# 计算机视觉

中国石油大学（华东）
青岛软件学院、计算机科学与技术学院
软件工程系
宫文娟

# 计算机视觉基础


- 第一章 概论
- 第二章 基础知识
- 第三章 图像分类
- 第四章 图像语义分割
- 第五章 目标检测
- 第六章 识别
- 第七章 目标跟踪
- 第八章 多目视觉
- 第九章 视觉问答

# 第一章 概论

- 1.1 计算机视觉的定义
- 1.2 计算机视觉的发展历史
- 1.3 计算机视觉的主要研究内容
- 1.4 计算机视觉的主要应用
- 1.5 计算机视觉的特点
- 1.6 实例：基于词袋的图像分类方法

# 1.1 计算机视觉的定义

计算机视觉方法也称为**机器视觉方法**，是基于图像处理、机器学习、模式识别等手段，让计算机（或者机器）能像人通过眼睛观察世界一样去理解世界的一类方法。

Computer vision is about “computing properties of the 3-D world from one or more digital images.” (Trucco & Verri, 1998)

# 1.1 计算机视觉的定义

# 1.1 计算机视觉的定义

[redacted]

# 1.1 计算机视觉的定义

**计算机视觉**

输入数据：二维数据或随时间变化的二维数据（通常是图像或者视频）

回答问题：输入数据中有什么物体，物体在什么地方，物体之间有什么关系等

# 1.1 计算机视觉的定义

**计算机图形学**

输入数据：三维数据信息

回答问题：如何构建三维数据以如何投影得到二维图像

# 1.2 计算机视觉的发展历史

1959年，两位神经生理学家David Hube 和Torsten Wiesel研究了猫的视皮质神经元的特性、反应、视觉经验对其皮质结构的影响。

1959年，Russel Kirsch等研制了设备，可以将影像转换成坐标方格样式的数字。

1963年，MIT的Larry Roberts提出将世界分解成简单几何体的机器感知模型。

20世纪60年代，人工智能成为一个学科。

1982年，英国神经科学家David Marr发表了《视觉：人类表示和处理视觉信息的计算研究》。

1986年，韦兰研究公司的Robert K. McConnell在申请的专利中提出定向梯度直方图（HOG）方法。

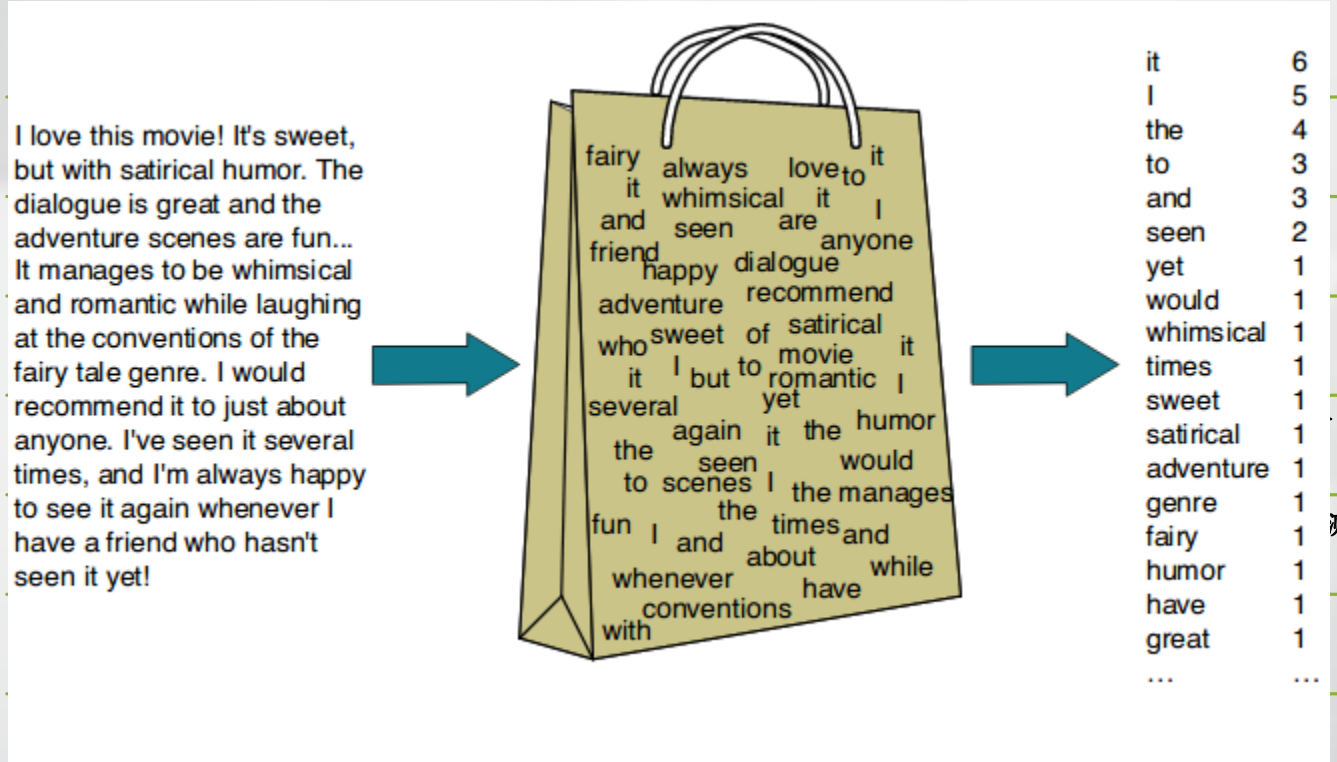
2005年，法国国家科学院的Navneet Dalal和Bill Triggs在CVPR上发表文章，用HOG进行行人检测，使得HOG得到广泛应用。

