



# 目录

## CONTENTS

9.1

专家系统概述

9.2

推理方法

9.3

一个简单的专家系统

9.4

非确定性推理

9.5

专家系统工具

9.6

专家系统的应用

9.7

专家系统的局限性





中國石油大學(华东)  
CHINA UNIVERSITY OF PETROLEUM

计算机科学与技术学院  
College of Computer Science & Technology

# 第十章 计算机视觉

人工智能课题组

智能科学系





# 目录

## CONTENTS

10.1

计算机视觉概述

10.2

数字图像的类型及表示

10.3

常用计算机视觉模型和关键技术

10.4

应用实例：人脸识别技术





## 10.1 计算机视觉概述

- 计算机视觉是一门研究如何对数字图像或视频进行**高层理解**的交叉学科。
- 类比人的视觉系统，摄像机等设备是机器的眼睛，而计算机视觉就是要实现人的大脑的视觉能力。计算机视觉赋予机器“看”的智能。
- 从工程视角来看，所谓理解图像或视频，就是用机器自动实现人类视觉系统的功能，包括**图像或视频的获取、处理、分析和理解**等诸多任务。

## 10.1 计算机视觉概述

假设我们为盲人设计一套导盲系统，盲人过马路时系统摄像机拍摄了如图所示图像，导盲系统需要完成哪些视觉任务？可能至少包括以下任务：

### 1、距离估计

距离估计是指计算输入图像中的每个点距离摄像机的物理距离。

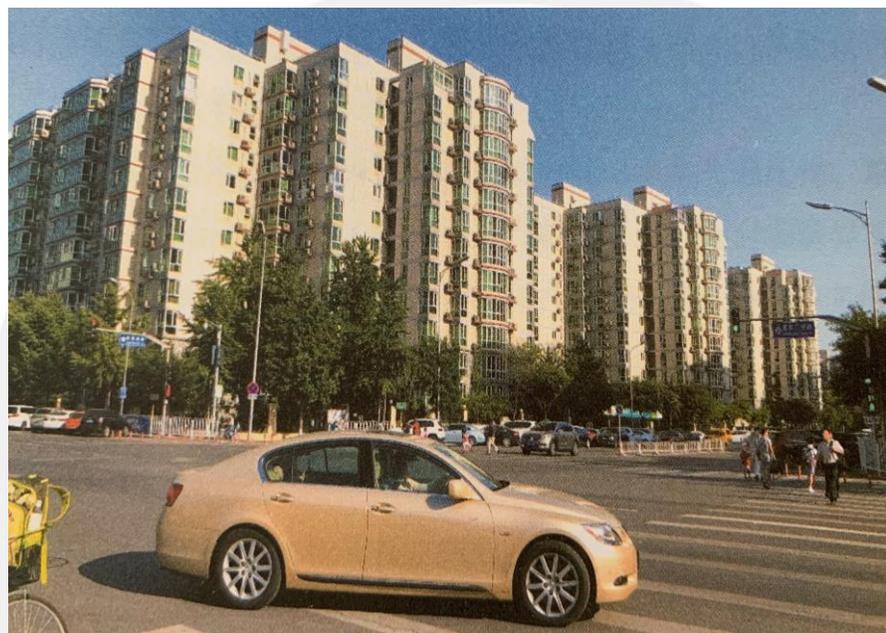
### 2、前背景分割和物体分割

将图像视频中前景物体所占据的区域或轮廓勾勒出来。

例如，视野中的车辆、斑马线、盲道的分割以及可行走区域。

### 3、目标检测

在图像视频中发现感兴趣的目标，并给出其位置和区域。



## 10.1 计算机视觉概述

假设我们为盲人设计一套导盲系统，盲人过马路时系统摄像机拍摄了如图所示图像，导盲系统需要完成哪些视觉任务？可能至少包括以下任务：

### 4、目标分类和识别

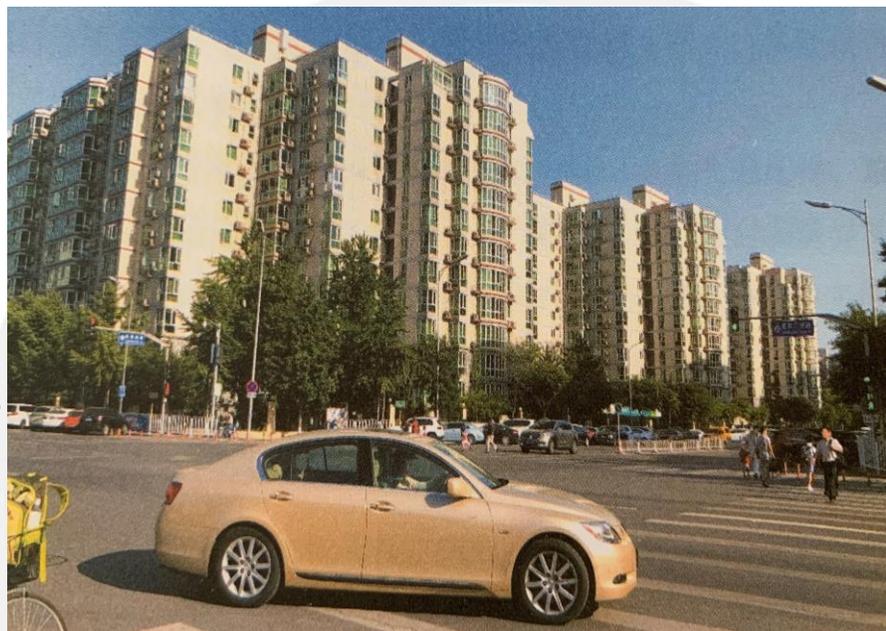
为图像视频中出现的目标，分配其所属类别的标签。

### 5、场景分类与识别

根据图像视频内容对拍摄环境进行分类。如：室内、室外、  
山景、海景、街景等。

### 6、文字检测与识别

识别场景中的各种文字。如：街道名、绿灯倒计时秒数、商店名称等。

