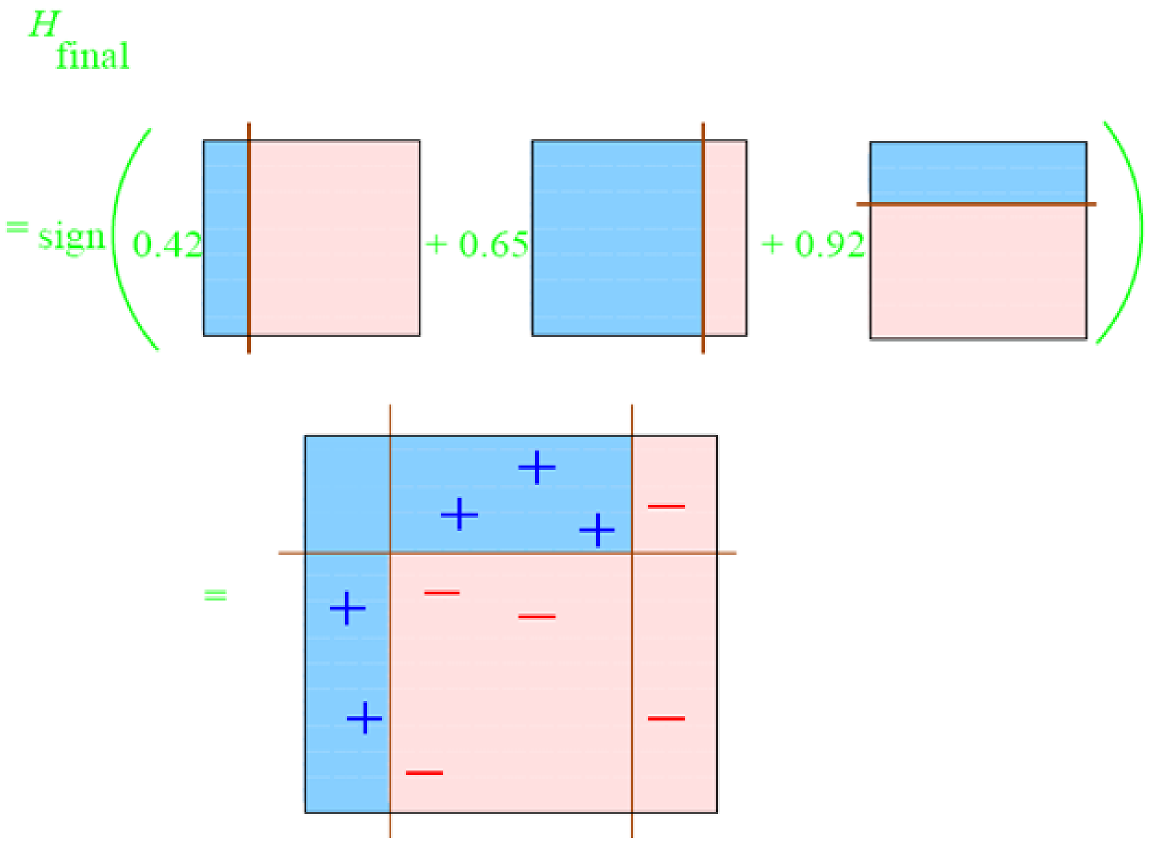
人脸检测

- 分类器
  - **AdaBoost（Adaptive Boosting）分类器**



人脸检测

- 分类器
  - **AdaBoost（Adaptive Boosting）分类器**



人脸检测

- 分类器
  - **AdaBoost（Adaptive Boosting）分类器**
  - 弱分类器
    - 决策桩(stumps)

$$h(x, f, p, \theta) = \begin{cases} 1 & \text{if } pf(x) < p\theta \\ 0 & \text{otherwise} \end{cases}$$

