



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

# 人工智能技术导论

## 第五章 计算机视觉

# 目录



**5.1 计算机视觉概述**

**5.2 数字图像的表达**

**5.3 计算机视觉关键技术**



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 5.1 计算机视觉概述



# 计算机视觉概述



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

如果要给计算机视觉一个统一的定义，可以说：计算机视觉是一门研究如何对**数字图像或者视频进行智能理解**的交叉学科，它模拟了人类的视觉系统，让机器具备“会看”的能力。

## 如何模拟呢？

- 人类视觉系统首先有一个精密的成像系统——眼睛，实现了“**看到**”。在计算机视觉中，这一功能一般由光学系统、摄像头来模拟。
- 人类视觉系统还包括大脑，实现视觉信息的高效分析和理解，实现“**看懂**”。比如在一个场景中，人类视觉可以轻松区分场景中的不同物品、不同人物的面容、判断距离、理解文字和图案等等。
- 在人工智能研究领域，计算机视觉主要研究的是模拟大脑如何“看懂”的问题。

假设我们要设计一个视觉辅助的自动驾驶系统，计算机需要具有哪些功能才能“看懂”？

**前景背景分割：**区分天空、路面

**物体识别：**区分场景中的各种物体

**距离估计：**当前位置和周围物体之间的距离

**场景文字分割和识别：**  
包含文字信息的部分单独处理

**动态物体轨迹判断：**汽车、行人的速度、前进方向等



**“看懂”非常复杂！**

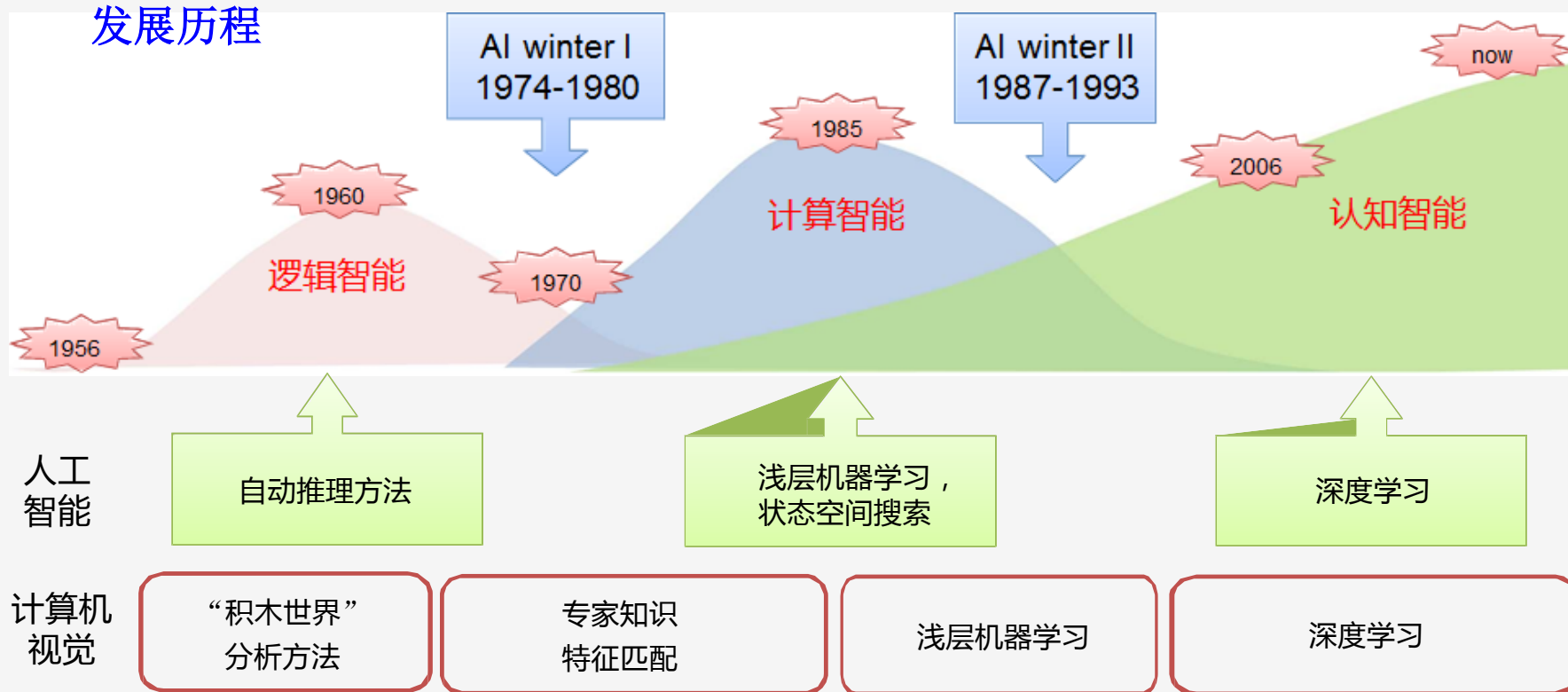


# 计算机视觉概述



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 发展历程



- 最早计算机视觉的研究，始于1960年代图像识别研究。
- 经历了几十年的变迁，计算机视觉研究的问题由简单到复杂，研究方法也随着人工智能的发展而变迁，但仍然具有典型的人工智能学科的特点：
  - (1) 从最开始就有明确的方向，甚至几十年研究的问题都一致，如**物体识别**、**人脸识别**、**三维重建**等。
  - (2) 方法更迭明显，在不同阶段均有主流方法。



# 计算机视觉概述



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 发展历程

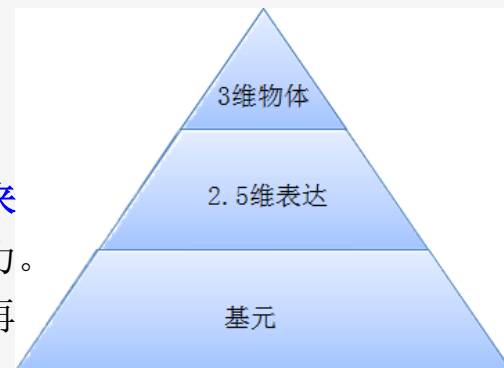
### ➤ 1960年代，积木世界分析

- 1963年，MIT的Roberts发表了一篇论文，通过计算机程序从数字图像中提取出诸如立方体、楔形体、棱柱体等多面体的三维结构。
- 该工作开创了以“**识别三维积木场景中的物体**”为目的的计算机视觉研究。学者们认为，如果积木世界中的物体可以被识别出来，则可推广到更复杂的三维场景物体识别中。
- 在当时，专家总结不同物体在图像中的特点，编写数据结构和规则，通过推理来实现识别。有人称为“**积木世界**”分析方法。

### ➤ 1977年，马尔视觉理论

- 1977年，David Marr提出计算机视觉理论，又称Marr视觉理论。
- 马尔认为，人的视觉主要功能在于“**从视网膜成像的二维图像来恢复空间物体的可见三维表面形状**”，称之为“**三维重建**”能力。因此，计算机视觉需要首先把三维结构从图像里面恢复出来，再去做理解和判断。
- 马尔认为，从图像到三维表达，要经过三个计算层次：

- (1) 从图像得到基元；
- (2) 通过立体视觉、运动、轮廓、光照等模块，得到2.5维表达；
- (3) 最后提升得到三维表达。







# 计算机视觉概述



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 发展历程

### ➤ 1980年代：特征匹配

- 1980年代的变革在于，人们发现要让计算机理解图像，不一定先要恢复物体的三维结构。直接从图像出发，同样可以实现视觉分析，最典型的的就是物体识别。

### 例如：让计算机识别苹果和橙子

- (1) 由专家来分析苹果和橙子的形状或其他特征，建立先验知识（特征库），如颜色、纹理、形状等。
- (2) 计算机计算图像中的物体的相应特征，并与先验知识匹配，一般为统计方法或者数值比较方法。
- (3) 如果满足匹配条件，就完成了识别。

#### (1) 建立专家知识



#### (2) 特征匹配



#### (3) 识别





# 计算机视觉概述



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

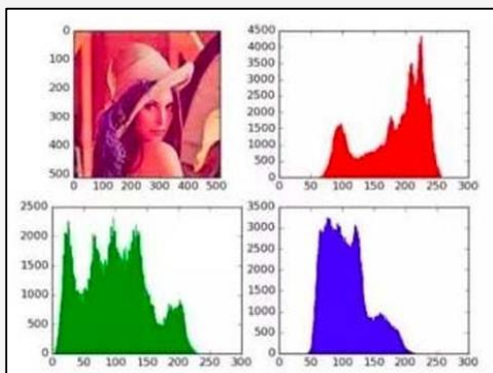
## 发展历程

### ➤ 1980年代：特征匹配

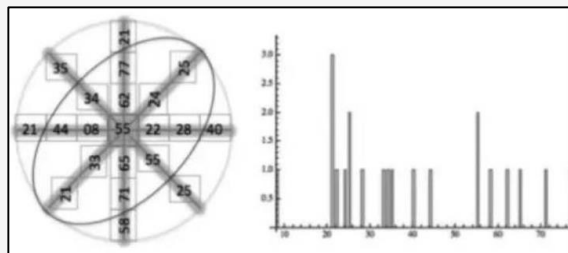
- 在这一阶段，围绕视觉特征，提出很多方法，将我们已知的物品转化成先验特征，然后用几何以及代数的方法在特征层面进行匹配。

常见的特征包括：

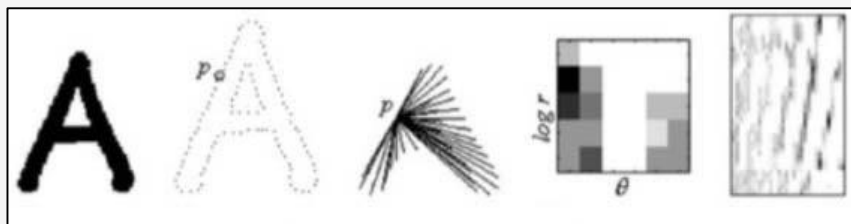
#### (1) 颜色特征



#### (2) 纹理特征



#### (3) 形状轮廓特征



- 同时，计算机视觉形成了以“特征”为核心的方法：找到一种合适的特征抽象方法，对一个待识别或者待分类问题进行特征表达，然后进行计算，给出识别结果。
- 到了1990年代，特征方法逐渐发展壮大，对物体的描述从整体特征细化为局部特征，通过统计大量局部特征得到整体特征，提高了识别准确率。产生了一系列实用系统。





# 计算机视觉概述



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 发展历程

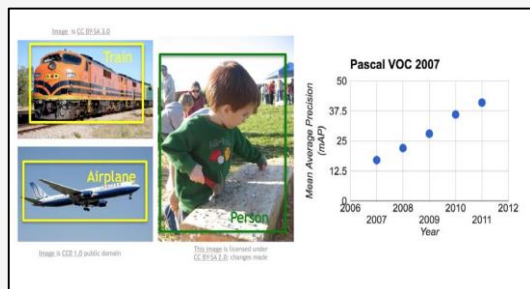
### ➤ 1990-2000年代，机器学习方法

- 到2000年左右，机器学习方法开始盛行，以其为核心的计算机视觉框架开始成为主流。
- 此前，实现计算机视觉需要专家规则、统计模型，**去匹配图像和特征**。专家既要选择特征知识，还需要制定判断规则。当引入机器学习以后，模型可以自行从海量数据里寻找最优的识别方法，从而提高了系统效率。
- 机器学习离不开大规模数据集。在这一时期，随着互联网时代的到来，视觉领域的大规模数据集也相伴而生。