1999年，Lowe提出尺度不变特征变换（SIFT）方法，并在2004年获得专利。

1954年，Zellig Harris提出词袋并应用在语境分析中。

# 1.2 计算机视觉的发展历史

1959年，两位神经生理学家David Hube 和Torsten Wiesel研究了猫的视皮质神经元的特性、反应、视觉经验对其皮质结构的影响。




法。

测，使得

# 1.2 计算机视觉的发展历史

1959年，两位神经生理学家David Hube 和Torsten Wiesel研究了猫的视皮质神经元的特性、反应、视觉经验对其皮质结构的影响。

1959年，Russel Kirsch等研制了设备，可以将影像转换成坐标方格样式的数字。

1963年，MIT的Larry Roberts提出将世界分解成简单几何体的机器感知模型。

20世纪60年代，人工智能成为一个学科。

1982年，英国神经科学家David Marr发表了《视觉：人类表示和处理视觉信息的计算研究》。

1986年，韦兰研究公司的Robert K. McConnell在申请的专利中提出定向梯度直方图（HOG）方法。

2005年，法国国家科学院的Navneet Dalal和Bill Triggs在CVPR上发表文章，用HOG进行行人检测，使得HOG得到广泛应用。

1999年，Lowe提出尺度不变特征变换（SIFT）方法，并在2004年获得专利。

1954年，Zellig Harris提出词袋（Bag of Words）并应用在语境分析中。

2009年，Josef Sivic采用词袋模型（Bag of Visual Words）解决计算机视觉问题。

# 1.2 计算机视觉的发展历史

1959年，两位
验对其皮质结
反应、视觉经

1959年，Russe

1963年，MIT的

20世纪60年代

1982年，英国
。

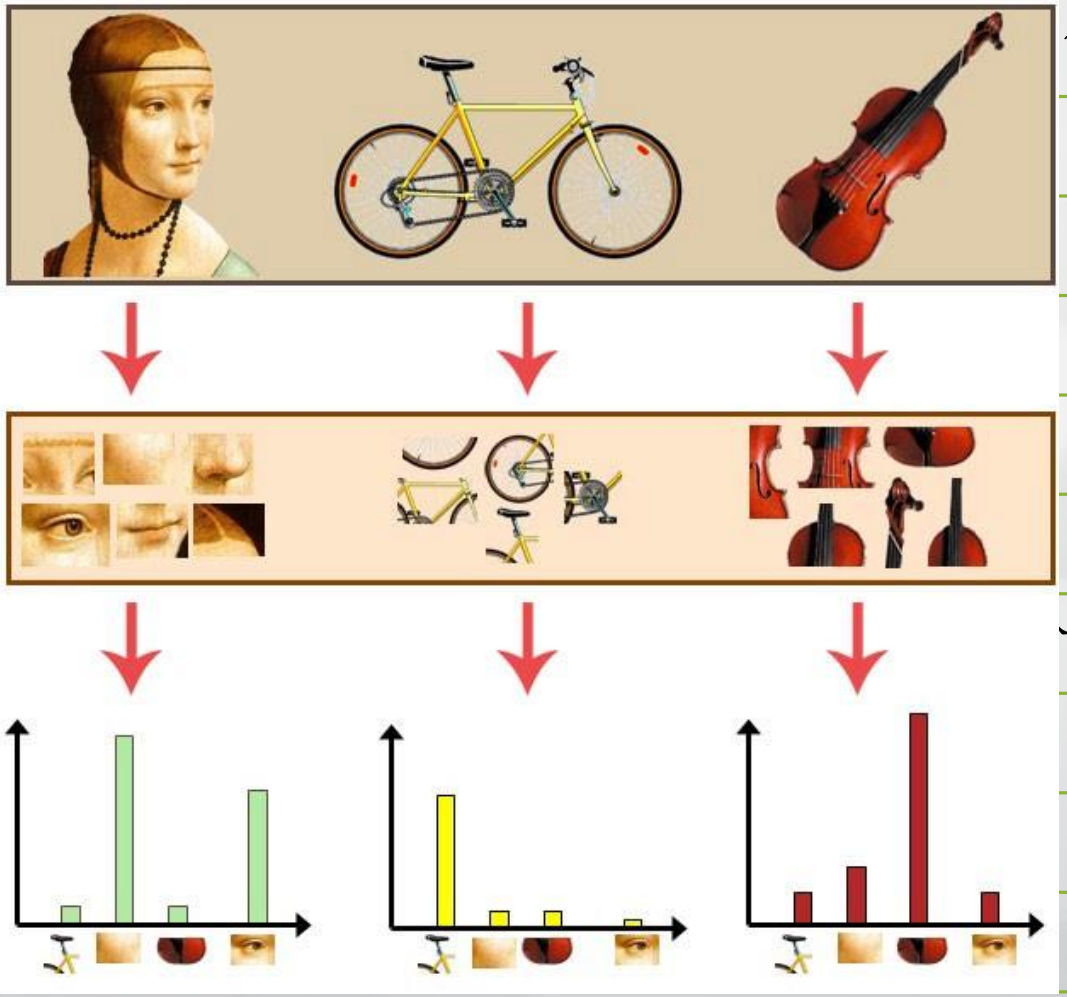
1986年，韦兰
方法。

2005年，法国
HOG得到广泛