    - 每一个弱分类器由一个特征组成

人脸检测

- 分类器
  - **AdaBoost（Adaptive Boosting）分类器**
    - 给定图像样本$(x_1, y_1), \ldots, (x_n, y_n)$，其中$y_\mathrm{i}$为正负样本标记，0表示负样本，1表示正样本
    - 初始化权重$w_{1,i} = \frac{1}{2m}, \frac{1}{2n}$，(针对负样本和样本)，其中m和 n 为负样本和正样本的数量。

人脸检测

- 分类器
  - **AdaBoost（Adaptive Boosting）分类器**
    - For $t = 1, \ldots, T$:

      1. Normalize the weights, $w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^{n} w_{t,j}}$
      2. Select the best weak classifier with respect to the weighted error

         $$\epsilon_t = \min_{f,p,\theta} \sum_i w_i \,|\, h(x_i, f, p, \theta) - y_i \,| \,.$$

      3. Define $h_t(x) = h(x, f_t, p_t, \theta_t)$ where f_t, p_t, and θ_t are the minimizers of ϵ_t.
      4. Update the weights:

         $$w_{t+1,i} = w_{t,i} \beta_t^{1-e_i}$$

         where $e_i = 0$ if example x_i is classified correctly, $e_i = 1$ otherwise, and $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.

人脸检测

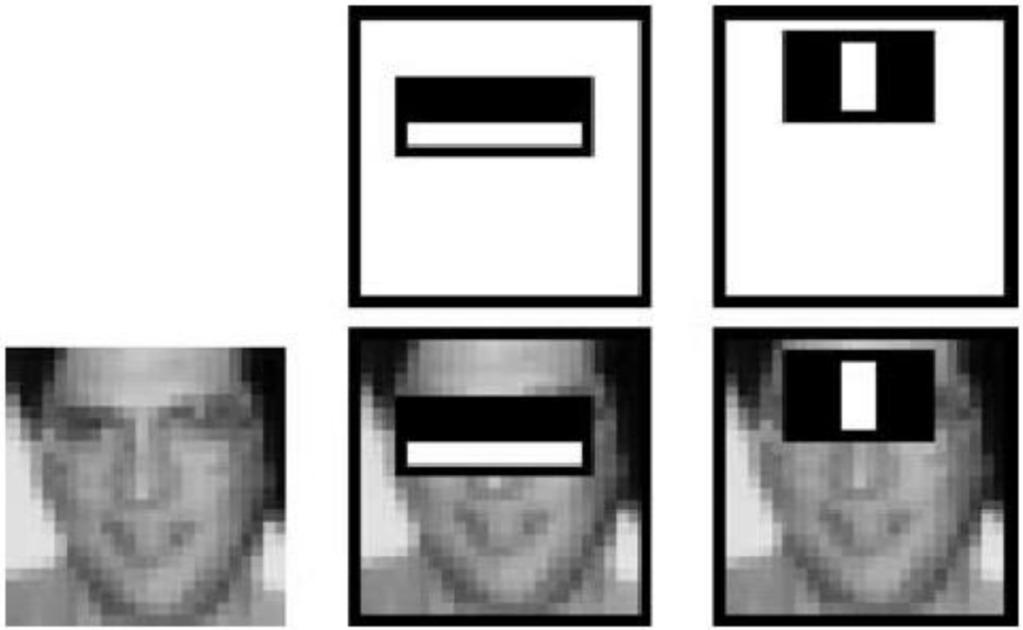
- 分类器
  - **AdaBoost（Adaptive Boosting）分类器**
    - **最后得到的分类器为：**

$$C(x) = \begin{cases} 1 & \sum_{t=1}^{T} \alpha_t h_t(x) \geq \frac{1}{2} \sum_{t=1}^{T} \alpha_t \\ 0 & \text{otherwise} \end{cases}$$