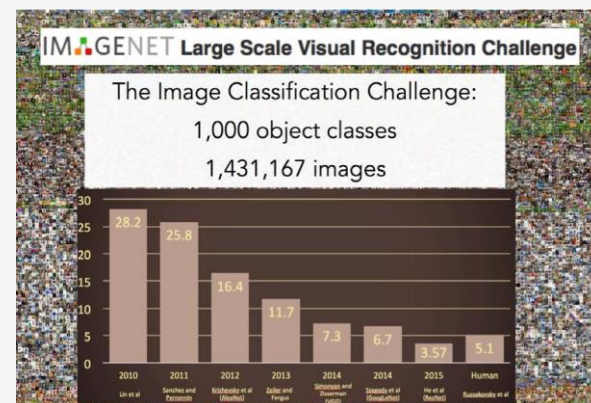
(1) 人脸检测的FDDB的数据集。这个数据集包含了超过5000多张人脸数据，每一张人脸，都人为的用框给框出来，机器就可以从这些框好的数据里面，通过机器学习的手段去学习人脸。



(2) PASCAL Visual Object Challenge, 该数据库中有20种类别图片，每种图片数量在一千至一万张不等。



(3) ImageNet 包含总计两万两千种类别，和一千四百余万张图片。长期成为计算机视觉的标准测试数据。





# 计算机视觉概述

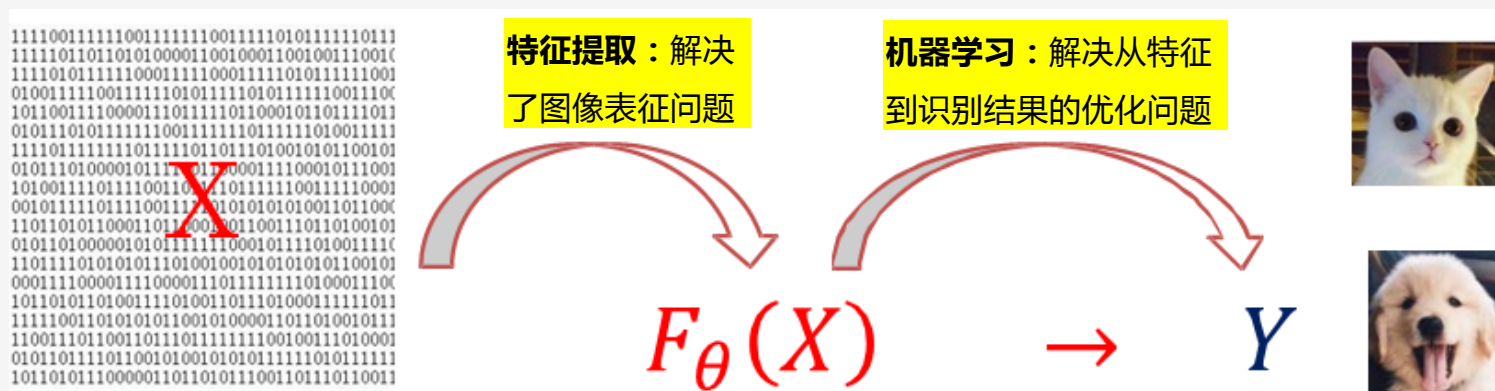


西安交通大学  
XI'AN JIAOTONG UNIVERSITY

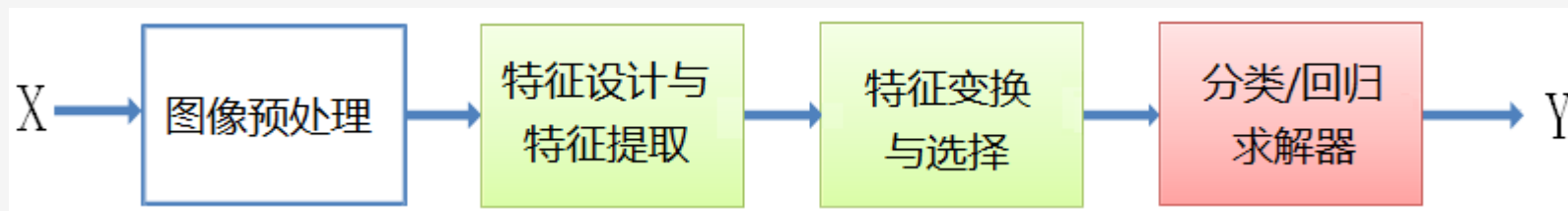
## 发展历程

### ➤ 1990-2000年代，机器学习方法

- 在这一时期，计算机视觉仍然以特征提取为核心。即，专家根据问题设计特征，实现原始图像的特征表达。然后借助大规模训练样本，使用统计机器学习方法实现特征的解析，求解最优的识别结果。



- 因此，在这个框架中，图像处理、特征提取、特征变换、机器学习，构成串联结构。





# 计算机视觉概述

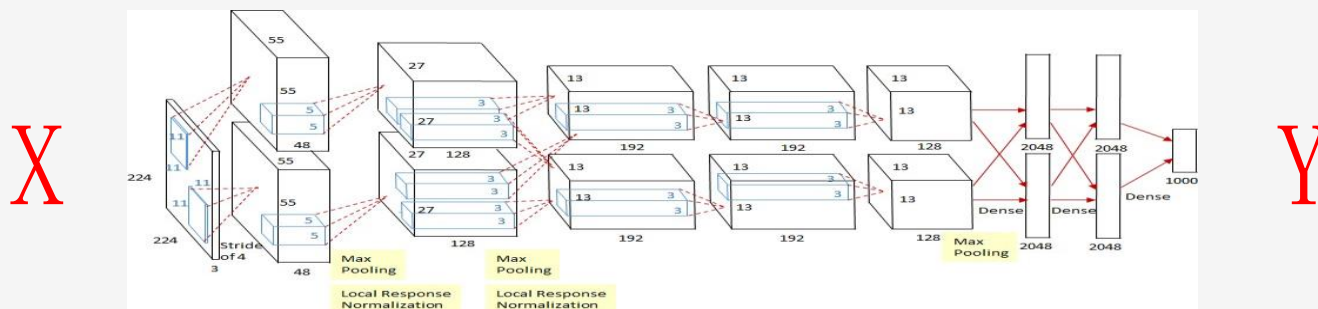
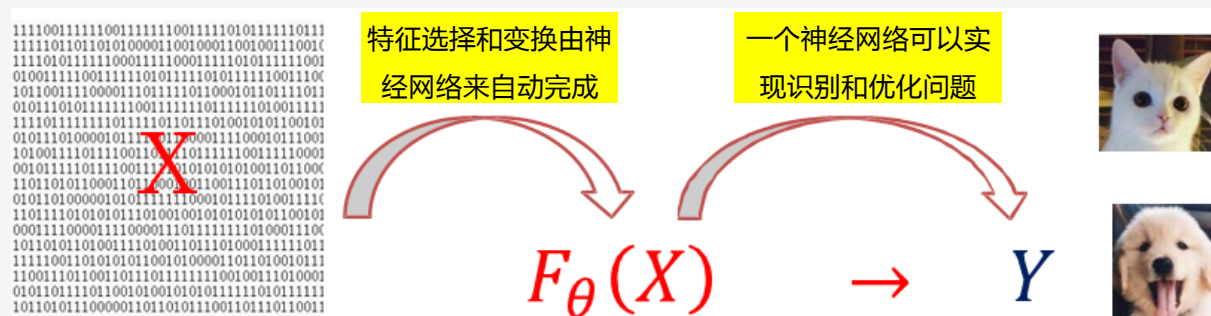


西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 发展历程

### ➤ 1990-2000年代，机器学习方法

- 机器学习方法在2010年后也逐渐遇到瓶颈。针对问题设计特征需要大量的经验，并且还需要根据应用场景进行大量的调试工作。随着问题的复杂化，机器学习模型的选择和调试也变得复杂，选择恰当特征、搭配恰当机器学习模型，达到最优的效果，变得日益困难。
- 此时计算机视觉有点脱离“模拟人类视觉”的初衷，陷入扩大数据，复杂化模型的循环。以现在的视角来看属于基于特征的浅层学习方法，随之而来的是基于深度学习的学习方法。



符合人类视觉系统结构



# 计算机视觉概述



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 应用领域

### ➤ 医学图像处理

- 计算机视觉最突出的应用领域是医疗和医学图像处理，从图像数据中提取用于使患者的医疗诊断的目的。通常，图像数据是显微镜图像，X射线图像，血管造影图像，超声图像和断层图像，可以从中提取的一个例子是检测肿瘤，动脉粥样硬化或其他恶性变化。它也可以是器官的尺寸，血流量等。这种应用领域还支持通过提供新的信息，医学研究的测量例如，对脑的结构，或医学治疗的质量。计算机视觉在医疗领域的应用还包括增强是由人类的解释，例如超声图像或X射线图像，以降低噪声的影响的图像。

### ➤ 工业领域

- 信息被提取为支撑的制造工序的目的。一个例子是质量控制，其中的信息或最终产品被以找到缺陷自动检测。另一个例子是，被拾取的位置和细节取向测量由机器臂。机器视觉也被大量用于农业的过程，如散装材料、食物的光学分拣。

### ➤ 军事应用

- 最明显的例子是探测士兵或车辆和制导。更先进的系统为制导发送区域，而不是一个特定目标，且当到达基于本地获取图像数据的区域的目标做出选择。现代军事概念，如“战场感知”，意味着各种传感器，包括图像传感器，提供了丰富的有关场景，可用于支持决策的信息。在这种情况下，数据的自动处理，用于减少复杂性和融合来自多个传感器的信息，以提高可靠性。

### ➤ 自主车

- 包括潜水，陆上车辆，高空作业车和无人机，其中基于计算机视觉的系统支持驱动程序或在不同情况下的试验。完全自主的汽车通常使用视觉导航时，用于生产的环境（地图SLAM）和用于检测障碍物。

### ➤ 其他领域

- 支持视觉特效制作的电影和广播，例如，摄像头跟踪（运动匹配）



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 5.2 数字图像表示





# 数字图像表示

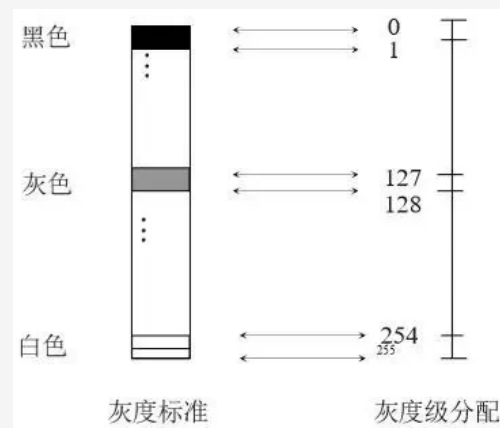
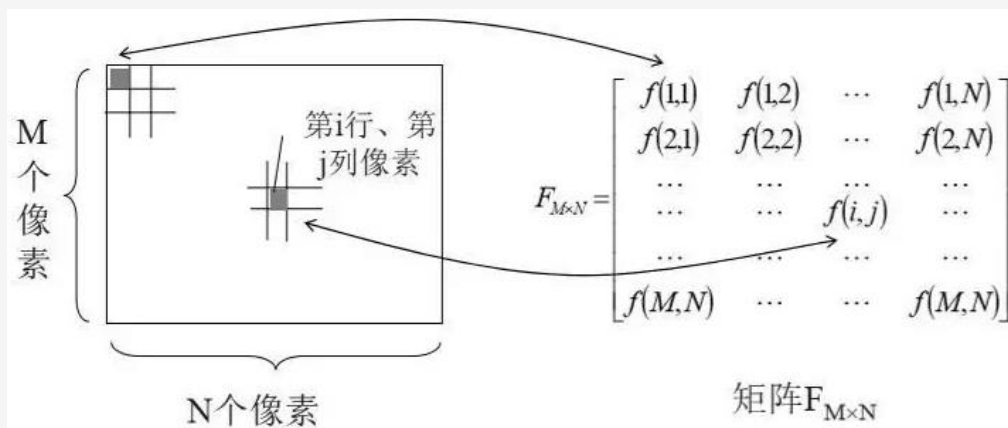