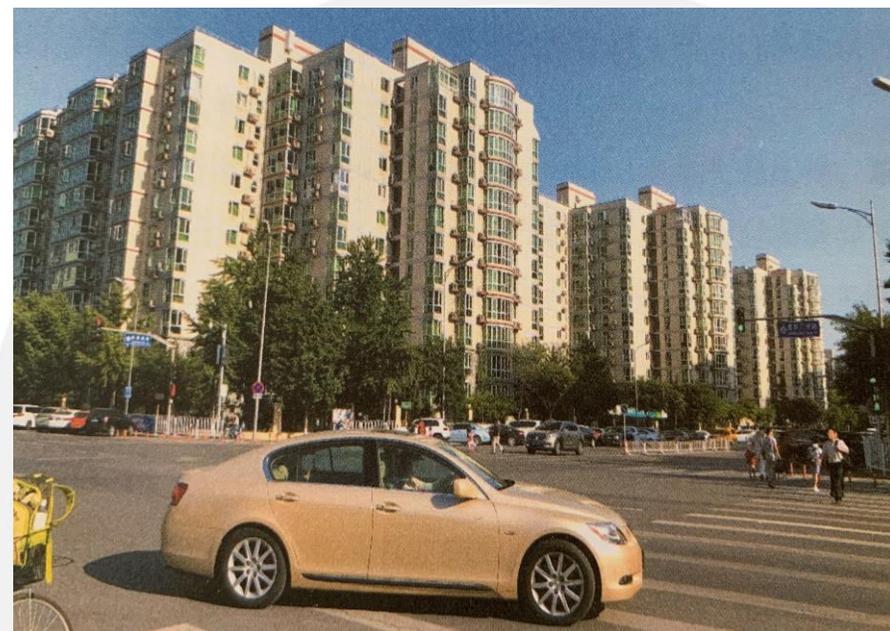


## 10.1 计算机视觉概述

假设我们为盲人设计一套导盲系统，盲人过马路时系统摄像机拍摄了如图所示图像，导盲系统需要完成哪些视觉任务？可能至少包括以下任务：

### 7、事件检测与识别

对视频中的人物和场景进行分析，识别人的行为或正在发生的事件（特别是异常事件）。如：**是否有车辆经过、闯红灯、逆行等。**



## 10.1 计算机视觉概述

8、其他：

8.1、3D重建：对画面中的场景和物体进行自动3D建模。

8.2、图像编辑：内容或风格进行修改，产生具有真实感的其他图像。如：把图像变成油画效果，甚至是变成某个艺术家的绘画风格图，也可以修改图像中的部分内容，如去掉照片中的某个垃圾桶。

8.3、自动标题：分析输入图像或视频的内容，并用自然语言进行描述。

8.4、视觉问答：对视觉图像的自然语言问答。给定图像或视频回答特定的问题。



Q: How many white objects in this picture ?

A: 9



Q: What color is the chair in front of the wall on the left side of the stacked chairs ?

A: blue



Q: What is the largest white object on the left side of the picture ?

A: printer



## 10.1 计算机视觉概述

计算机视觉与很多学科都有密切关系：

- **数字图像处理**：可以看作是偏低级的计算机视觉，多数情况下其输入和输出都是图像，而计算机视觉系统的输出一般是模型、结构或符号信息
- **模式识别**：以图像为输入的模式识别任务多数可以看作是计算机视觉的研究范畴。
- **机器学习**：为计算机视觉提供了分析、识别和理解的方法和工具。
- **计算机图形学**：计算机图形学研究的是如何从模型生成图像或视频的“正”问题；计算机视觉研究的是如何从输入图像中解析出模型的“反”问题。



## 10.2 数字图像的类型及表示

- 数字图像由一个个点组成，这些点称为**像素**。
- 每个像素的亮度、颜色或距离等属性在计算机内表示为一个或多个数字。
- 如果是黑白图像（又称**灰度图像**），每个像素由一个亮度值表示，最小值为0，最大值为255。
- 如果是彩色图像，每个像素的颜色通常用分别代表红绿蓝的三个字节表示。