where $\alpha_t = \log \frac{1}{\beta_t}$

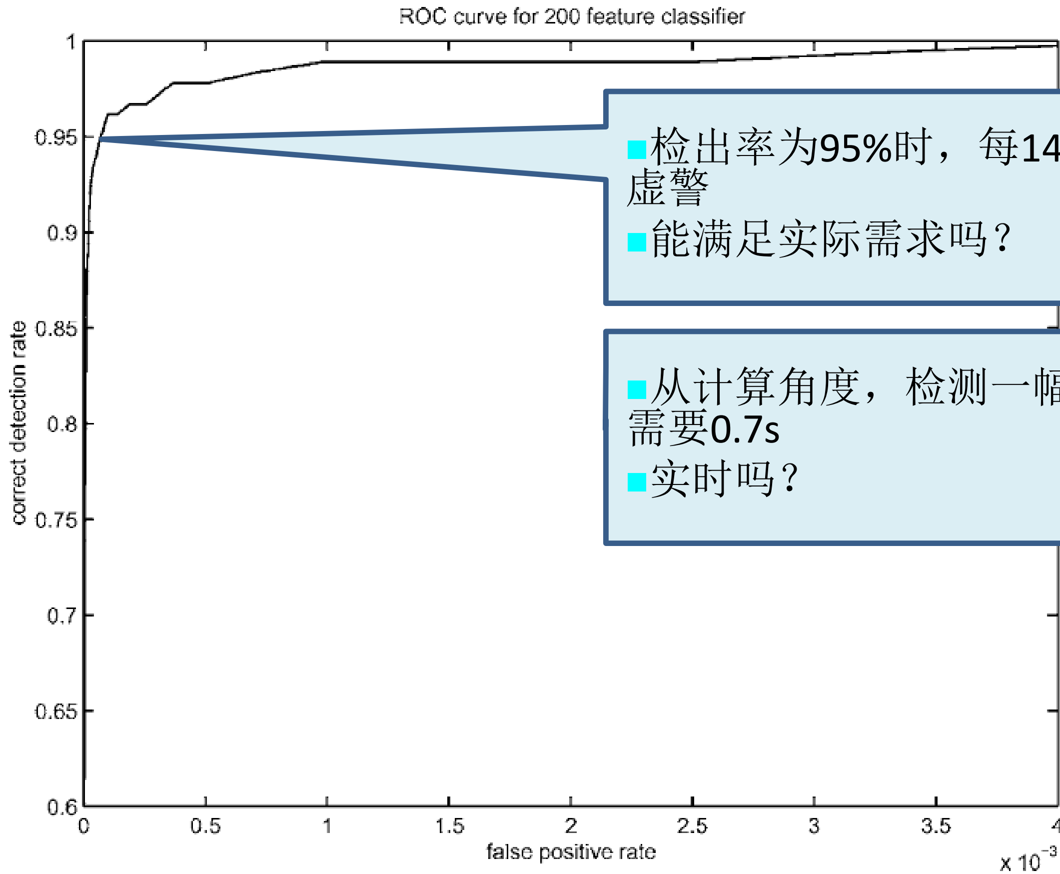
人脸检测

- 分类器
  - **AdaBoost（Adaptive Boosting）分类器**
  - 分类器学习 & 特征选择



人脸检测

- 分类器
  - 分类结果



■检出率为95%时，每14084次检测有1个虚警