检测，使得

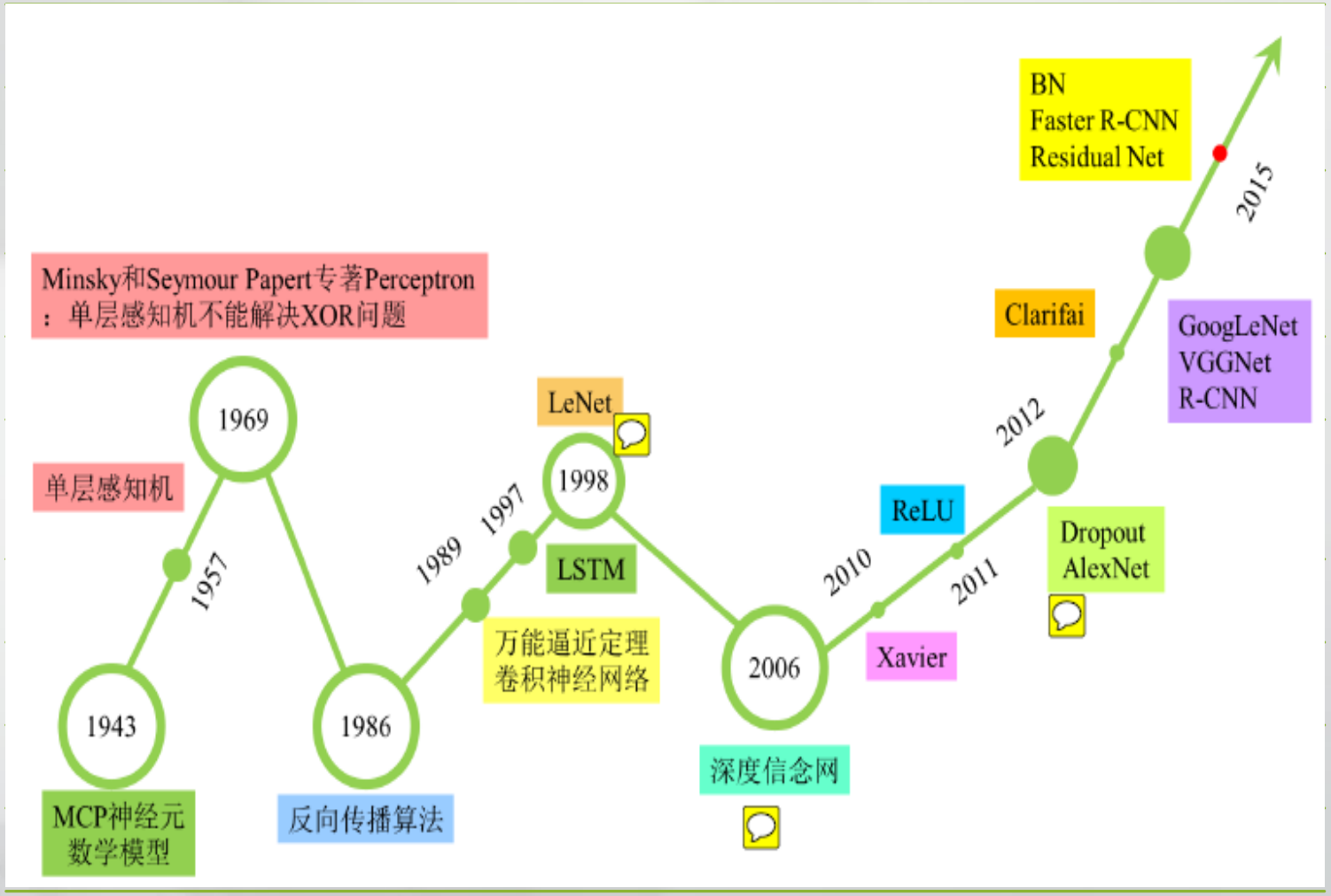
1999年，Lowe

1954年，Zellig

2009年，Josef

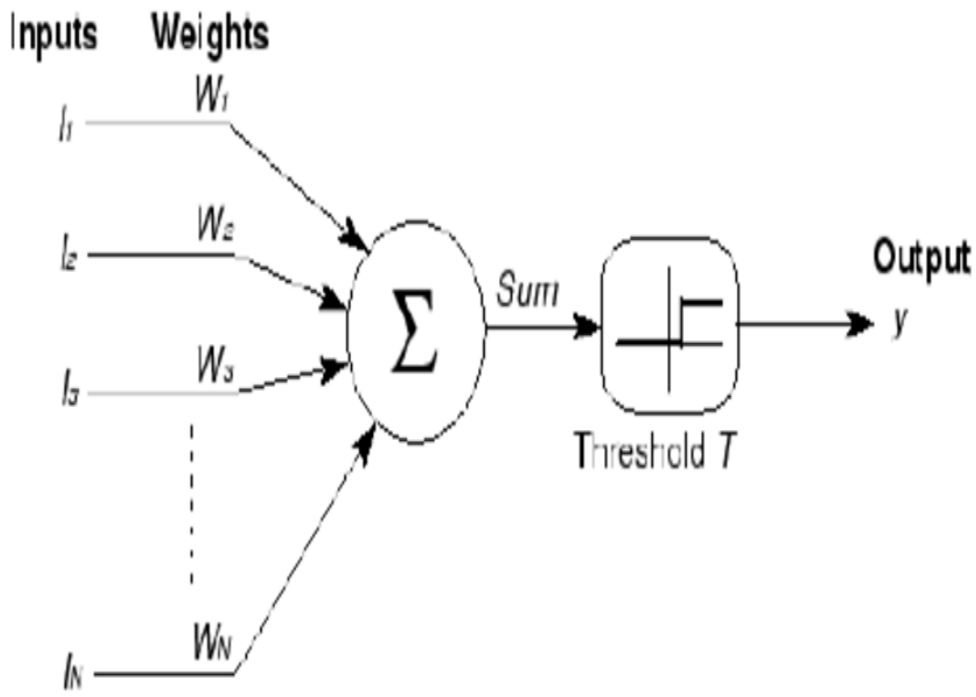


# 1.2 计算机视觉的发展历史



# 1.2 计算机视觉的发展历史

1943年，神经科学家麦卡洛克W.S.McCilloch 和数学家皮兹W.Pitts提出MCP神经元模型。




$$y_k = \varphi\left(\sum_{j=0}^{m} w_{kj} x_j\right)$$

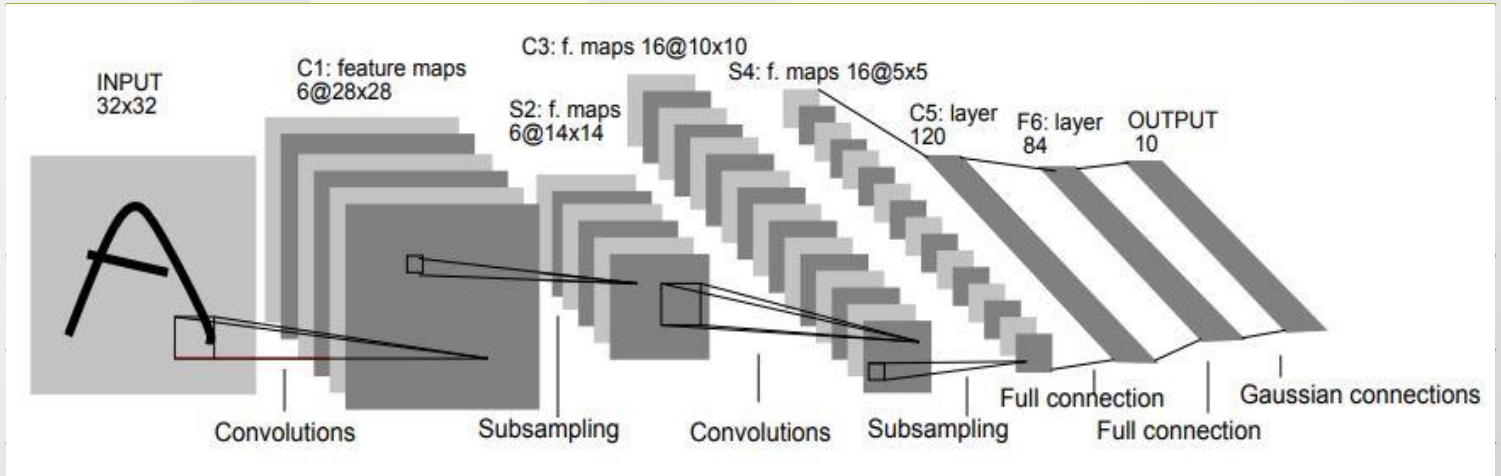
# 1.2 计算机视觉的发展历史

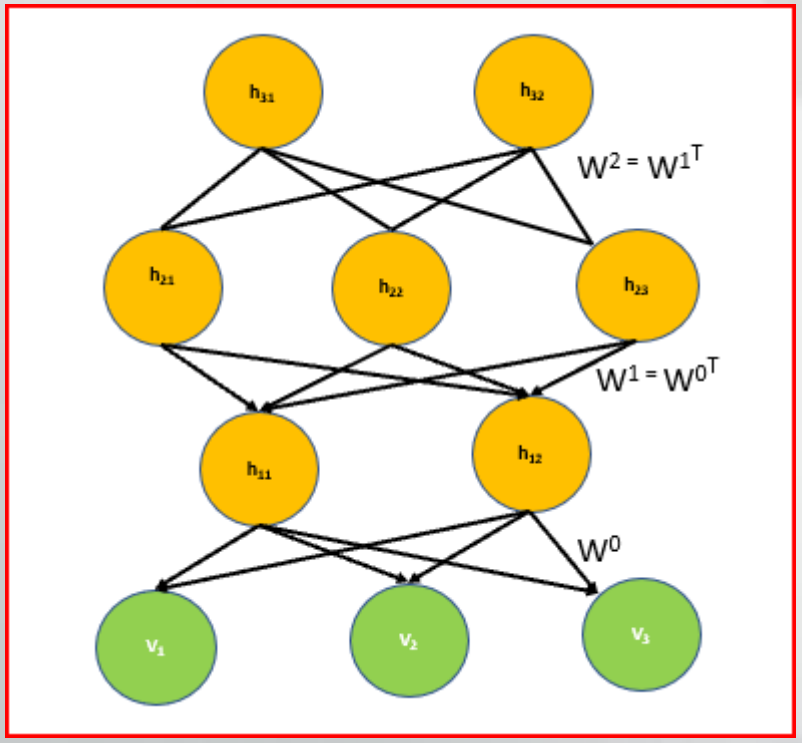
**1943年**，神经科学家麦卡洛克W.S.McCilloch 和数学家皮兹W.Pitts提出MCP神经元模型。