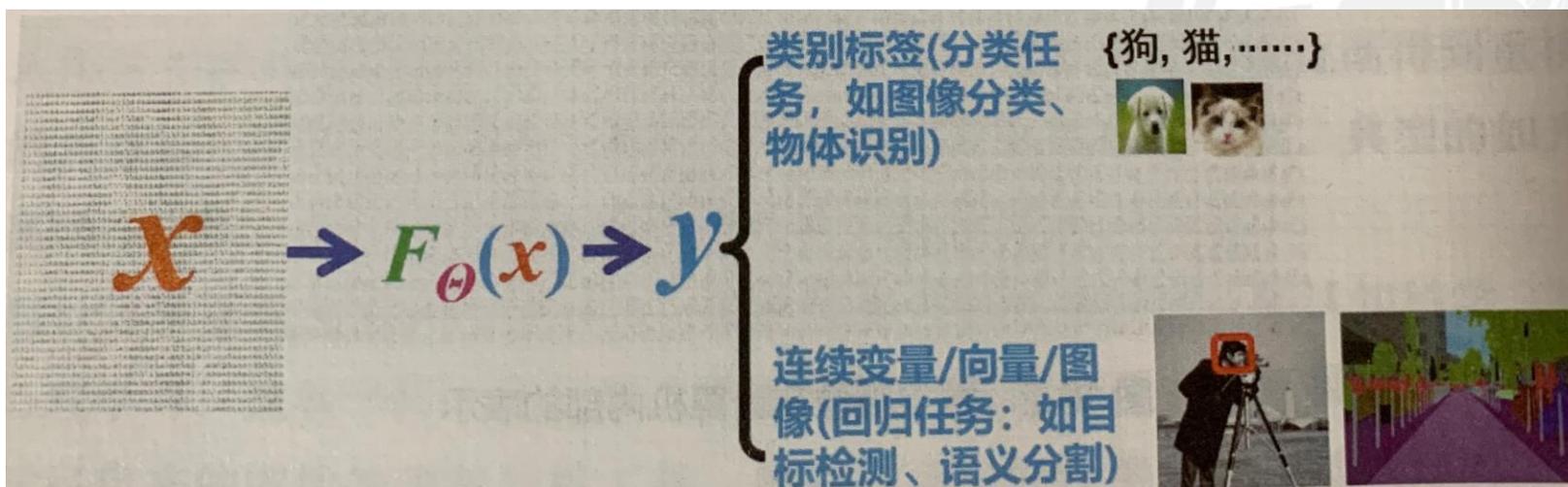
## 10.2 数字图像的类型及表示

- 除了**黑白**和**彩色图像**，还有一类特殊的相机可以采集深度信息，即**RGBD图像**。RGBD图像对每个像素，除了赋予红绿蓝彩色信息之外，还会有一个值表达深度，即该像素距离摄像机的距离。这类相机在**体感游戏、自动驾驶、机器人导航**等领域有广泛的应用价值。
- 此外，计算机视觉处理的图像或视频还可能来自超越人眼所能感知可见光（例如红外、紫外、X光成像等）的成像设备。这些设备及其视觉处理算法在医疗、军事、工业等领域有非常广泛的应用。

## 10.3 常用计算机视觉模型和关键技术

大多数计算机视觉任务可以建模为广义的函数拟合问题，如图所示。对任意输入图像 $x$ ，需要学习一个以 $\theta$ 为参数的函数 $F$ ，使得 $y = F_{\theta}(x)$ ，其中 $y$ 可能有两类：

- 1、 $y$ 为类别标签，属于“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。
- 2、 $y$ 为连续变量或向量或矩阵，属于“回归”问题。如：**距离估计、目标检测、语义分割**等视觉任务。





## 10.3 常用计算机视觉模型和关键技术

多数视觉模型和方法可以被分成两大类：

10.3.1 基于浅层模型的方法

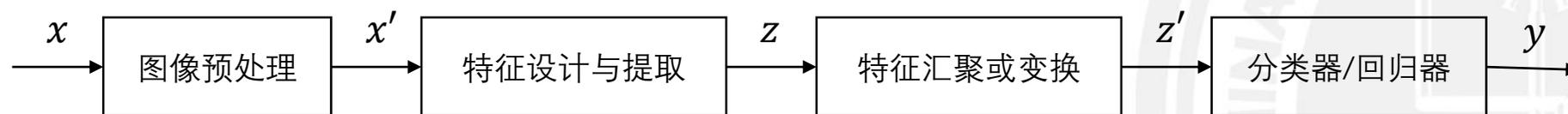
10.3.2 基于深度模型的方法



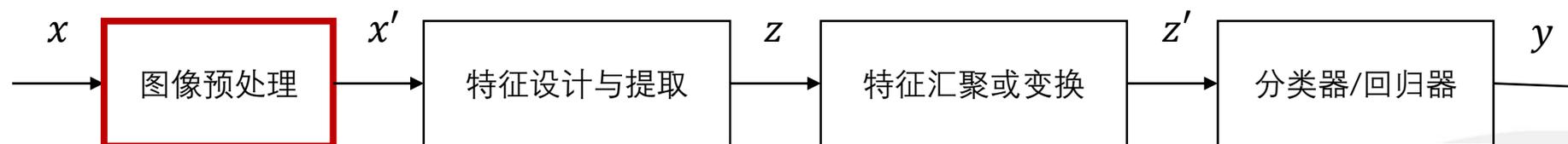


## 10.3.1 基于浅层模型的方法

- 采用“分而治之”的思想，对视觉任务进行求解。
- 一个典型的视觉任务实现流程包括以下四个步骤：
  1. 图像预处理过程 $p$ 。
  2. 特征设计与提取过程 $q$ 。
  3. 特征汇聚或变换 $h$ 。
  4. 分类器或回归器函数 $g$ 的设计与训练。