■能满足实际需求吗？

■从计算角度，检测一幅384*288的图像需要0.7s

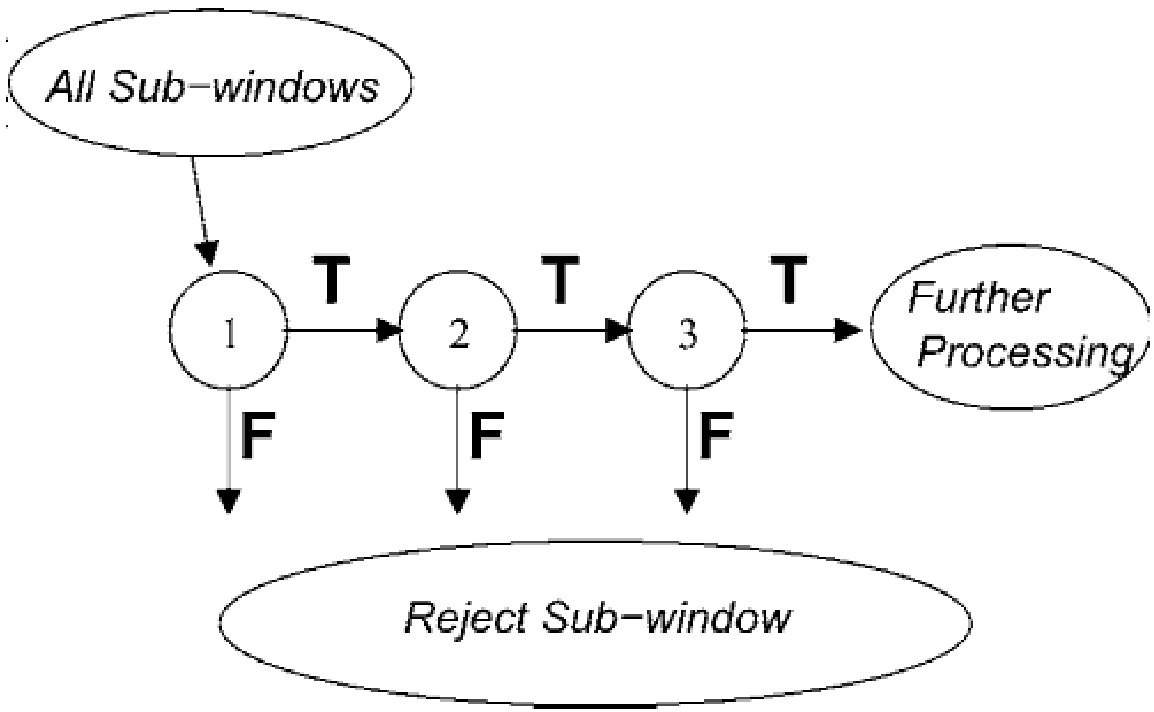
■实时吗？

人脸检测

- 层级式(Cascade)分类器
  - 对于一幅图像，只有极少量扫描窗中存在人脸。也就是说大多数扫描窗中都是非人脸。
  - 使用少量的特征的组成的分类器排除掉大部分非人脸的区域。
  - “先易后难”