西安交通大学  
XI'AN JIAOTONG UNIVERSITY

要在计算机中处理图像，必须先把真实的图像（照片、画报、图书、图纸等）通过数字化转变成计算机能够接受的显示和存储格式，然后再用计算机进行分析处理。

具体就是在成像过程中，将一幅连续分布的图像先经过电视摄像机、转鼓、CCD电荷耦合器件和密度计等装置进行采样来获得**离散化空间位置坐标后的离散的像素**，再通过量化将**像素灰度转换成离散的整数值**，然后进行编码的过程。

空间坐标(x, y)的数字化称为图像采样，而幅值数字化被称为灰度级量化。经过数字化后的图像称为数字图像（或离散图像）。







# 数字图像的表达



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 数字图像与物理图像

- 数字图像是离散的，物理图像是连续的函数。
- **采样** —— 图像在空间上的离散化处理，即使空间上连续变化的图像离散化。
- **量化** —— 经过取样的图像，只是在空间上被离散为像素（样本）的阵列，而每一个样本灰度值还是一个有无穷多个取值的连续变化量，必须将其转化为有限个离散值，赋予不同码字才能真正成为数字图像，再由数字计算机或其它数字设备进行处理运算，这样的转化过程称其为量化。
- **分辨率** —— 灰度分辨率和空间分辨率。
  - 灰度分辨率是指值的单位幅度上包含的灰度级数，即在灰度级数中可分辨的最小变化。若用8比特来存储一幅数字图像，其灰度级为256。
  - 空间分辨率是指图像中可辨别的最小细节，采样间隔决定空间分辨率的主要参数。一般情况下，如果没有必要实际度量所涉及像素的物理分辨率和在原始场景中分析细节等级时，通常将图像大小 $M \times N$ ，灰度为 $L$ 级的数字图像称为空间分辨率为 $M \times N$ ，灰度级分辨率为 $L$ 级的图像。



# 数字图像的表示



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 数字图像与物理图像

- 图像噪声——妨碍人们感觉器官对所接收的信息理解的因素种类很多，如加性噪声、乘性噪声、白噪声和高斯噪声。

假定一幅图像取 $M \times N$ 个样点，对样点值进行 $Q$ 级分档取整。那么对 $M$ 、 $N$ 和 $Q$ 如何取值呢？

- 为了存取的方便， $Q$ 一般总是取成2的整数次幂，如 $Q=2^b$ ， $b$ 为正整数，通常称为对图像进行 $b$ 比特量化。
- 对 $b$ 来讲，取值越大，重建图像失真越小。
- 对 $M \times N$ 的取值，主要的依据是取样的约束条件，也就是在 $M \times N$ 达到满足取样定理的情况下，重建图像就不会产生失真，否则就会因取样点数不够而产生所谓混淆失真。





# 数字图像表示



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 灰度图像表示

$$f(x, y) = \begin{bmatrix} f(0,0) & f(0,1) & \cdots & f(0,N-1) \\ f(1,0) & f(1,1) & \cdots & f(1,N-1) \\ \vdots & \vdots & & \vdots \\ f(N-1,0) & f(N-1,1) & \cdots & f(N-1,N-1) \end{bmatrix}$$

像素 (pixel)

灰度图像 (128x128) 及其对应的数值矩阵 (仅列出一部分)



```

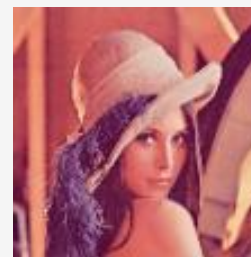
125, 153, 158, 157, 127,
70, 103, 120, 129, 144, 144, 150, 150, 147, 150, 160, 165,
160, 164, 165, 167, 175, 175, 166, 133, 60,
133, 154, 158, 100, 116, 120, 97, 74, 54,
74, 118, 146, 148, 150, 145, 157, 164, 157, 158, 162, 165.

```

## 彩色图像表示

- 彩色图像可以用红、绿、蓝三元组的二维矩阵来表示。
- 三元组的每个数值也是在0到255之间，0表示相应的基色在该像素中没有，而255则代表相应的基色在该像素中取得最大值，这种情况下每个像素可用三个字节来表示。

彩色图像 (128x128) 及其对应的数值矩阵 (仅列出一部分)



```

(207, 137, 130) (220, 179, 163) (215, 169, 161)
(210, 179, 172) (210, 179, 172) (207, 154, 146)
(217, 124, 121) (226, 144, 133) (226, 144, 133)
(224, 137, 124) (227, 151, 136) (227, 151, 136)
(226, 159, 142) (227, 151, 136) (230, 170, 154)
(231, 178, 163) (231, 178, 163) (231, 178, 163)
(236, 187, 171) (236, 187, 171) (239, 195, 176)

```



# 数字图像表示



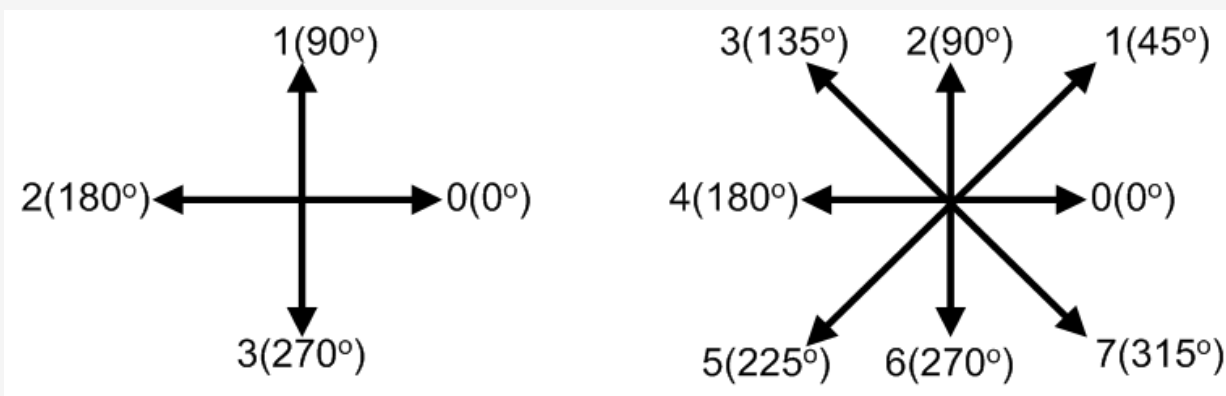
西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 二值图像表示

- 为了减少计算量，常将灰度图像转为二值图像处理。

**定义：**只有黑白两个灰度级，即像素灰度级非1即0，如文字图片，其数字图像可用每个像素1bit的矩阵表示。

- **二值图像的特殊表示法：**如链码 (Freeman码) —— 适合表示直线和曲线组成的二值图像，以及描述图像的边界轮廓。
- 采用链码节省很多的比特数 —— 规定了链的起点坐标和链的斜率序列，这样就可以完全描述曲线和直线。



八向链码



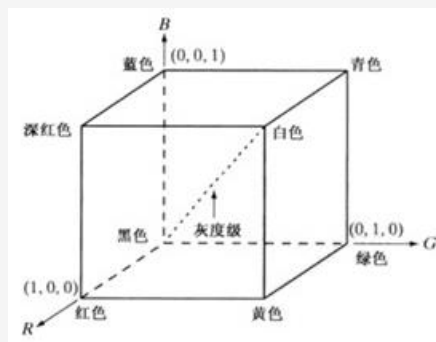
# 数字图像的表示



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## RGB彩色空间

- RGB颜色模型主要应用于监视器和图形刷新设备中。该彩色空间是最普遍的，R、G、B三色之间存在强烈的相关性。

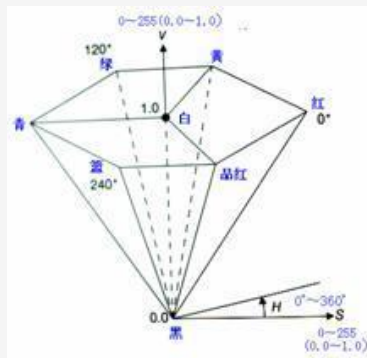


## RGBD=RGB+Depth Map

- **RGB:** RGB色彩模式是工业界的一种颜色标准，是通过对红(R)、绿(G)、蓝(B)三个颜色通道的变化以及它们相互之间的叠加来得到各式各样的颜色的，RGB即是代表红、绿、蓝三个通道的颜色。
- **Depth Map:** 在3D计算机图形中，Depth Map（深度图）是包含与视点的场景对象的表面的距离有关的信息的图像或图像通道中，Depth Map类似于灰度图像，只是它的每个像素值是传感器距离物体的实际距离。通常RGB图像和Depth图像是配准的，因而像素点之间具有一对一的对应关系。

## HSV颜色空间

- HSV（色调Hue，饱和度Saturation，明度Value）是根据颜色直观特性由A. R. Smith 在1978年创建的一种颜色空间，也称六角锥体模型(Hexcone Model)。



RGB --> HSV 颜色表

Color	Color name	Hex	(R,G,B)	(H,S,V)
	Black	#000000	(0,0,0)	(0°,0%,0%)
	White	#FFFFFF	(255,255,255)	(0°,0%,100%)
	Red	#FF0000	(255,0,0)	(0°,100%,100%)
	Lime	#00FF00	(0,255,0)	(120°,100%,100%)
	Blue	#0000FF	(0,0,255)	(240°,100%,100%)
	Yellow	#FFFF00	(255,255,0)	(60°,100%,100%)
	Cyan	#00FFFF	(0,255,255)	(180°,100%,100%)
	Magenta	#FF00FF	(255,0,255)	(300°,100%,100%)
	Silver	#C0C0C0	(192,192,192)	(0°,0%,75%)
	Gray	#808080	(128,128,128)	(0°,0%,50%)
	Maroon	#800000	(128,0,0)	(0°,100%,50%)
	Olive	#808000	(128,128,0)	(60°,100%,50%)
	Green	#008000	(0,128,0)	(120°,100%,50%)
	Purple	#800080	(128,0,128)	(300°,100%,50%)
	Teal	#008080	(0,128,128)	(180°,100%,50%)
	Navy	#000080	(0,0,128)	(240°,100%,50%)