## 10.3.1 基于浅层模型的方法

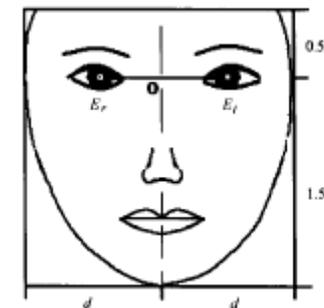


### 1、图像预处理过程p。

用于实现**目标对齐**、**几何归一化**等处理，从而提高数据的一致性。该过程一般人为设定。

- **目标对齐**

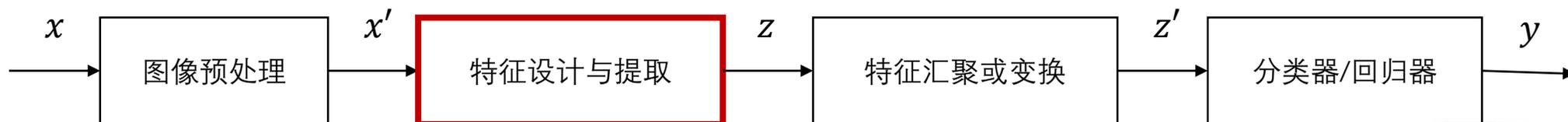
如：人脸对齐任务即根据输入的人脸图像，自动定位出面部关键特征点，如眼睛、鼻尖、嘴角点、眉毛以及人脸各部件轮廓点等。



- **几何归一化**

如：人脸识别任务中，几何归一化将表情图像变换为统一的尺寸，有利于表情特征的提取。

## 10.3.1 基于浅层模型的方法



### 2、特征设计与提取过程q。

从预处理后的图像 $x'$ 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。

- 一般根据专家知识进行人工设计。
- 目前，多数人工设计的特征有两大类
  - ✓ 全局特征
  - ✓ 局部特征



## 10.3.1 基于浅层模型的方法

- **全局特征**：对图像中全部像素或多个不同区域像素中所蕴含的信息建模。
  - 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
    - 傅里叶频谱
    - **GIST特征**：对图像场景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。
  - 特点：全局特征粒度比较粗
  - 适合高效而无须精细分类的任务，比如**场景分类**或**大规模图像检索**等。



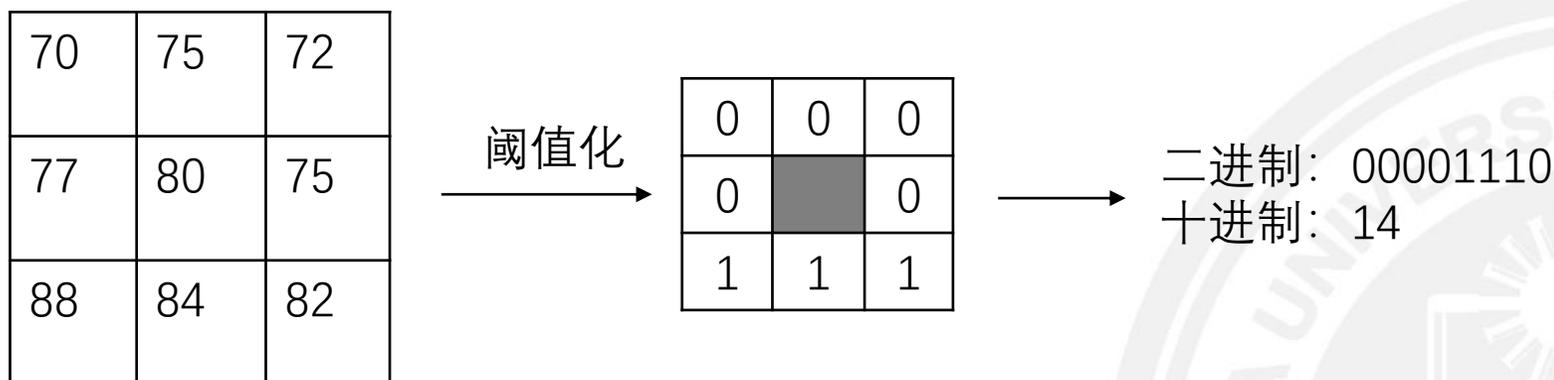
## 10.3.1 基于浅层模型的方法

- **局部特征**：只从一个局部区域内的少量像素中提取信息。
  - 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。
  - 典型的局部特征有：
    - SIFT
    - GIST
    - SURF
    - HOG
    - **LBP (local binary pattern)**
    - Gabor等



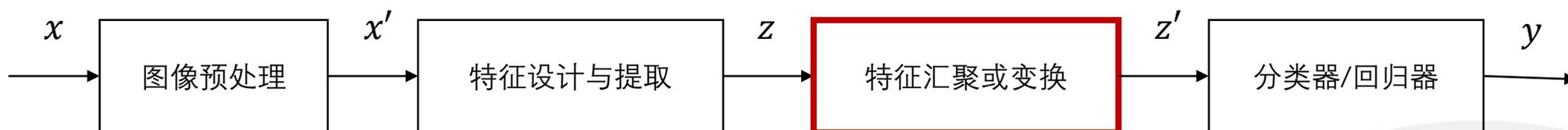
## 10.3.1 基于浅层模型的方法

- 局部二值模式 (**local binary patterns, LBP**) 局部特征
  - LBP 只关注中心像素与其邻域像素的关系，例如：



- 256种模式以不同的概率出现。为了提高鲁棒性、减少模式类，模式分为：
  - **均衡模式**：0/1串中最多包含2次0-1或1-0跳变，如01111000
  - **非均衡模式**：如00110011（3次）
  - 非均衡模式在自然图像中出现的非常少，被强制归为一类模式。  
⇒59种模式。