人脸检测

- 层级式(Cascade)分类器



人脸检测

- 层级式(Cascade)分类器
  - 检测率和虚警率
  - False Positive Rate（假阳率）

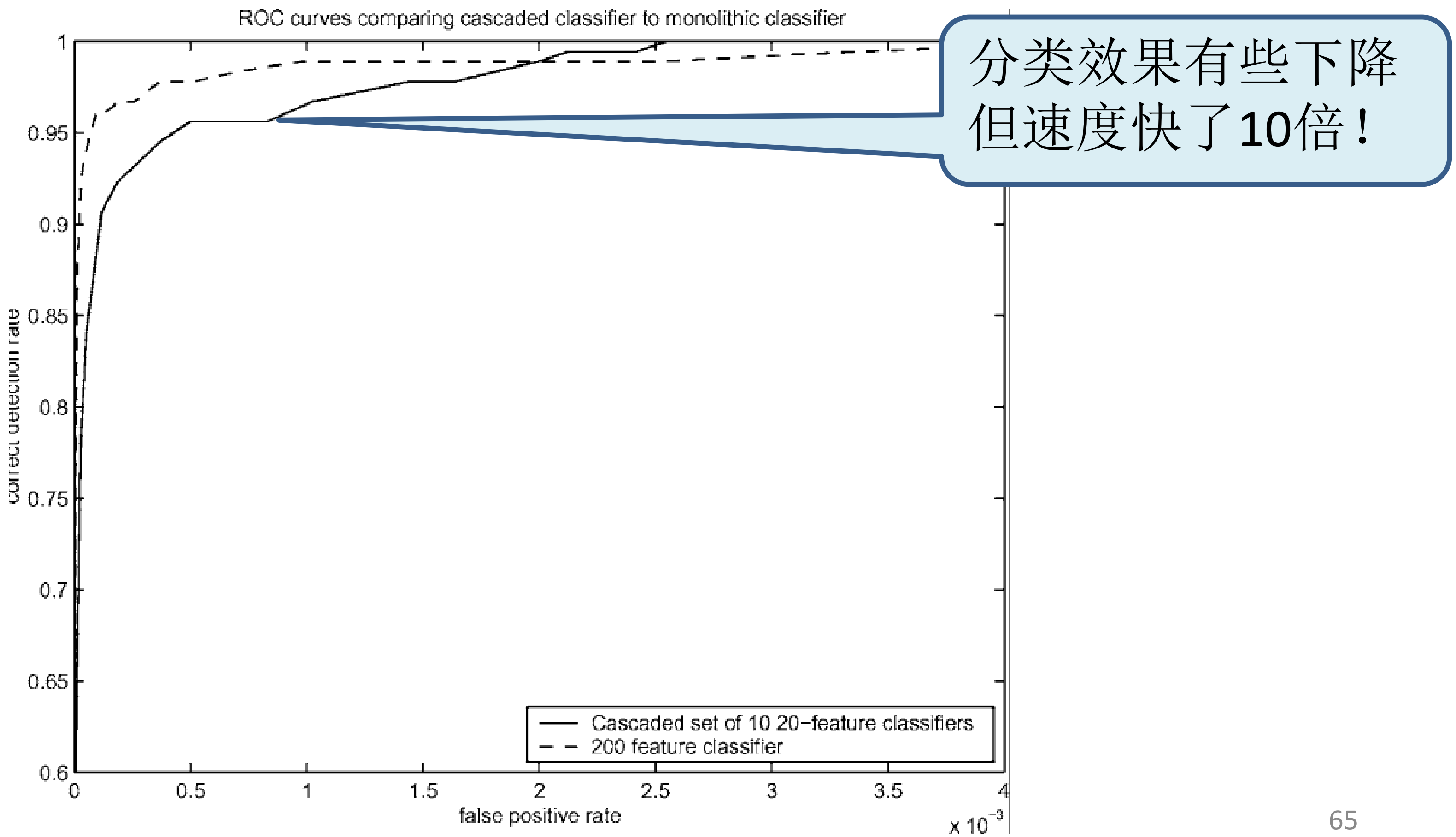
$$F = \prod_{i=1}^{K} f_i,$$

  - Detection Rate

$$D = \prod_{i=1}^{K} d_i,$$

人脸检测

- 层级式(Cascade)分类器



人脸检测

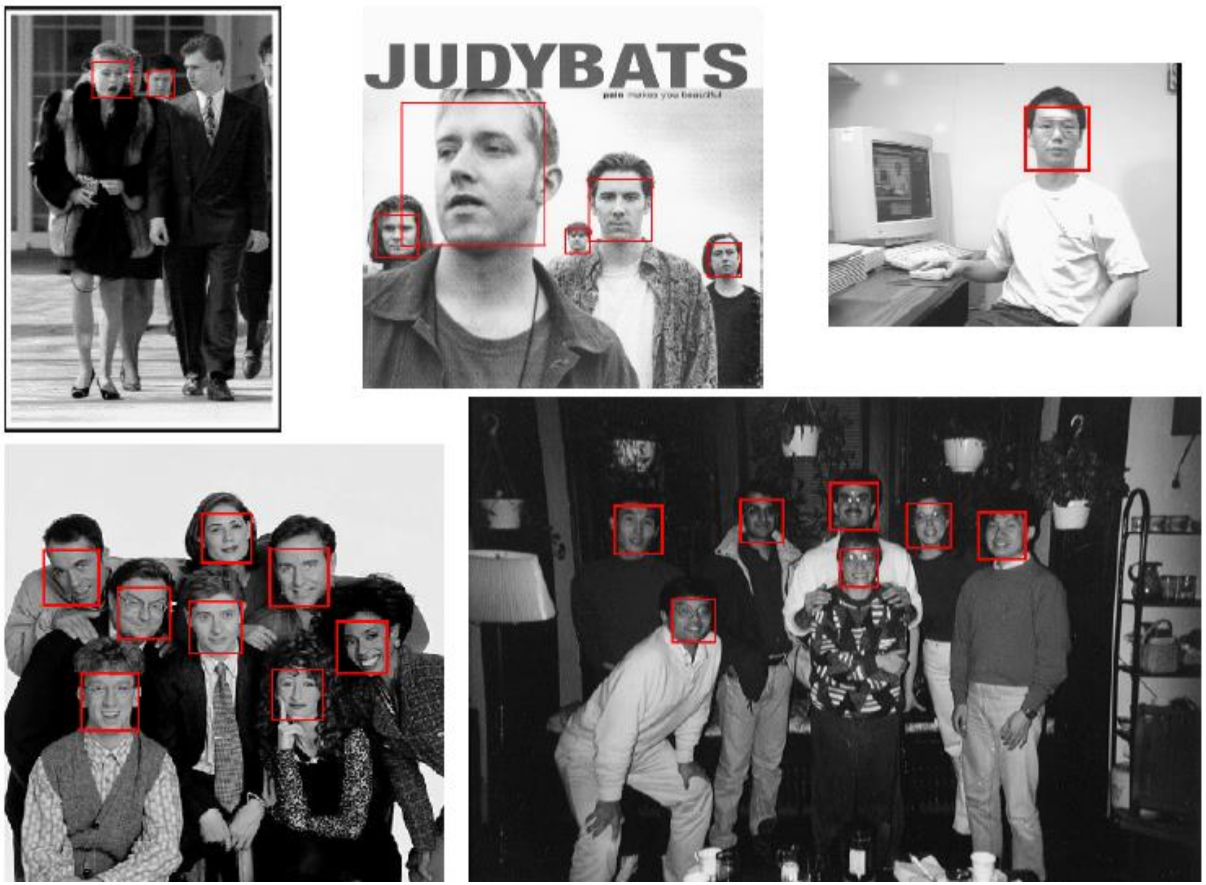
- 层级式(Cascade)分类器
  - 一般来说，使用较多的特征会得到检测率高、虚警率低的分类器。
  - 使用的特征越多，计算耗时越多
  - 如何折中？

人脸检测

- 层级式(Cascade)分类器
  - User selects values for f, the maximum acceptable false positive rate per layer and d, the minimum acceptable detection rate per layer.
  - User selects target overall false positive rate, F_{target}.
  - P = set of positive examples
  - N = set of negative examples
  - $F_0 = 1.0$; $D_0 = 1.0$
  - $i = 0$
  - while $F_i > F_{target}$
    - $i \leftarrow i + 1$
    - $n_i = 0$; $F_i = F_{i-1}$
    - while $F_i > f \times F_{i-1}$
      - $n_i \leftarrow n_i + 1$
      - Use P and N to train a classifier with n_i features using AdaBoost
      - Evaluate current cascaded classifier on validation set to determine F_i and D_i.
      - Decrease threshold for the ith classifier until the current cascaded classifier has a detection rate of at least $d \times D_{i-1}$ (this also affects F_i)
    - $N \leftarrow \emptyset$
    - If $F_i > F_{target}$ then evaluate the current cascaded detector on the set of non-face images and put any false detections into the set N