**1969年**，计算机科学家 Rosenblatt提出了两层神经元组成的神经网络，也称为单层感知机Perceptron。Minsky和Seymour Papert研究发现单层感知机不能解决XOR问题。

**1986年**，神经网络之父 Geoffrey Hinton 发明了适用于多层感知器MLP的BP（Backpropagation）算法反向传播。

**1998年**，Yan Le Cunn提出LeNet，用于手写字体识别。





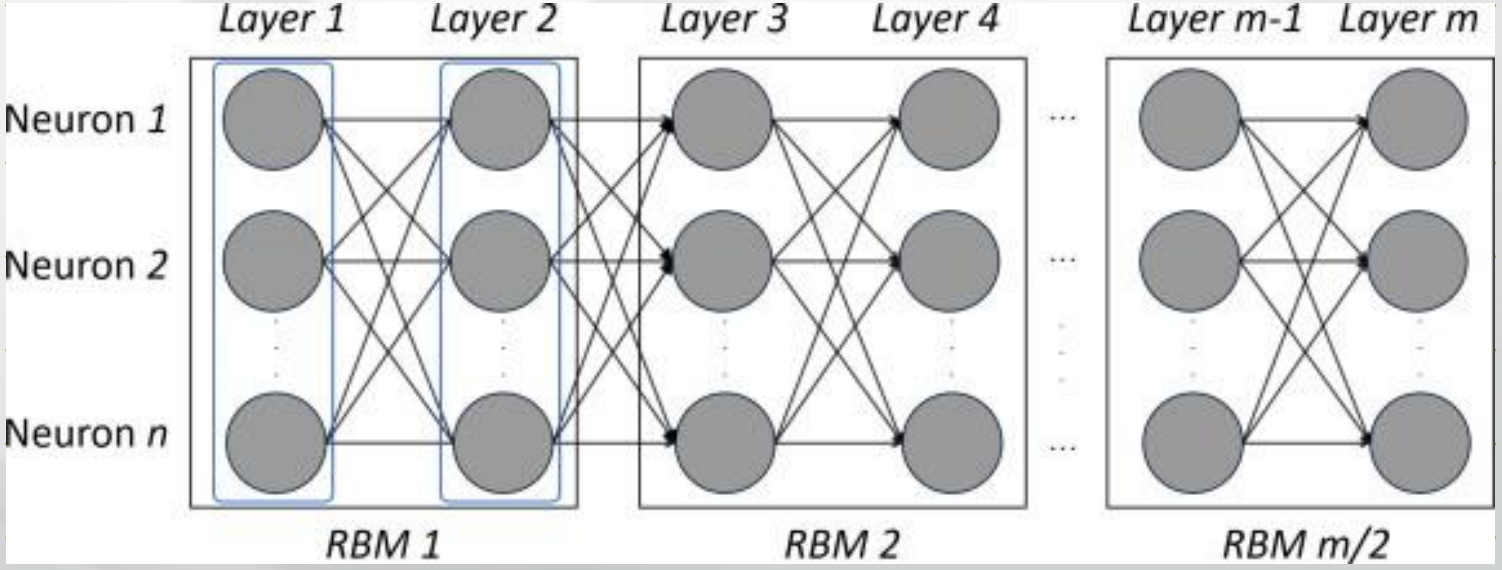

**1943年**，神经科学家麦卡洛克W.S.McCilloch 和数学家皮兹W.Pitts提出MCP神经元模型。

**1969年**，计算机科学家 Rosenblatt提出了两层神经元组成的神经网络，也称为单层感知机Perceptron。Minsky和Seymour Papert研究发现单层感知机不能解决XOR问题。

**1986年**，神经网络之父 Geoffrey Hinton 发明了适用于多层感知器MLP的BP（Backpropagation）算法反向传播。

**1998年**，Yan Le Cunn提出LeNet，用于手写字体识别。

**2006年**，Geoffrey Hinton和他的学生提出了深度置信网络。



# 1.2 计算机视觉的发展历史

**1943年**，神经科学家麦卡洛克W.S.McCilloch 和数学家皮兹W.Pitts提出MCP神经元模型。

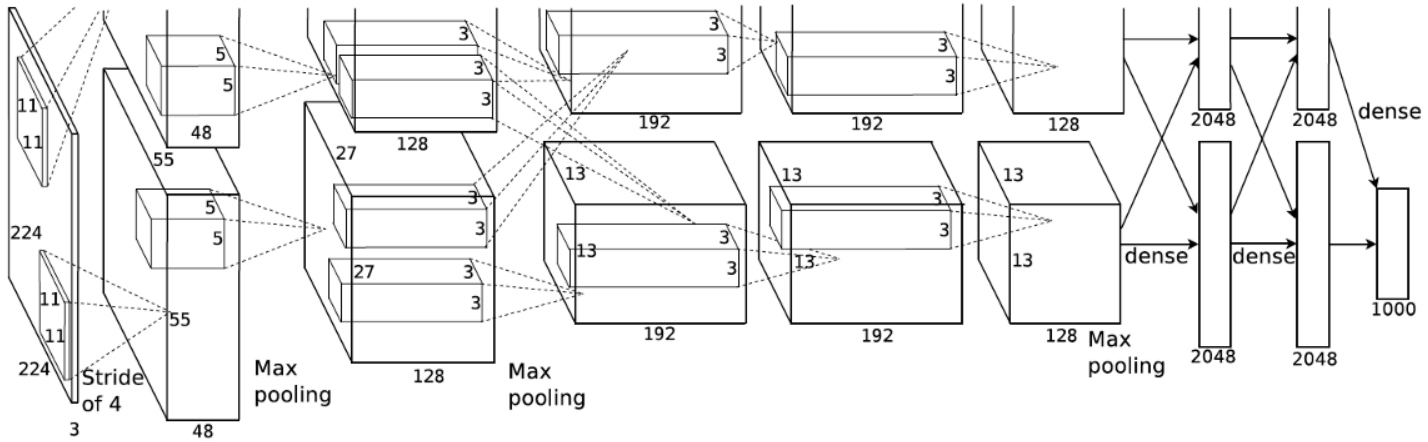