## 10.3.1 基于浅层模型的方法



### 3、特征汇聚或变换h。

其功能是对前面提取的局部特征进行统计汇聚或降维处理。

- 一般通过专家设计的统计建模方法实现
- 一类是**特征汇聚**方法，典型的方法包括：
  - ✓ 词袋模型
  - ✓ Fisher向量
  - ✓ 局部聚合向量(VLAD)
- 另一类是**特征变换**方法，又称子空间分析法，典型的方法包括：
  - ✓ 主成分分析
  - ✓ 线性判别分析
  - ✓ 核方法
  - ✓ 流形学习

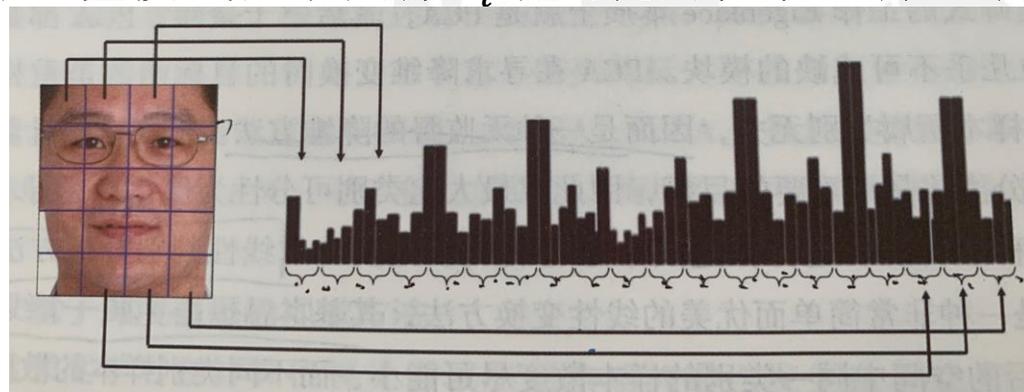
## 10.3.1 基于浅层模型的方法

### • 特征汇聚

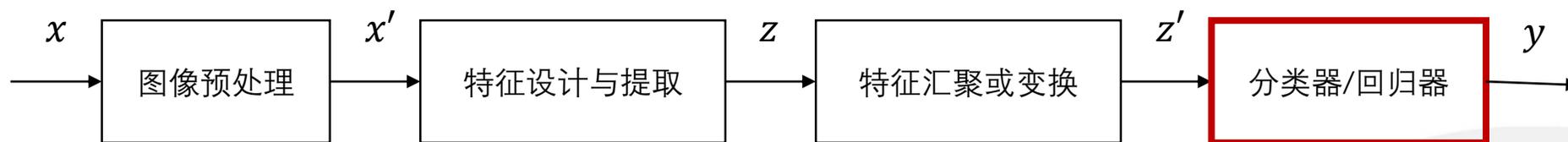
#### • 词袋模型

- 最早出现在自然语言处理（NLP）和信息检索（IR）领域，忽略掉文本的语法和语序，用一组无序的单词来表达一段文字或一个文档。
- 词袋模型扩展到计算机视觉中，称为**视觉词袋模型**。
- 例如：利用LBP提取特征之后，对特征进行直方图统计。

给定一幅人脸图像（设为128\*160像素），首先将其划分为4\*5个子图像，对每个子图像中的每个像素点计算LBP，统计59种模式的频数，即得到一个59维的直方图 $H_i = (h_{i,1}, h_{i,2}, \dots, h_{i,59})$ ，其中， $h_{i,j}$ 表示第 $j$ 种二值模式在子图像 $B_i$ 中出现的次数。将20个直方图串接可得到整个人像的描述特征。



## 10.3.1 基于浅层模型的方法



### 4、分类器/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，基于一个有监督的训练集

$\{x_i, y_i, i = 1, \dots, N\}$ （其中， $x_i$ 是训练图像的特征， $y_i$ 是其类别标签），通过有监督的机器学习方法来实现。例如：

- 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



## 10.3.2 基于深度模型的视觉方法

- 2012年，Hinton教授研究组设计了深度卷积神经网络模型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%，传统的方法错误率高达26.2%。
- 2014年，Google利用22层的GoogLeNet将错误率降到6.6%。
- 2015年，微软亚洲研究院何凯明等人设计了一个152层的ResNet模型，错误率降到3.6%。

1、基于深度模型的目标检测技术

2、基于全卷积网络的图像分割

3、融合图像和语言模型的自动图题生成

## 10.3.2 基于深度模型的视觉方法

### 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的一个基础问题，其定义某些感兴趣的特定类别组成**前景**，其他类别为**背景**。
- 目标检测器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。物体的位置一般用长方形物体边框描述。
- ✓ 目标检测问题可以简化为图像区域的分类问题。
  - 在图像中提取可能物体的候选位置，对所有候选位置进行分类。
  - 实际操作中，需要引入一个边框回归器修正候选框的位置。
  - 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。