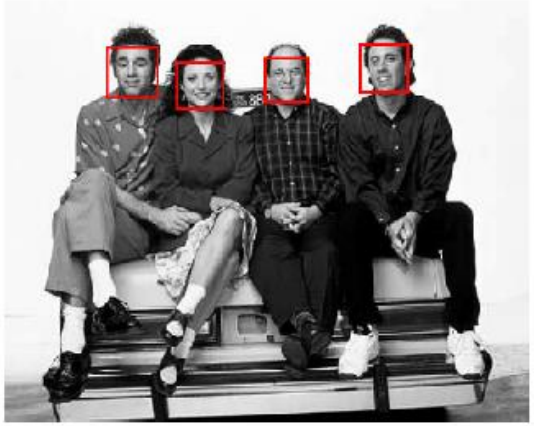
人脸检测

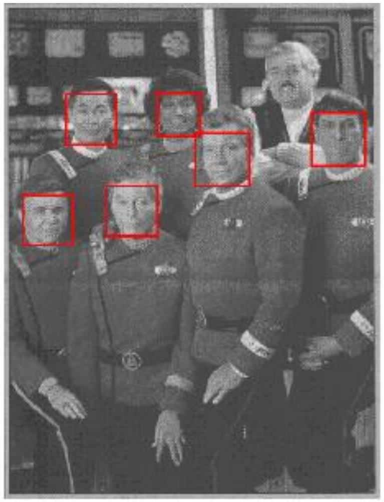
- 人脸检测结果

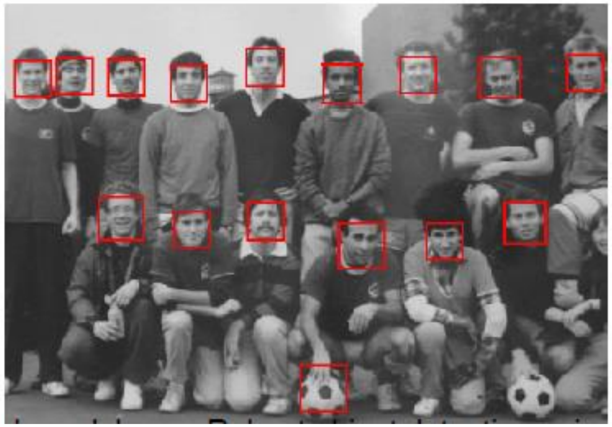


人脸检测

- 人脸检测结果





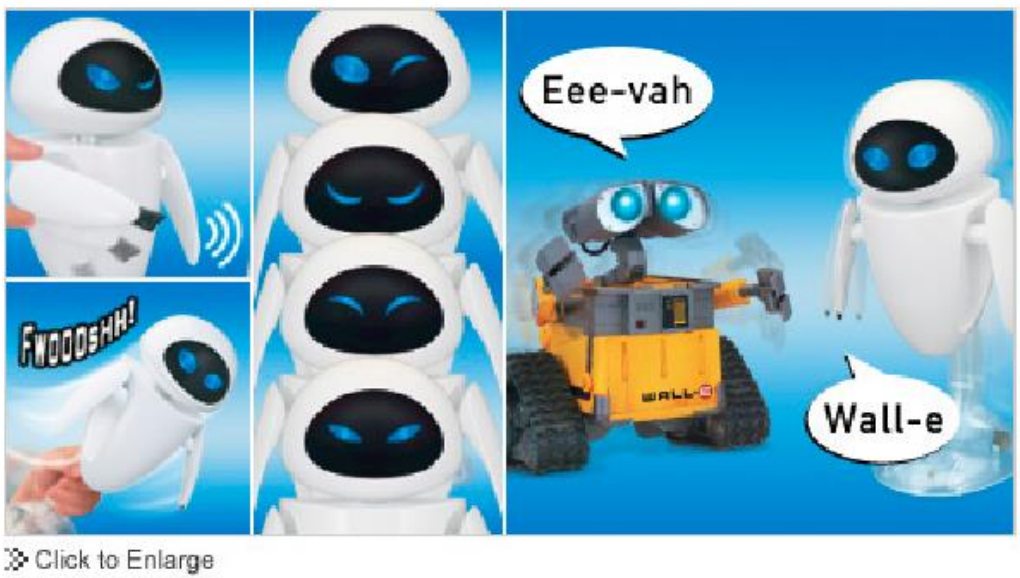


人脸检测

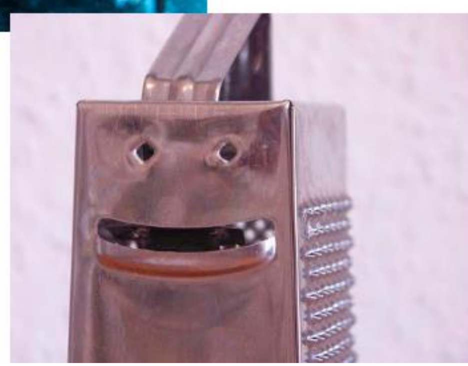
- 算法贡献总结
  - 特征（Haar-like, 过完备特征集）
  - 积分图像
  - 基于AdaBoost的特征选择方法。选择弱分类器即选择特征