**1969年**，计算机科学家 Rosenblatt提出了两层神经元组成的神经网络，也称为单层感知机Perceptron。Minsky和Seymour Papert研究发现单层感知机不能解决XOR问题。

**1986年**，神经网络之父 Geoffrey Hinton 发明了适用于多层感知器MLP的BP（Backpropagation）算法反向传播。

**1998年**，Yan Le Cunn提出LeNet，用于手写字体识别。

**2006年**，Geoffrey Hinton和他的学生提出了深度置信网络。

**2012年**，Hinton课题组为了证明深度学习的潜力，首次参加ImageNet图像识别比赛，其通过构建的CNN网络AlexNet一举夺得冠军，且碾压第二名（SVM方法）的分类性能。也正是由于该比赛，CNN吸引到了众多研究者的注意。



# 1.2 计算机视觉的发展历史

**1943年**，神经科学家麦卡洛克W.S.McCilloch 和数学家皮兹W.Pitts提出MCP神经元模型。

**1969年**，计算机科学家 Rosenblatt提出了两层神经元组成的神经网络，也称为单层感知机Perceptron。Minsky和Seymour Papert研究发现单层感知机不能解决XOR问题。

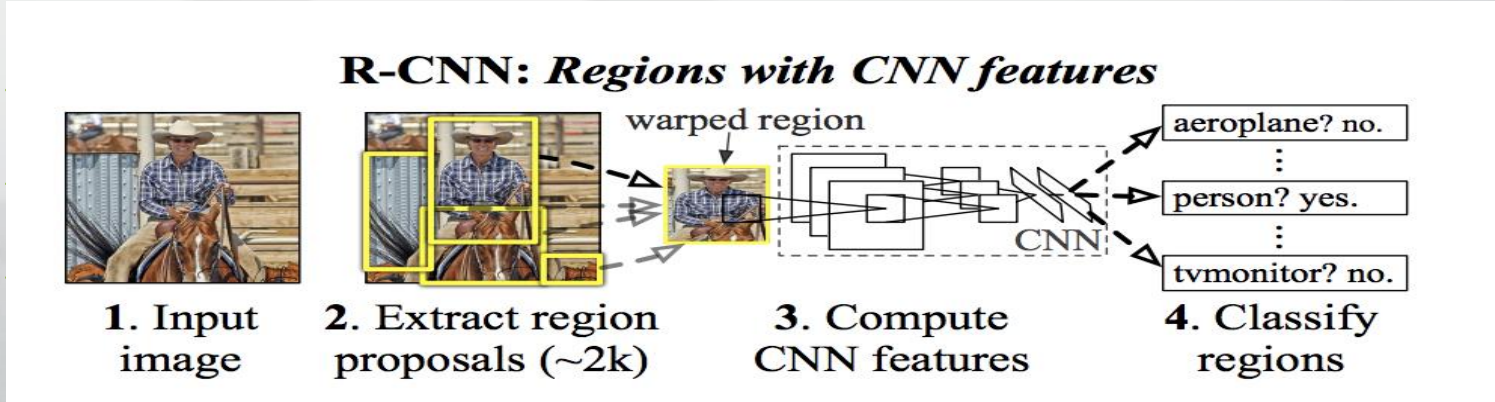
**1986年**，神经网络之父 Geoffrey Hinton 发明了适用于多层感知器MLP的BP（Backpropagation）算法反向传播。