## 10.3.2 基于深度模型的视觉方法

### 1、基于深度模型的目标检测技术

- R-CNN

R-CNN最早将深度学习应用在目标检测中，如图所示，步骤包括：

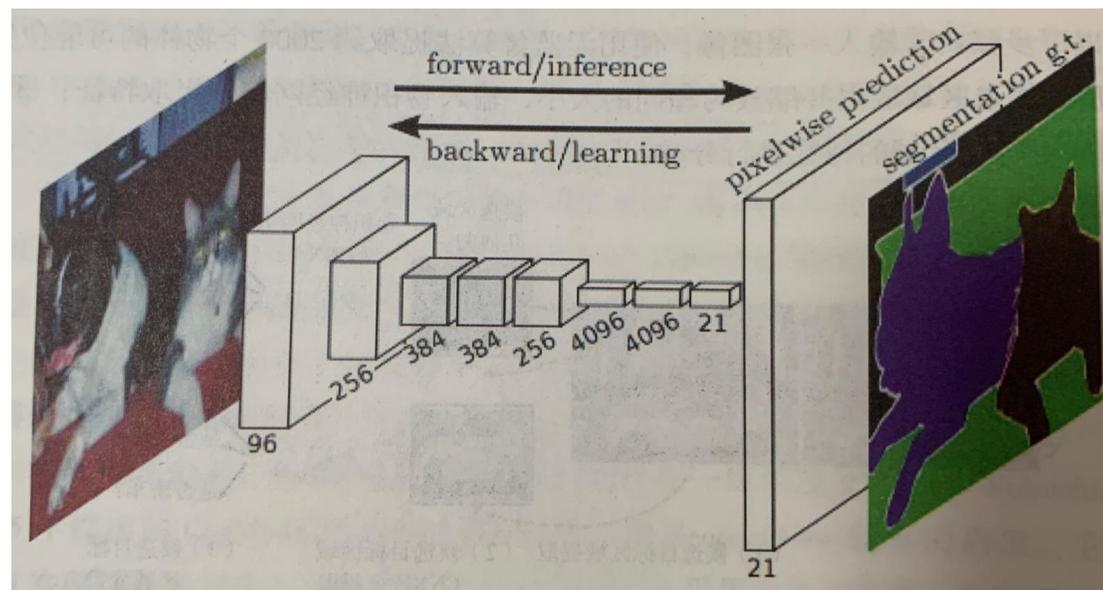


- ① 输入一张图像，使用无监督算法提取约2000个物体的可能位置
- ② 将所有候选区域取出并缩放为相同的大小，输入卷积神经网络中提取特征
- ③ 使用SVM对每个区域的特征进行分类

## 10.3.2 基于深度模型的视觉方法

### 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（FCN）。
- 全卷积神经网络的基本模块：**卷积**。



- 经典的DCNN在卷积层之后使用了全连接层，而全连接层中每个神经元的感受野是整张输入图片，破坏了神经元之间的空间关系，因此不适用于像素级的处理任务。FCN去掉了全连接层，代之以 $1 \times 1$ 的卷积核和反卷积层，从而在保持神经元之间的空间关系的前提下，通过反卷积操作获得与输入图像大小相同的输出。



## 10.3.2 基于深度模型的视觉方法

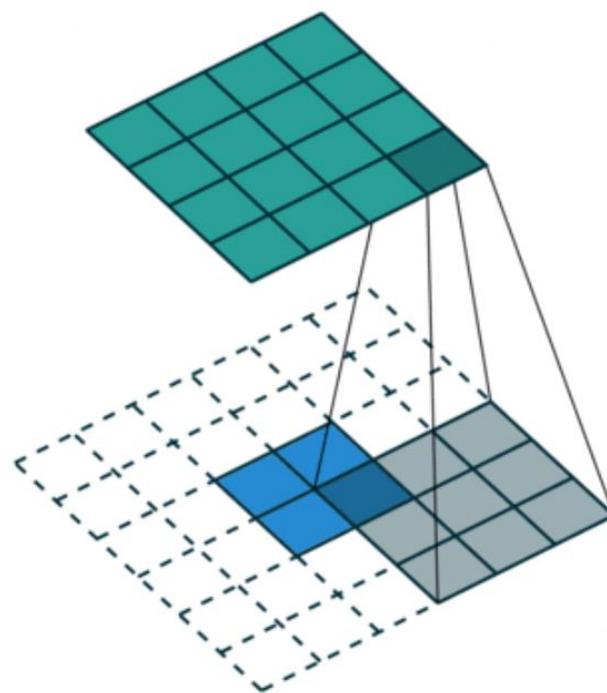
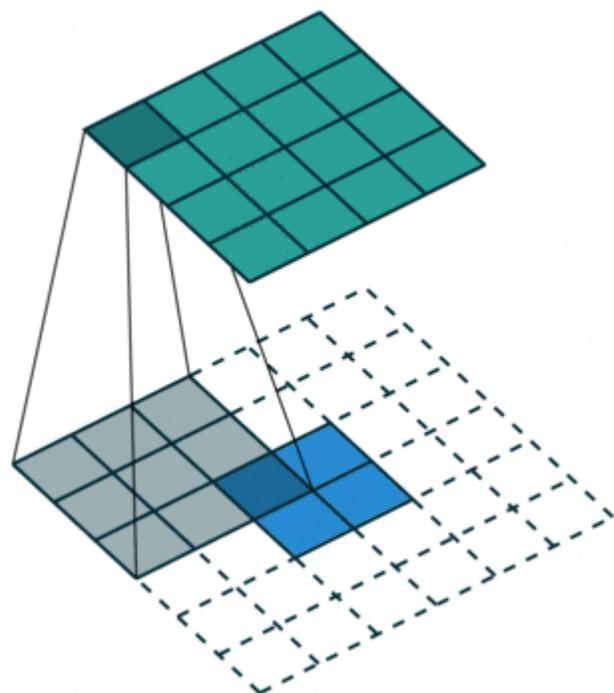
### 2、基于全卷积网络的图像分割