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 5.3 计算机视觉关键技术





# 计算机视觉关键技术

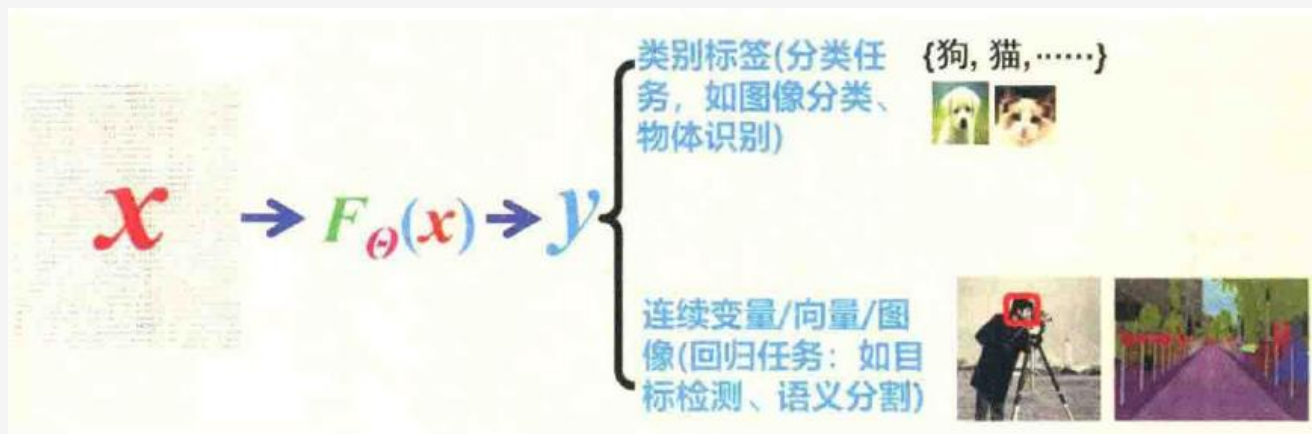


西安交通大学  
XI'AN JIAOTONG UNIVERSITY

尽管计算机视觉任务繁多，但大多数本质上可以建模为广义的函数拟合问题，即对任意输入图像 $x$ ，需要学习一个以 $\theta$ 为参数函数 $F$ ，使得 $y=F_{\theta}(x)$ ，其中 $y$ 可能有两大类：

**(1)  $y$ 为类别标签：**对应模式识别或机器学习中的“分类”问题，如场景分类、图像分类、物体识别、精细物体类识别、人脸识别等视觉任务。这类任务的特点是输出 $y$ 为有限种类的离散型变量。

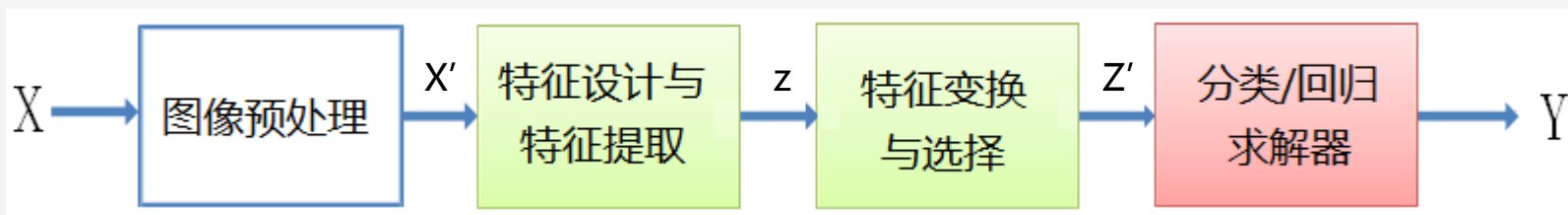
**(2)  $y$ 为连续变量或向量或矩阵：**对应模式识别或机器学习中的“回归”问题，如距离估计、目标检测、语义分割等视觉任务。在这些任务中， $y$ 或者是连续的变量(如距离、年龄、角度等)，或者是一个向量(如物体的横纵坐标位置和长宽)，或者是每个像素有一个所属物体类别的编号(如分割结果)。





## 基于浅层模型的方法

实现上述视觉任务的函数  $F_\theta$  通常都是非常复杂的。为此，一种可能的解法是遵循“分而治之”的思想，对其进行分步、分阶段求解，如下图所示，一个典型的视觉任务实现流程包括以下四个步骤。



常见的浅层视觉模型处理流程

### ➤ 步骤1：图像预处理过程 $p$ 。

用于实现目标对齐、几何归一化、亮度或颜色矫正等处理，从而提高数据的一致性，该过程一般由人为设定。

### ➤ 步骤2：特征设计与提取过程 $q$ 。

其功能是预处理后的图像  $x'$  中提取描述图像内容的特征，这些特征可能反映图像的低层(如边缘)、中层(如部件)或高层(如场景)特性，一般依据专家知识进行人工设计。



## 基于浅层模型的方法

### ➤ 步骤3：特征汇聚或特征变换h。

其功能是对前步提取的局部特征 $z$  (一般是向量) 进行统计汇聚或降维处理，从而得到维度更低、更利于后续分类或回归过程的特征 $z$ 。该过程一般通过专家设计的统计建模方法实现。例如，一种常用的模型是线性模型，即 $z' = Wz$ ，其中 $W$ 为矩阵形式表达的线性变换，一般需要在训练集合进行学习得到。

### ➤ 步骤4：分类器或回归器函数 $g$ 的设计与训练。

其功能是采用机器学习或模式识别的方法，基于一个有导师的训练集 $\{((x_i, y_i): i=1, \dots, N)\}$  (其中 $x_i$ 是训练图像， $y_i$ 是其类别标签) 学习得到，通过有监督的机器学习方法来实现。例如，假设我们采用线性模型，即 $y = Wz'$ ，则可以通过优化 $W^* = \arg \min_W \sum_{i=1}^N \|y_i - Wz'_i\|_2$  得到，其中 $z'_i$ 为通过步骤3得到的 $x_i$ 的特征。

上述流程可理解为通过序贯执行 $p, q, h, g$ 四个函数实现需要的 $y = F_\theta(x)$ ，即 $y = g(h(q(p(x))))$ 。不难发现该流程带有强烈的“人工设计”色彩，不仅依赖专家知识进行步骤划分，更依赖专家知识选择和设计各步骤函数，这与后来出现的深度学习方法依赖大量数据进行端到端的自动学习(即直接学习 $F_\theta$ 函数)形成了鲜明对比。为了深度学习在概念上进行区分，通常称这些模型为**浅层视觉模型**。考虑到步骤1的图像预处理往往依赖于图像类型和任务，接下来仅对后面三个步骤进行概要阐述。



## 基于浅层模型的方法

### 1. 特征设计与提取方法

人工设计特征**本质是一种专家知识驱动**的方法，即研究者自己或通过咨询特定领域专家，根据对所研究问题或目标的理解，设计某种流程来提取专家觉得“好”的特征。例如，在人脸识别研究早期，研究人员普遍认为应用面部关键特征点的相对距离、角度或器官面积等就可以区分不同的人脸，但后来的实践很快证明了这些特征并不好。

目前，多数人工设计的特征有两大类，即**全局特征和局部特征**。前者通常建模的是图像中全部像素或多个不同区域像素中所蕴含的信息，后者则通常只从一个局部区域内的少量像素中提取信息。

典型的全局特征对颜色、全图结构或形状等进行建模，例如在全图上计算颜色直方图，傅立叶频谱也可以看作全局特征。

另种典型的全局场景特征是20年Ande oliva和Anonio Torralba提出的GIST特征，它主要对图像场最的空间形状属性进行建模，如自然度、开放度、粗糙度、扩张度和崎岖度等。与局部特征相比，全局特征往往粒度比较粗，适合于需要高效而无须精细分类的任务，比如场景分类成大规模图像检索等。



## 基于浅层模型的方法

### 1. 特征设计与提取方法

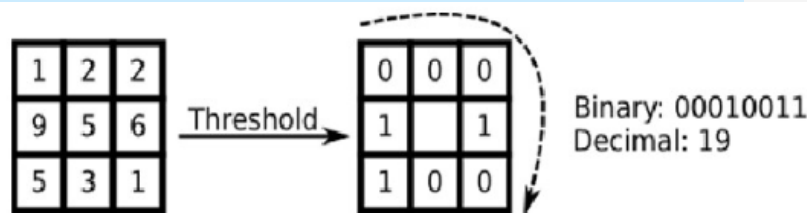
相对而言、局部特征可以提取更为精细的特征，应用更为广泛，也因此2000年之后的十年得到充分发展，研究人员设计出了数以百计的局部特征。这些局部特征大多数以建模边缘、梯度、纹理等为目标，采用的手段包括滤波器设计、局部统计量计算、直方图等。最典型的局部特征有SIFT、SURF、HOG、LBP、Cabor滤波器、DAISY、BRIEF、ORB、BRISK等数十种，下面以LBP为例详细介绍其提取方法。

#### ◆ 局部二值模式( local binary patterns, LBP)

LBP是一种简单有效地编码图像局部区域内变化模式(即微纹理)的局部描述子。与其他对图像梯度强度和方向进行精细统计的特征不同，LBP只关注梯度的符号，换句话说，它只关注中心像素与其邻域像素的明暗关系。

以3x3邻域组成的9个像素关系为例，LBP比较中心像素与其8邻域像素的亮度值大小：某邻域像素值大于等于中心像素值则赋1，否则赋0，从而得到8个0/1位，串接成1个字节即得到一个 [0, 255] 区间内的十进制数。

不难理解，这相当于把3x3共9个像素组成的局部邻域编码成了256种不同的模式类型。



LBP算子

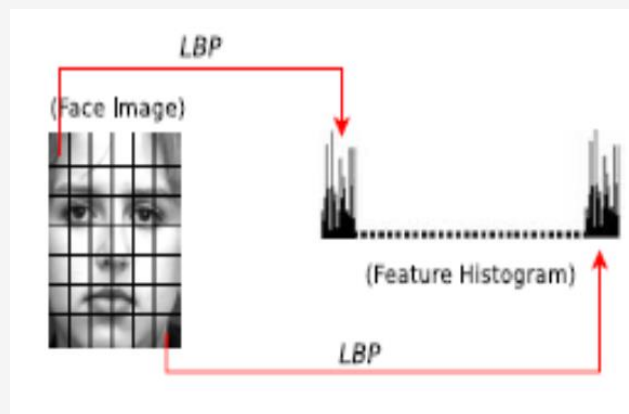


## 基于浅层模型的方法

### 1. 特征设计与提取方法

#### ◆ 局部二值模式 (local binary patterns, LBP)

得到了每个像素分割的LBP。建立一个直方图，每个LBP的种类作横轴，每种LBP出现的次数为纵轴。这张图可以反映该区域的纹理信息。

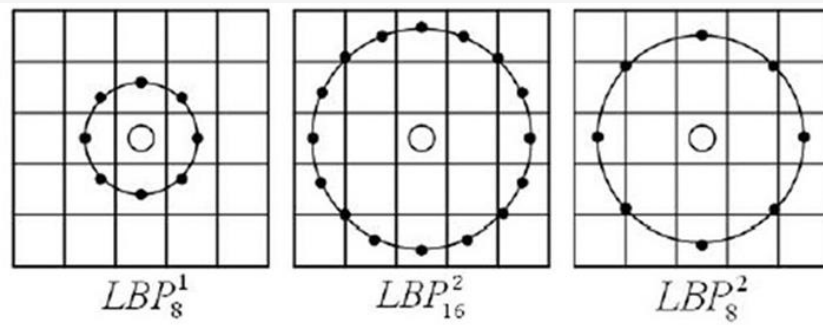


由于基本LBP算子的局限性，研究人员不断对其进行改进，比如：

#### ① 圆形LBP算子

基本的LBP算子的最大缺陷在于它只覆盖了一个固定半径范围内的小区域，这显然不能满足不同尺寸和频率纹理的需要。为了适应不同尺度的纹理特征，并达到灰度和旋转不变性的要求，Ojala等对 LBP 算子进行了改进。