**1998年**，Yan Le Cunn提出LeNet，用于手写字体识别。

**2006年**，Geoffrey Hinton和他的学生提出了深度置信网络。

**2012年**，Hinton课题组为了证明深度学习的潜力，首次参加ImageNet图像识别比赛，其通过构建的CNN网络AlexNet一举夺得冠军，且碾压第二名（SVM方法）的分类性能。也正是由于该比赛，CNN吸引到了众多研究者的注意。

**2014年**，谷歌提出GoogleLeNet；牛津大学的视觉几何组(Visual Geometry Group)提出VGG；Ross Girshick等人提出R-CNN



# 1.2 计算机视觉的发展历史

**1943年**，神经科学家麦卡洛克W.S.McCilloch 和数学家皮兹W.Pitts提出MCP神经元模型。

**1969年**，计算机科学家 Rosenblatt提出了两层神经元组成的神经网络，也称为单层感知机Perceptron。Minsky和Seymour Papert研究发现单层感知机不能解决XOR问题。

**1986年**，神经网络之父 Geoffrey Hinton 发明了适用于多层感知器MLP的BP（Backpropagation）算法反向传播。

**1998年**，Yan Le Cunn提出LeNet，用于手写字体识别。

**2006年**，Geoffrey Hinton和他的学生提出了深度置信网络。

**2012年**，Hinton课题组为了证明深度学习的潜力，首次参加ImageNet图像识别比赛，其通过构建的CNN网络AlexNet一举夺得冠军，且碾压第二名（SVM方法）的分类性能。也正是由于该比赛，CNN吸引到了众多研究者的注意。

**2014年**，谷歌提出GoogleLeNet；牛津大学的视觉几何组(Visual Geometry Group)提出VGG；Ross Girshick等人提出R-CNN

**2015年**，何凯明等提出了ResNet。

# 1.3 计算机视觉的主要研究内容

- 图像分类（第三章）
- 图像分割（第四章）
- 物体检测（第五章）
- 识别（第六章）
- 运动物体跟踪（第七章）
- 多目视觉（第八章）
- 视觉问答（第九章）

# 1.3 计算机视觉的主要研究内容

图像分割（第四章）

# 1.3 计算机视觉的主要研究内容

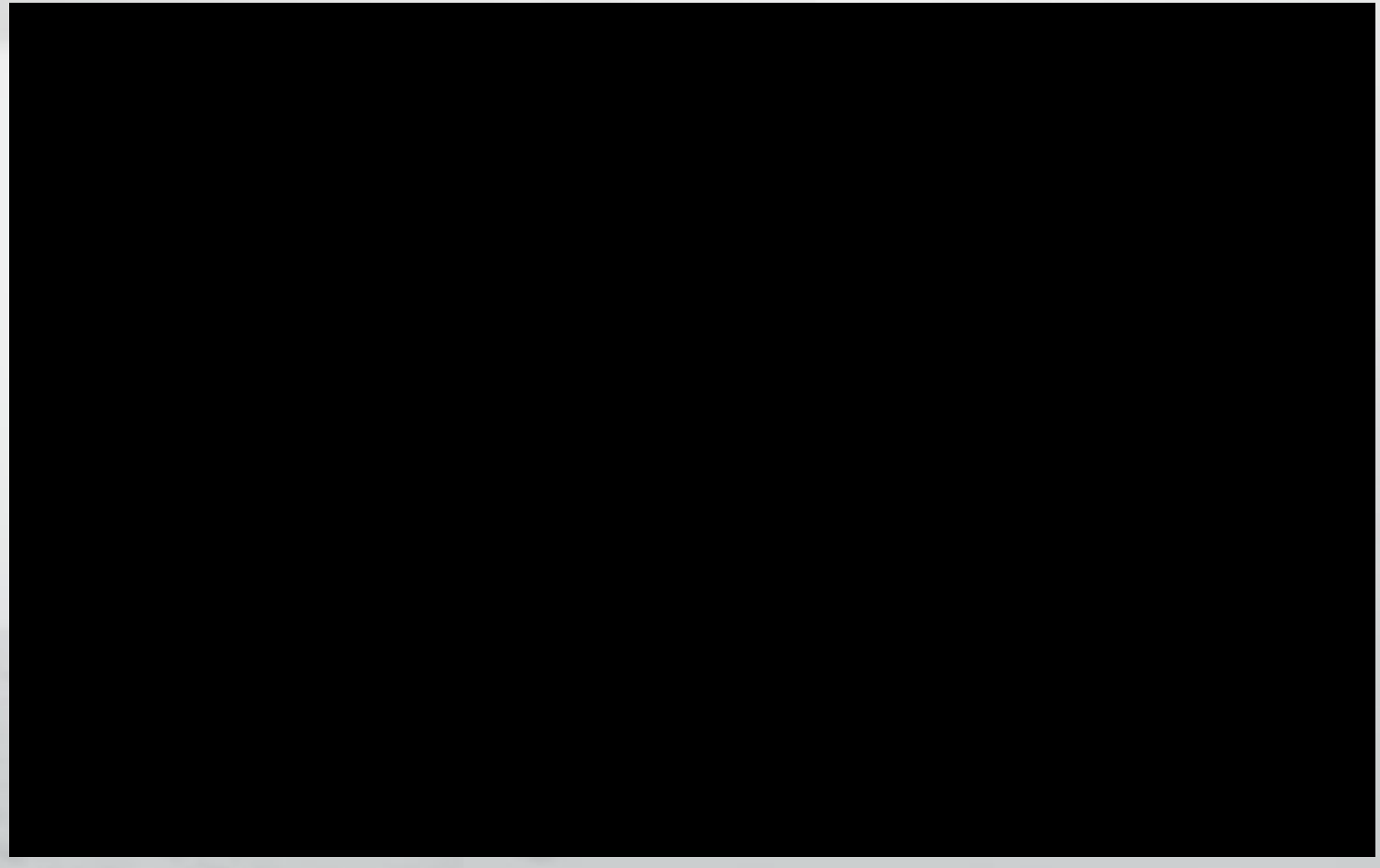
[redacted]

物体检测（第五章）

# 1.3 计算机视觉的主要研究内容

运动物体跟踪（第七章）

# 1.3 计算机视觉的主要研究内容



多目视觉（第八章）

# 1.3 计算机视觉的主要研究内容

[redacted]