- 多个卷积核：采用一个卷积核进行卷积抽取特征不充分，通常使用多个卷积核
- 多通道卷积：处理彩色图像时，分别使用不同的卷积核对RGB三个通道进行卷积，然后使用线性或非线性激活函数将相同位置的卷积特征合并
- 边界填充：大小为 $m$ 的卷积核在大小为 $n$ 的输入向量上进行操作之后得到的卷积特征向量大小会缩小为 $n - m + 1$ 。卷积层数增加，特征向量减小的速度为 $m - 1$ 。叠加到 $\lfloor \frac{n}{m-1} \rfloor$ 个卷积层之后，特征向量坍塌为标量。为了解决这个问题，通常采用边界填充0的方式。

## 10.3.2 基于深度模型的视觉方法

### 2、基于全卷积网络的图像分割

- 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。

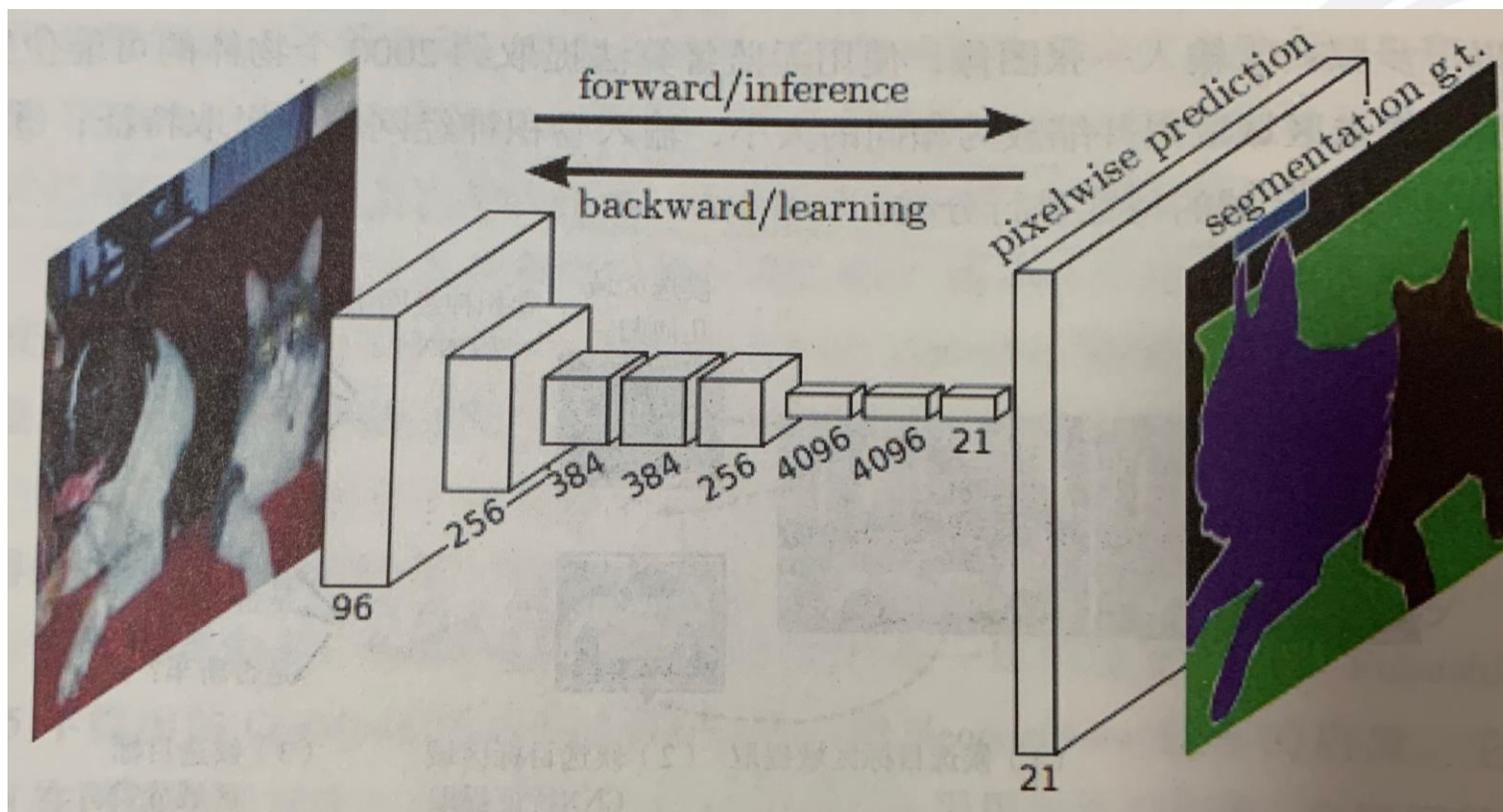


Upsampling Via Deconvolution (Blue: Input, Green: Output)