将  $3 \times 3$  邻域扩展到任意邻域，并用圆形邻域代替了正方形邻域，改进后的 LBP 算子允许在半径为  $R$  的圆形邻域内有任意多个像素点。从而得到了诸如半径为  $R$  的圆形区域内含有  $P$  个采样点的LBP算子。



几种LBP算子

符号  $LBP_P^R$  表示半径为  $R$  的圆形邻域内有  $P$  个像素点  $g_p$  ( $p=0, 1, \dots, P$ ) 的LBP算子。





## 基于浅层模型的方法

### 1. 特征设计与提取方法

#### ◆ 局部二值模式( local binary patterns, LBP)

##### ② LBP等价模式

一个LBP算子可以产生不同的二进制模式，对于半径为 $R$ 的圆形区域内含有 $P$ 个采样点的LBP算子将会产生 $2^P$ 种模式。随着邻域集内采样点数的增加，二进制模式的种类是急剧增加的。例如： $5 \times 5$ 邻域内20个采样点，有 $2^{20} = 1,048,576$ 种二进制模式。

如此多的二值模式无论对于纹理提取还是对于纹理识别、分类及信息的存取都是不利的，同时对于纹理的表达也是不利的。例如，将LBP算子用于纹理分类或人脸识别时，常采用LBP模式的统计直方图来表达图像信息，而较多的模式种类将使得数据量过大，且直方图过于稀疏。因此，需要对原始的LBP模式进行降维，使得数据量减少的情况下能最好的代表图像的信息。

上述256种二值模式出现的概率是有差异的，为了获得鲁棒性并减少模式类别数，LBP的设计者们定义了均衡模式(uniform patterns)和非等价模式的概念。所谓等价模式，是指0/1串中最多包含2次1-0或0-1跳变的模式，例如00000000、00001111、01111000都是均衡模式，分别包含0次、1次、2次跳变，这样可以得到58种不同的均衡模式。而01010000、00110011、01001101分别包含了4次、3次、5次跳变，不是均衡模式。鉴于非均衡模式在自然图像中出现的非常少，它们被强制归为一类模式，从而共得到59种不同的二值模式。这使得特征向量的维数更少，并且可以减少高频噪声带来的影响。



## 基于浅层模型的方法

### 1. 特征设计与提取方法

#### ◆ 局部二值模式( local binary patterns, LBP)

##### ③ 旋转不变模式

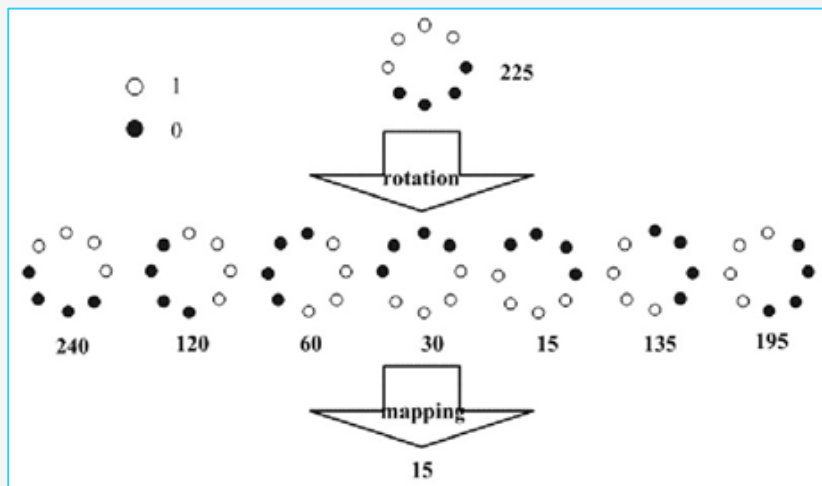
从LBP 的定义可以看出, LBP 算子是灰度不变的, 但却不是旋转不变的。图像的旋转就会得到不同的 LBP值。Maenpaa等人又将 LBP算子进行了扩展, 提出了具有旋转不变性的 LBP 算子, 即不断旋转圆形邻域得到一系列初始定义的 LBP值, 取其最小值作为该邻域的 LBP 值。

用公式表示如下:

$$LBP_{P,R}^{ri} = \min(ROR(LBP_{P,R}^{ri}, i) | i = 0, 1, \dots, P-1)$$

其中,  $LBP^{ri}$  表示旋转不变的 LBP 算子,  $ROR(x, i)$ 函数为旋转函数, 表示将  $x$ 循环右移  $i(i < P)$  位。

右图给出了求取旋转不变的LBP的过程示意图, 图中算子下方的数字表示该算子对应的LBP值, 图中所示的8种LBP模式, 经过旋转不变的处理, 最终得到的具有旋转不变性的LBP值为15。也就是说, 图中的8种LBP模式对应的旋转不变的LBP模式都是00001111。





## 基于浅层模型的方法

### 1. 特征设计与提取方法

#### ◆ 局部二值模式( local binary patterns, LBP)

##### ③ 旋转不变模式

通过引入旋转不变的定义, LBP 算子不仅对于图像旋转表现得更为鲁棒, 并且 LBP 模式的种类进一步减少, 使得纹理识别更加容易。从另一方面来说, 旋转不变的 LBP 算子的最大缺陷在于其丢失了方向信息, 而对于某些场合来说, 方向信息是十分重要的信息。然而, 对于同类纹理图像的分析, 算子已经被证明是有效的。

LBP<sup>rii</sup> 算子还可以与等价模式联合起来, 将等价模式类进行旋转得到旋转不变的等价模式, 这使得可能的模式种类由2P类减少为P+1类。所有的非等价模式被归为第 P+1类, 如右式所示。

$$LBP_{P,R}^{riu2} = \begin{cases} \sum_{p=0}^{P-1} s(g_p - g_c), & U(G_p) \leq 2 \\ P+1, & U(G_p) > 2 \end{cases}$$

其中,  $U(G_p)$ 表示 0 到 1 或 1 到 0 跳变的次数, LBP<sup>riu2</sup> 被称为旋转不变的等价模式。

等价模式和旋转不变改进后, LBP 算子不仅具有旋转不变性和灰度不变性的显著特点, 原始 LBP 模式种类也大幅度减少。表中显示了不同 LBP 算子的模式维数。可以看出, 经过等价模式和旋转不变等价模式的改进, 原始 LBP 模式的种类得到大幅度降低。

几种 LBP 算子的维数比较

	原始模式数	等价模式	旋转不变等价模式
$LBP_8^1$	256	58(+1)	9
$LBP_{16}^2$	65536	242(+1)	17
$LBP_{24}^3$	16777216	554(+1)	25



## 基于浅层模型的方法

### 1. 特征设计与提取方法

#### ◆ LBP用于检测的原理

对图像采用LBP算子提取特征后得到的仍然是一幅图像，经过LBP运算后，所改变的只是各个像素点的取值，我们将经过LBP运算后的图像称为LBP图谱。

在LBP的应用中，如纹理分类、人脸分析等，一般都不将LBP图谱作为特征向量用于分类识别，而是采用LBP特征谱的统计直方图作为特征向量用于分类识别。常用非参数统计的方法比较两个样本间LBP直方图分布的相似性，采用这种方法的优点在于可以避免对特征分布做假设。

#### 对LBP特征向量进行提取的步骤

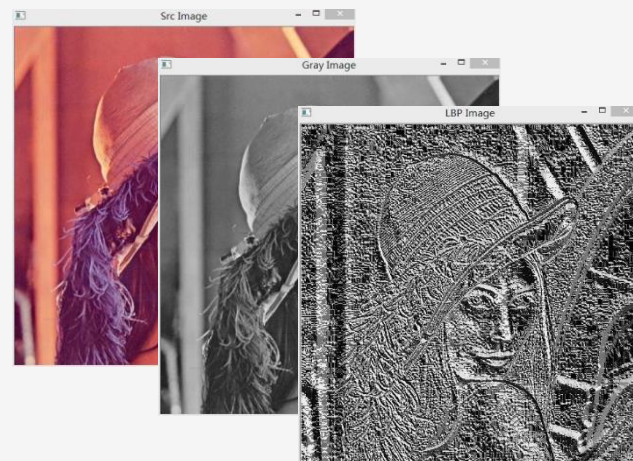
- ✓ 首先将检测窗口划分为 $16 \times 16$ 的小区域（cell）；
- ✓ 对于每个cell中的一个像素，将相邻8个像素的灰度值与其比较，若周围像素值大于中心像素值，则该像素点位置被标记为1，否则为0。这样， $3 \times 3$ 邻域内8个点经比较可产生8位二进制数，即得到该窗口中心像素点LBP值；
- ✓ 然后计算每个cell的直方图，即每个数字（假定是十进制数LBP值）出现的频率；然后对该直方图进行归一化处理。
- ✓ 最后将得到的每个cell的统计直方图进行连接成为一个特征向量，也就是整幅图的LBP纹理特征向量；
- ✓ 最后便可利用SVM或者其他机器学习算法进行分类了



(a) 原始图像



(b) LBP图像



LBP算子具有旋转不变和灰度不变等优点，还支持多尺度分析，因此，LBP算子在模式识别和纹理分析等方面得到了极大的推广和发展。LBP已经被应用在工业检测、图像恢复、人脸分析和场景分析等方面。



## 基于浅层模型的方法

### 2. 特征汇聚与特征变换

前序步骤提取的人工设计特征往往非常多，给后续计算带来困难。更重要的是，这些特征在设计之初并未充分考虑随后的任务或目标。例如，用于分类时未必具有非常好的判别能力，即区分不同目标的能力。因此，在进行图像分类、检索或识别等任务时，在将它们输入到分类器或回归器之前，一般还需要对这些特征进行进一步处理——特征汇聚与特征变换，以便把高维特征进一步编码到某个维度更低或者具有更好判别能力的新空间。

实现上述目的的方法有两大类：

一类是特征汇聚方法。典型的方法包括视觉词袋模型、Fisher 向量和局部聚合向量（VLAD）方法。其中，词袋模型（bag-of-words, BOW）最早出现在自然语言处理（NLP）和信息检索（IR）领域。该模型忽略掉文本的语法和语序，用一组无序的单词（words）来表达一段文字或一个文档。受此启发，研究人员将词袋模型扩展到计算机视觉中，并称之为视觉词袋模型（bag-of-visual-words, BOVW）。简而言之，图像可以被看作文档，而图像中的局部视觉特征（visual feature）可以看作是单词（word）的实例，从而可以直接应用BOW方法实现大规模图像检索等任务。





## 基于浅层模型的方法

### 2. 特征汇聚与特征变换

另一类是特征变换方法，又称子空间分析法。这类方法特别多，典型的方法包括主成分分析（PCA）、线性判别分析、核方法、流形学习等。

- PCA一种在最小均方误差意义下最优的线性变换降维方法，在计算机视觉中应用极为广泛。例如：1990年发表的人脸识别领域最具里程碑式的工作Eigenface本质上就是PCA，其后二十余年，PCA都是人脸识别系统中几乎不可或缺模块。**PCA在寻求降维变换时的目标函数是重构误差最小化，与样本所属类别无关，因而是一种无监督的降维方法。**
- 但在众多计算机视觉应用中，分类才是最重要的目标，因此以最大化类别可分性为优化目标寻求特征变换成为一种最自然的选择，这其中最著名的是**费舍尔线性判别分析方法（FLDA）**，它也是种非常简单而优美的线性变换方法，其基本思想是寻求一个线性变换，使得变换后的空间中同一类别的样本散度尽可能小，而不同类别样本的散度尽可能大，即所谓“**类内散度小，类间散度大**”。
- 核方法曾经是实现非线性变换的重要手段之一。核方法并不试图直接构造或学习非线性映射函数本身，而是在原始特征空间内通过核函数（kernel function）来定义目标“高维隐特征空间”中的内积。换句话说，**核函数实现了一种隐式的非线性映射，将原始特征映射到新的高维空间**，从而可以在无须显式得到映射函数和目标空间的情况下，计算该空间内模式向量的距离或相似度，完成模式分类或回归任务。前述的PCA和FLDA均可以Kernel化，以实现“非线性”的特征提取。