  - 层级式分类器

人脸检测

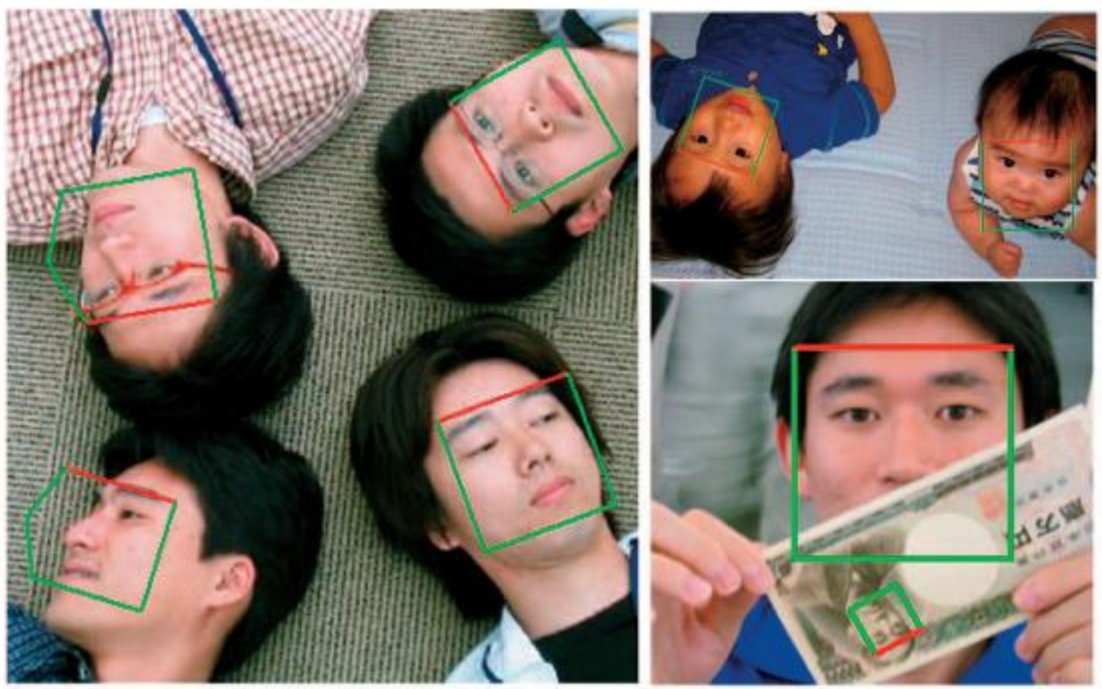

WALL·E
06.27.08
Disney · PIXAR


FWOOOSHH!
Eee-vah
Wall-e
Click to Enlarge

人脸检测



人脸检测



作业

- 分析HOG、颜色直方图、PCA、shapecontext、特征点+BOW这几种表达的平移、旋转（平面内）、尺度、遮挡不变性，并分析原因

	平移	旋转	尺度	遮挡
HOG				
颜色直方图				
PCA				
shapecontext				
特征点+BOW				

The end！