## 10.3.2 基于深度模型的视觉方法

### 2、基于全卷积网络的图像分割

- 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。





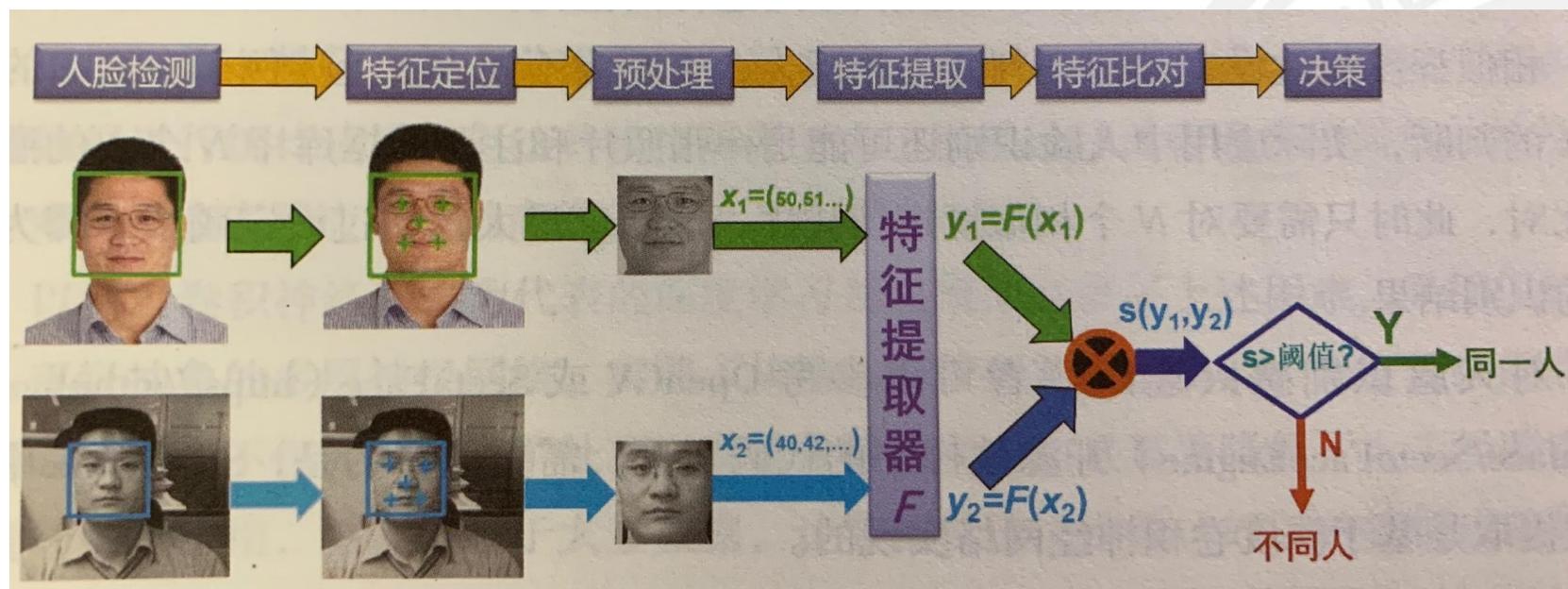
## 10.3.2 基于深度模型的视觉方法

### 3、融合图像和语言模型的自动图题生成

- 图像自动标题的目标是生成输入图像的文字描述，即“看图说话”。
- 深度学习方法应用于该问题的思路是使用CNN学习图像表示，然后采用循环神经网络RNN或长短时记忆模型LSTM学习语言模型，并以CNN特征输入初始化RNN/LSTM的隐层节点，组成混合网络进行端到端的训练。

## 10.4 应用：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。在金融、交通、公共安全等行业有广泛应用。
- 人脸识别的本质是对两张照片中人脸的相似度进行计算。下图给出了人脸识别系统的典型流程。





## 10.4 应用：人脸识别技术

人脸识别系统包括六部分：

1. 人脸检测：从输入图像中判断是否有人脸，如果有，给出人脸的位置和大小（即流程图中的矩形框）。2000年Viola和Jones提出基于AdaBoost的人脸检测方法。
2. 特征点定位：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。
  - ✓ 采用通用的目标检测技术
  - ✓ 或者采用回归方法
3. 面部子图预处理，实现人脸子图的归一化：
  - ✓ 把关键点对齐，即把所有人脸的关键点放到差不多接近的位置，消除人脸大小、旋转等影响
  - ✓ 对人脸核心区域子图进行光亮度方面的处理，消除光强弱、偏光等影响



## 10.4 应用：人脸识别技术

人脸识别系统包括六部分：

4. 特征提取：提取可以区分不同人的特征。
5. 特征对比：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
6. 决策：对相似度或距离进行阈值化。最简单的方法是，相似程度超过设定阈值的判断为相同人，否则为不同人。



**THANKS**