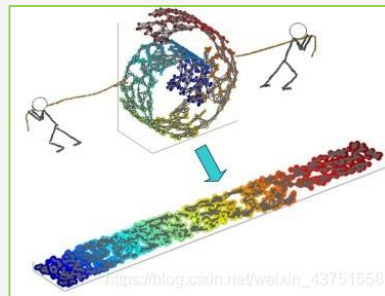
## 基于浅层模型的方法

### 2. 特征汇聚与特征变换

另一类是特征变换方法，又称子空间分析法。这类方法特别多，典型的方法包括主成分分析（PCA）、线性判别分析、核方法、流形学习等。

- 实现非线性映射的另外一类方法是流形学习（manifold learning），其主要思想是寻求将高维的数据映射到低维本征空间的低维嵌入，要求该低维空间中的数据能够保持原高维数据的某些本质结构特征。根据要保持的结构特征的不同，2000年之后出现了很多流形学习方法，其中最著名的是2000年发表在同一期Science杂志上的等距映射ISOMAP和局部线性嵌入（locally linear embedding, LLE）。ISOMAP保持的是测地距离，其基本策略是首先通过最短路径方法计算数据点之间的测地距离，然后通过MDS得到满足数据点之间测地距离的低维空间。而LLE方法则假设每个数据点可以由其近邻点重构，通过优化方法寻求一个低维嵌入，使所有数据仍能保持原空间邻域关系和重构系数。略显不足的是，多数流形学习方法都不易得到一个显式的非线性映射，因而往往难以将没有出现在训练集中的样本变换到低维空间，只能采取一些近似策略，但效果并不理想。

所以直观上来讲，一个流形好比是一个 $d$ 维的空间，在一个 $m$ 维空间中（ $m > d$ ）被扭曲之后的结果。需要注意的是流形并不是一个形状，而是一个空间。举个例子来说（如图所示），比如说一块布，可以把它看成一个二维的平面，这是一个二维的空间，现在我们把它扭一扭（三维空间），它就变成了一个流形，当然不扭的时候，它也是一个流形，欧式空间是流形的一种特殊情况。

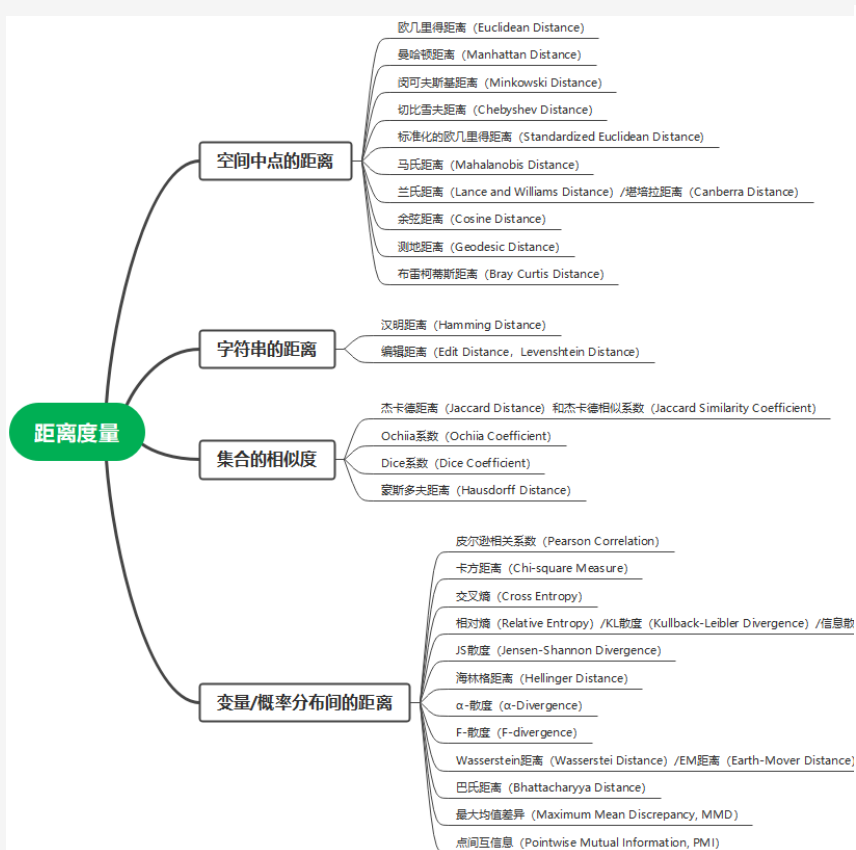




## 基于浅层模型的方法

### 3. 分类器或回归器设计

前面介绍了面向浅层模型的人工设计特征及其对它们进一步汇聚或变换的方法。一旦得到这些特征，剩下的步骤就是分类器或回归函数的设计和学习了。



计算机视觉中的分类器基本都借鉴模式识别或机器学习领域，如最近邻分类器线性感知机、决策树、随机森林、支持向量机、AdaBoost、神经网络等都是适用的。

需要特别注意的是，根据前述特征的属性不同，分类器或回归器中涉及的距离度量方法也有所差异。例如，对于直方图类特征，一些面向分布的距离如直方图交、KLD、卡方距离等可能更实用；对PCA、FLDA 变换后的特征，欧氏距离或Cosine相似度可能更佳；对一此二值化的特征，海明距离可能带来更优的性能。



## 基于深度模型的方法

点燃深度学习在计算机视觉领域应用热潮的爆点发生在2012年。

这一年，Hinton教授研究组设计了**深度卷积神经网络（DCNN）模型AlexNet**，利用ImageNet提供的大规模训练数据并采用两块GPU卡进行训练，将ImageNet大规模视觉识别竞赛（ILSVRC）之“图像分类”任务的**Top5错误率降低到了15.3%**，而传统方法的错误率高达26.2%（且仅比2011年降低了2个百分点）。

这一结果让研究者看到了深度学习的巨大威力，以致2013年当这个竞赛再次举行时，成绩靠前的队伍几乎全部采用了深度学习方法，其中图像分类任务的冠军来自纽约大学Fergus 研究组，他们将**Top5错误率降到了11.7%**，所采用的模型亦是进一步优化的深度 CNN。

2014年，在同一竞赛中，Google依靠一个加深为22层的深度卷积网络GoogLeNet将**Top5错误率降低到了6.6%**。到2015年，微软亚洲研究院的何凯明等人设计了一个深达152层的ResNet模型，并将这一错误率刷新到了3.6%。

四年内，ImageNet图像分类任务的**Top5错误率从26.2%降到3.6%**，几乎每年错误率都下降50%，这显然是一次跨越式的进步。



## 基于深度模型的方法

实际上，**深度学习是多层神经网络的复兴而非革命。**

20世纪90年代之后，神经网络研究陷入了低潮期，但实际上神经网络的研究并未完全中断。LeCun 等在1989年提出了卷积神经网络（CNN），并在此基础上于1998年设计了LeNet-5卷积神经网络，通过大量数据的训练，该模型成功应用于美国邮政手写数字识别系统中。

引爆深度学习在计算机视觉领域应用热潮的**AlexNet即是LeNet-5网络的扩展和改进**，而后来的GoogleNet、VGG、ResNet、DenseNet等深度模型在基本结构上都是CNN，只是在网络层数、卷积层结构、非线性激活函数、连接方式、Loss函数、优化方法等方面有了新的发展。其中，ResNet 通过跨层跳连（shortcut结构），使得优化非常深的模型成为可能。

事实上，深度学习中的深度卷积神经网络（DCNN）也是通过滤波器提取局部特征，然后通过逐层卷积和汇聚，逐渐将“小局部”特征扩大为“越来越大的局部”特征，甚至最终通过全连接形成“全局特征”。但**与浅层模型相比，深度模型的滤波器参数（权重）不是人为设定的，而是通过神经网络的BP算法等训练学习而来**，而且DCNN模型以统一的卷积作为手段，实现了从小局部到大局部（即所谓层级感受野）特征的提取。



## 基于深度模型的方法

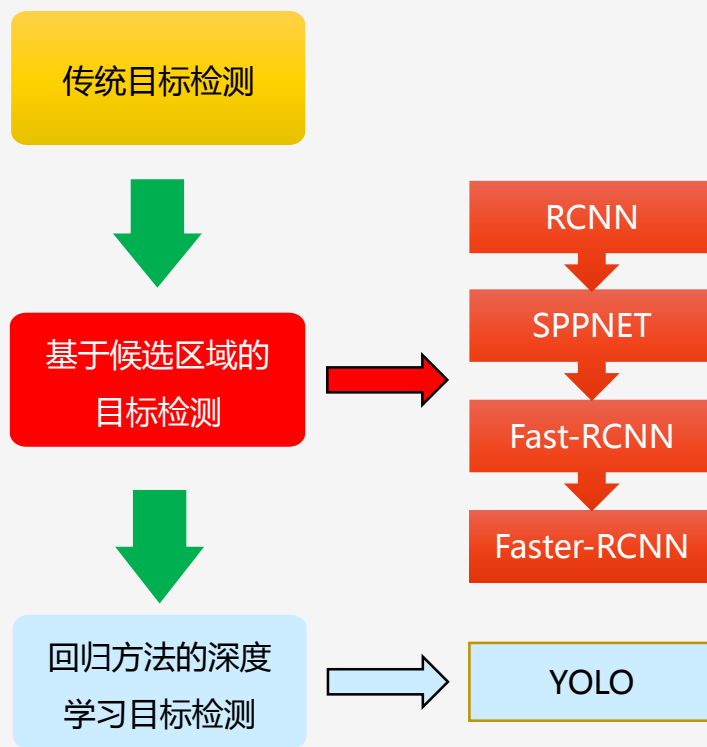
### 1. 基于深度模型的目标检测技术

目标检测是计算机视觉中一个基础问题，其定义某感兴趣的特定类别组成前景，其他类别为背景。需要设计一个目标检测器，它可以在输入图像中找到所有前景物体的位置以及它们所属的具体类别。物体的位置用长方形物体边框描述。

实际上，目标检测问题可以简化为图像区域的分类问题，如果在一张图像中提取足够多可能的物体候选位置，那么只需要将所有候选位置进行分类，即可找到含有物体的位置。

在实际操作中，常常再引入一个边框回归器用来修正候选框的位置，并在检测器后接入一个后处理操作去除属于同一物体的重复检测框。

自从深度学习引入目标检测问题后，目标检测正确率大大提升。





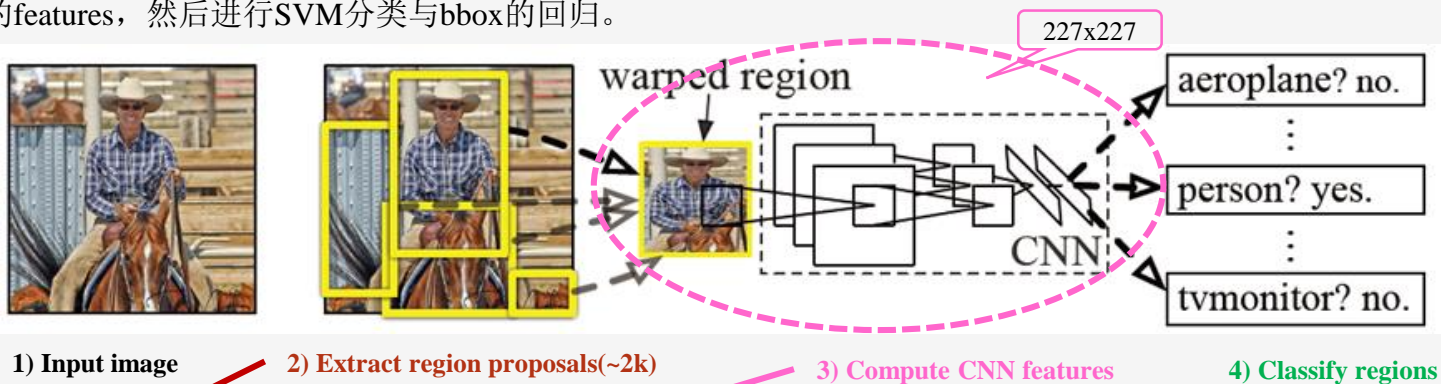


## 基于深度模型的方法

### 1. 基于深度模型的目标检测技术

#### ◆ R-CNN

R-CNN全称region with CNN features, 其实它的名字就是一个很好的解释。用CNN提取出Region Proposals中的features, 然后进行SVM分类与bbox的回归。



通过Selective Search从原始图片提取2000个左右区域候选框;

每个候选区域被resize成固定大小( $227 \times 227$ ) Warp, 并送入CNN中, 通过CNN网络提取特征(在特征层基础上添加两个全连接层)

特征向量送入SVM分类器, 预测出候选区域中所含物体的属于每个类的概率值。每个类别训练一个SVM分类器, 从特征向量中推断其属于该类别的概率大小。

#### R-CNN的缺点:

- **重复计算:** R-CNN虽然不再是穷举, 但依然有两千个左右的候选框, 这些候选框都需要进行CNN操作, 计算量依然很大, 其中有不少其实是重复计算;
- **SVM模型:** 而且还是线性模型, 在标注数据不缺的时候显然不是最好的选择;
- **训练测试分为多步:** 区域提名、特征提取、分类、回归都是断开的训练的过程, 中间数据还需单独保存; 训练的空间和时间代价很高。

使用SVM结合NMS(非极大值抑制)获得区域边框(BBox), 最后用DPM中类似的线性回归方法精修(Refine)边框位置。为了提升定位准确性, R-CNN最后又训练了一个边界框回归模型, 通过边框回归模型对框的准确位置进行修正。



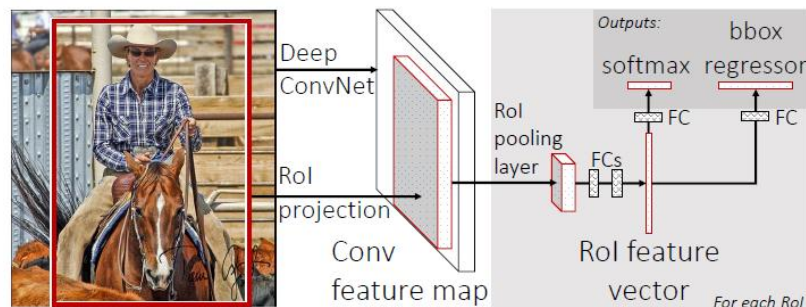


## 基于深度模型的方法

### 1. 基于深度模型的目标检测技术

#### ◆ Fast R-CNN

Fast R-CNN是在R-CNN基础上采纳了SPP Net (Spatial Pyramid Pooling, 空间金字塔池化) 方法, 使性能进一步提高。



Fast R-CNN对同一张图像只提取一次卷积特征, 此后接入ROI pooling层, 将特征图上不同尺寸的感兴趣区域取出并池化为固定尺寸的特征, 再将这些特征用Softmax进行分类。

此外, Fast RCNN还利用多任务学习, 将ROI pooling层后的特征输入一个边界框的回归器来学习更准确的位置。后来, 为了降低提取候选位置所消耗的运算时间, Faster R-CNN进一步简化流程, 在特征提取器后设计了RPN (Region Proposal Network), 用于修正和筛选预定义在固定位置的候选框, 将上述所有步骤集成于一个整体框架中, 从而进一步加快了目标检测速度。

Fast R-CNN与R-CNN相比, 主要有两点不同:

(1) **最后一个卷积层加入了ROI pooling layer**。ROI pooling layer实际上是SPP Net的一个精简版, SPP Net对每个proposal使用了不同大小的金字塔映射, 而ROI pooling layer只需要采样到一个7x7的特征图。对于VGG16网络conv5\_3有512个特征图, 这样所有region proposal对应了一个7\*7\*512维度的特征向量作为全连接层的输入。也就是说, 它可以把不同大小的输入映射到一个固定尺度的特征向量。

(2) **损失函数使用了多任务损失函数 (multi-task loss)**, 将边框回归Bounding Box Regression直接加入到CNN网络中训练。R-CNN训练过程分为了三个阶段, 而Fast R-CNN直接使用了softmax替代SVM分类, 同时利用多任务损失函数边框回归也加入到网络中, 整个训练过程就是端到端的 (region proposal提取除外)

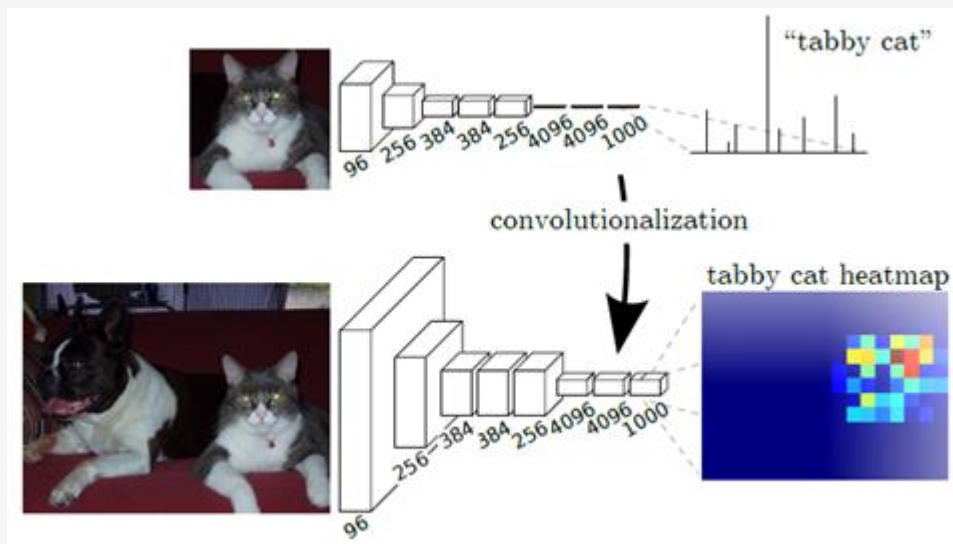
也就是说, R-CNN的处理流程是先提proposal, 然后CNN提取特征, 之后用SVM分类器, 最后再做box regression。而在Fast R-CNN中, 把box regression与region分类合并成multi-task模型, 实际实验也证明, 这两个任务能够共享卷积特征, 并相互促进。



## 基于深度模型的方法

### 2. 基于全卷积网络的图像分割

对于像素级的分类和回归任务（如图像分割或边缘检测），代表性的深度网络模型是全卷积网络（Fully Convolutional Network, FCN）。经典的DCNN在卷积层之后使用了全连接层，而全连接层中单个神经元的感受也是整张输入图像，破坏了神经元之间的空间关系，因此不适用于像素级的视觉处理任务。为此，**FCN去掉了全连接层，代之以 $1 \times 1$ 的卷积核和反卷积层**，从而能够在保持神经元空间关系的前提下，通过反卷积操作获得与输入图像大小相同的输出。进一步，FCN通过不同层、多尺度卷积特征图的融合为像素级的分类和回归任务提供了一个高效的框架。



FCN将这3层表示为卷积层，滤波器的大小（通道数，宽，高）分别为 $(4096, 7, 7)$ 、 $(4096, 1, 1)$ 、 $(1000, 1, 1)$ 。

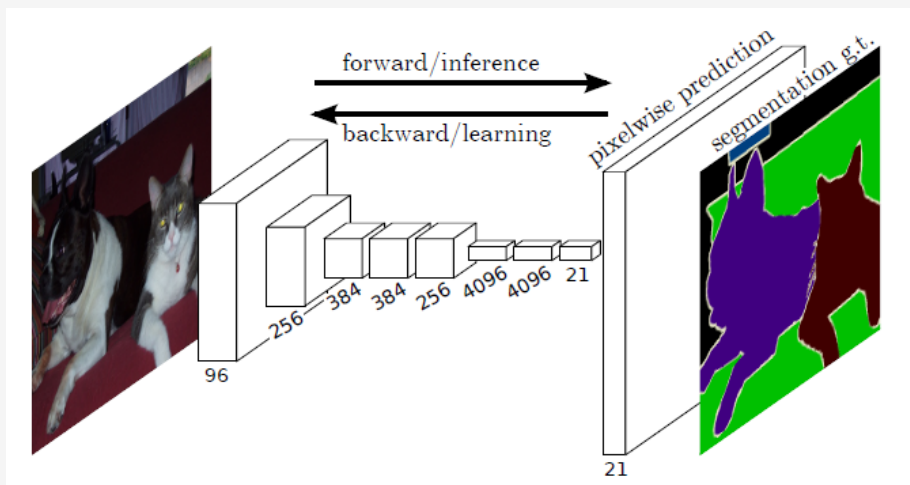
看上去数字上并没有什么差别，但是卷积跟全连接是不一样的概念和计算过程，使用的是之前CNN已经训练好的权值和偏置，但是不一样的在于权值和偏置是有自己的范围，属于自己的一个卷积核。



## 基于深度模型的方法

### 2. 基于全卷积网络的图像分割

经过多次卷积和池化以后，得到的图像越来越小，分辨率越来越低。产生图叫做heatmap热图，热图即为高维特征图，然后进行上采样，把图像进行放大、放大、放大，到原图像的大小，确保识别出来的类别能和原来每一个像素点的位置对上。



- 为了对每个像素进行分类预测成最后已经进行语义分割的图像，就对每个像素点分别在21个通道中进行softmax得到属于各类的概率，然后比较这21个通道中的概率哪个最大就选哪个类作为该像素的分类，因此产生了一张已经分类好的图。
- 直接利用ground truth作为监督信息训练一个端到端的网络。
- 最后逐个像素计算softmax分类的损失。

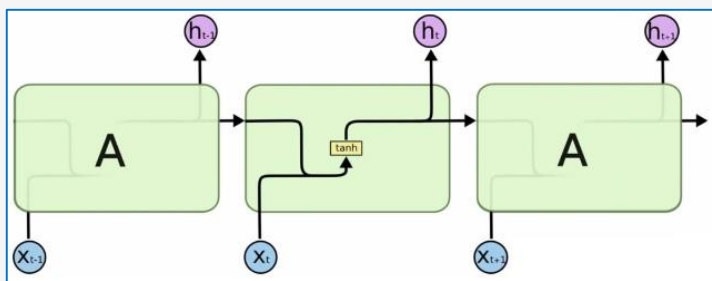


## 基于深度模型的方法

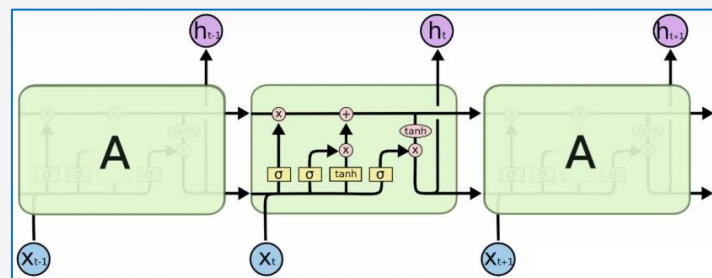
### 3. 融合图像和语言模型的自动图题生成

图像自动标题（Image Captioning）的目标是生成输入图像的文字描述，即我们常说的“看图说话”，也是一个因深度学习取得重要进展的研究方向。

深度学习方法应用于该问题的代表性思路是使用CNN学习图像表示，然后采用循环神经网络RNN或长短期记忆模型LSTM学习语言模型，并以CNN特征输入初始化RNN/LSRM的隐层节点，组成混合网络进行端到端的训练。通过这种方法，有些系统在MS COCO数据集上的部分结果甚至优于人类给出的语言描述。



RNN



LSTM

- LSTM模型是为了解决RNN神经网络中的梯度爆炸问题。
- RNN是想把所有信息都记住，不管是有用的信息还是没用的信息。LSTM设计一个记忆细胞（具备选择性记忆的功能），可以选择记忆重要信息，过滤掉噪声信息，减轻记忆负担。
- LSTM内部更加复杂，且RNN只有一条链路串联起各个神经元，而LSTM是由两条链路串联起来的。





# 计算机视觉关键技术

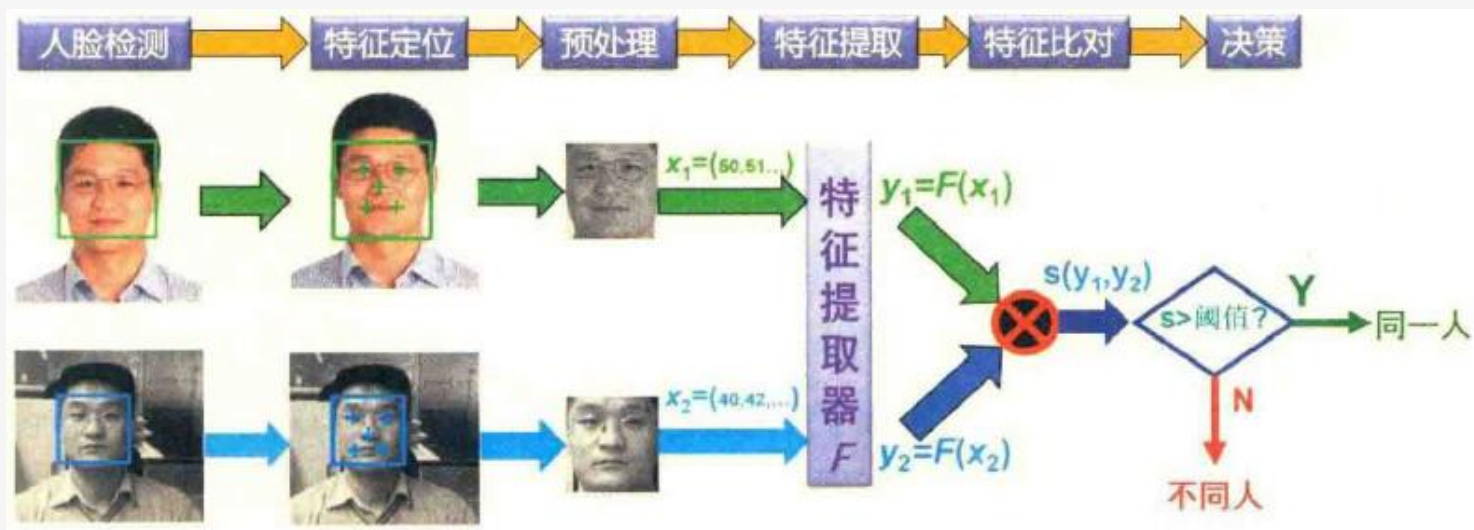


西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 应用实例：人脸识别技术

人脸识别是计算机视觉领域的典型研究课题，不仅可以作为计算机视觉、模式识别、机器学习等学科领域理论和方法的验证案例，还在金融、交通、公共安全等行业有非常广泛的应用价值。特别是近年来，人脸识别技术逐渐成熟，基于人脸识别的身份认证、门禁、考勤等系统开始大量部署。

如图所示，**人脸识别的本质是对两张照片中人脸相似度进行计算**。为了计算该相似度，一套典型的人脸识别系统包括6个步骤：人脸检测、特征点定位、面部子图预处理、特征提取、特征比对和决策。

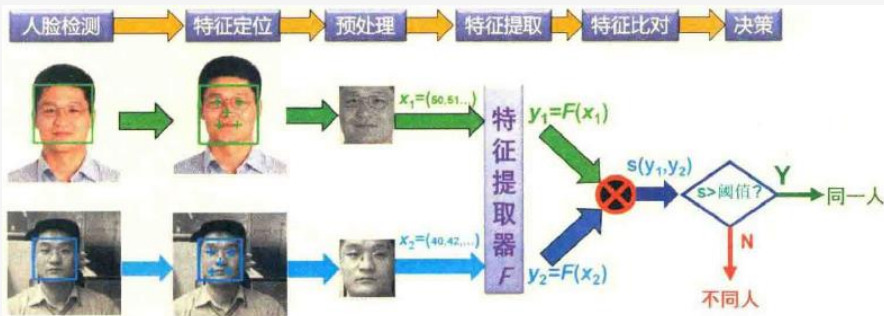


人脸识别典型流程



## 应用实例：人脸识别技术

- **步骤 1：人脸检测**，即从输入图像中判断是否有人脸，如果有的话，**给出人脸的位置和大小（图中的矩形框）**。作为一类特殊目标，人脸检测用前述的基于深度学习的通用目标检测技术实现。
- **步骤 2：特征点定位**，即在人脸检测给出的矩形框内进一步找到**眼睛中心、鼻尖和嘴角等关键特征点**，以便进行后续的预处理操作。理论上，也可以采用通用的目标检测技术实现对眼睛、鼻子和嘴巴等目标的检测。此外，也可以采用回归方法，直接用深度学习方法实现从检测到的人脸子图到这些关键特征点坐标位置的回归。
- **步骤 3：面部子图预处理**，即实现对面脸子图的归一化，主要包括两部分：①把**关键点进行对齐**，即把所有人脸关键点放到差不多接近的位置，以消除人脸大小、旋转等影响；②对人脸**核心区域子图进行光亮度方面的处理**，以消除光强弱、偏光等影响。该步骤的处理结果是一个标准大小（比如 100x100 像素大小）的人脸核心区子图像。
- **步骤 4：特征提取**，是人脸识别的核心，其功能是从步骤3输出的人脸子图中**提取可以区分不同人的特征**。在采用深度学习之前，典型方法是采用前述的“特征设计与提取”及“特征汇聚与特征变换”两个步骤来实现。例如，采用LBP 特征，最终可以形成由若干区域局部二值模式直方图串接而成的特征。







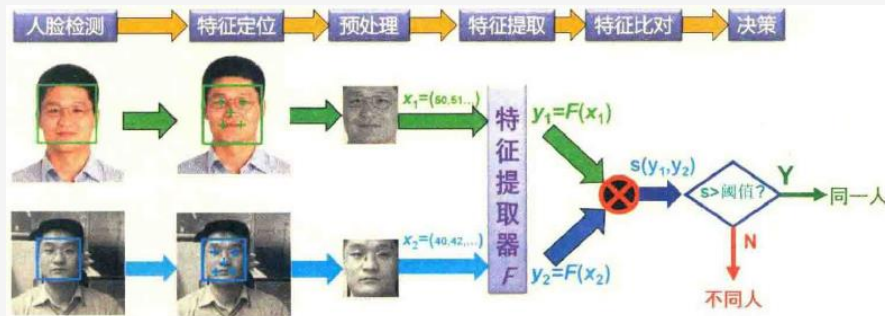
# 计算机视觉关键技术



西安交通大学  
XI'AN JIAOTONG UNIVERSITY

## 应用实例：人脸识别技术

- **步骤 5：特征比对**，即对两幅图像所提取的特征进行距离或相似度的计算，如欧氏距离、cosine 相似度等。如果采用的是 LBP 直方图特征，则直方图交是常用的相似度度量。

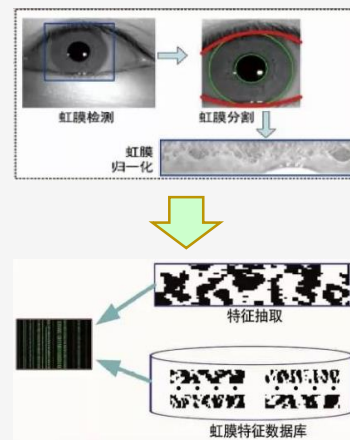


例如，设人脸图像 A 得到的 LBP 直方图特征为： $H^A = (h_{1,1}^A, h_{1,2}^A, \dots, h_{1,p}^A; h_{2,1}^A, h_{2,2}^A, \dots, h_{2,p}^A; \dots; h_{m,1}^A, h_{m,2}^A, \dots, h_{m,p}^A)$   
设人脸图像B得到的 LBP 直方图特征为： $H^B = (h_{1,1}^B, h_{1,2}^B, \dots, h_{1,p}^B; h_{2,1}^B, h_{2,2}^B, \dots, h_{2,p}^B; \dots; h_{m,1}^B, h_{m,2}^B, \dots, h_{m,p}^B)$   
则它们之间的直方图交计算为： $S(H^A, H^B) = \sum_{i=1}^m \sum_{j=1}^p \min(h_{i,j}^A, h_{i,j}^B)$

- **步骤 6：决策**，即对前述相似度或距离进行阈值化。最简单的做法是采用阈值法，相似程度超过设定阈值则判断为相同人，否则为不同人。上述例子中给出的是1:1的判断，实际应用中人脸识别还可能是一张照片和注册数据库中N个人的照片的比对，此时只需要对 N个相似度进行排序，相似度最大且超过设定阈值者即为输出的识别结果。

### 相似度扩展：虹膜识别等其他生物特征识别技术

人脸识别作为生物特征识别技术已经得到广泛应用，而基于计算机视觉的生物特征识别技术还有很多，如指纹识别、虹膜识别、掌纹识别、指静脉识别等。其中指纹识别是大家最熟悉，也是相对最成熟的。虹膜识别是利用眼睛虹膜区域的随机纹理特性来区分不同人的技术。虹膜是人眼瞳孔和巩膜之间的环状区域，在较高分辨率、用户高度配合等良好采集条件下，可以采集到纹理丰富细腻的辐射状虹膜细节，虹膜采集往往采用主动近红外采集。虹膜识别的典型过程与人脸识别类似，也需要检测并分割出环状虹膜区域并预处理，然后进行特征提取和比对等步骤。





西安交通大学  
XI'AN JIAOTONG UNIVERSITY

# 谢谢大家