视觉问答（第九章）

# 1.3 计算机视觉的主要研究内容

[redacted]

# 1.4 计算机视觉的主要应用

| | | |
|---|---|---|
| 航空航天和军事 | 医学 | 工业 |
| 安防 | 交通 | 体育 |
| 娱乐 | 商业 | 公众服务 |

# 1.5 计算机视觉的特点

很多计算机视觉问题采用数字图像作为输入，最普遍是RGB图像。

很多计算机视觉问题本身很主观，不能被很好的量化定义，例如：

- 表情识别中对一个人微笑、愤怒、开心、难过的定义
- 对一副图像是否美观的定义

# 1.6实例：基于词袋的图像分类方法

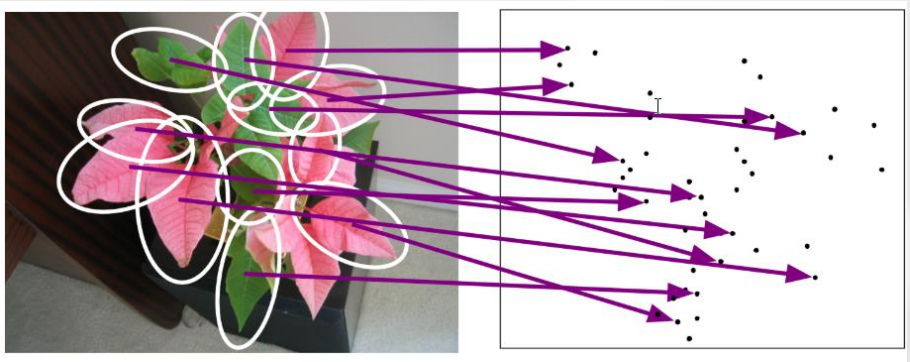
（a）原始图像采样得到（b）图像块

对这些图像块提取特征点（c）

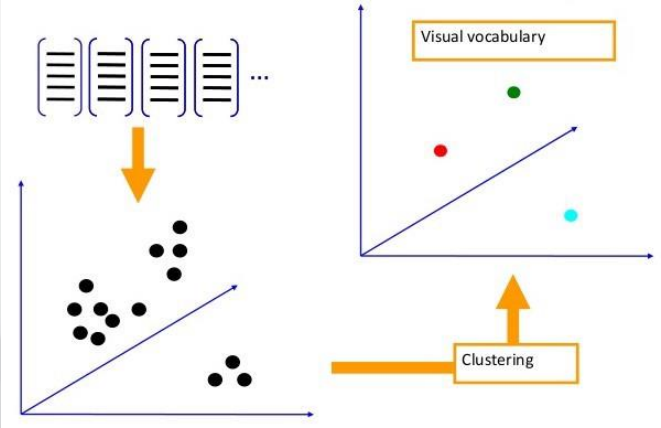
对提取的特征点进行特征描述（d）

对所有的训练图像进行聚类，得到的聚类中心作为词典（e）

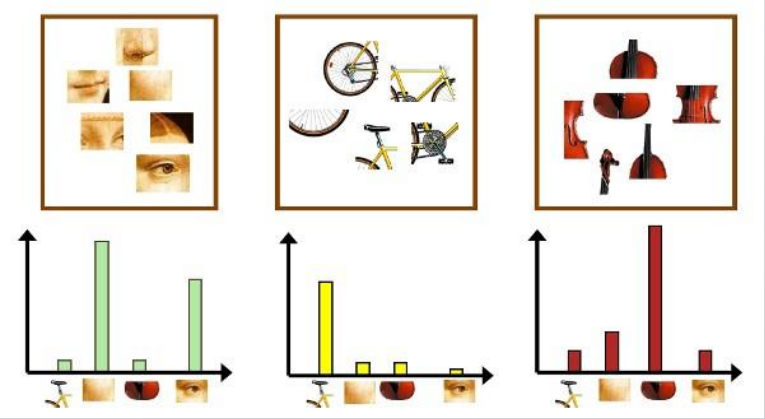
则每幅图像可以重新表示成基于词典的直方图（f）



（c）提取SIFT特征点




（de）提取SIFT特征点



（f）计算直方图

图1-2 基于词袋模型的表示方法的主要流程

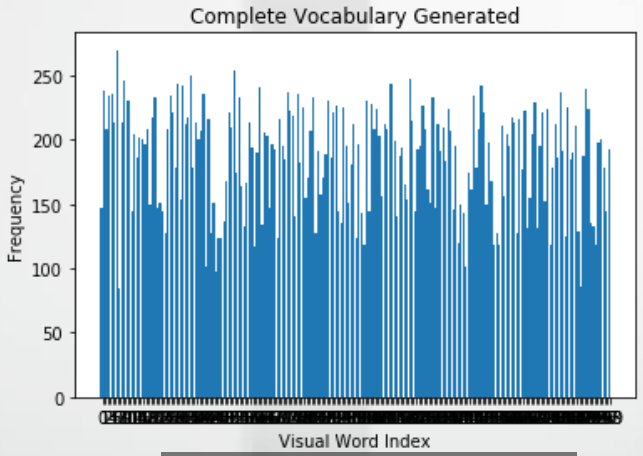
# 1.6实例：基于词袋的图像分类方法

- city
- face
- green
- house_building
- house_indoor
- office
- sea

数据库

```
Train images path detected.
Descriptors vstacked.
Descriptors clustered.
Images features extracted.
Train images normalized.
```

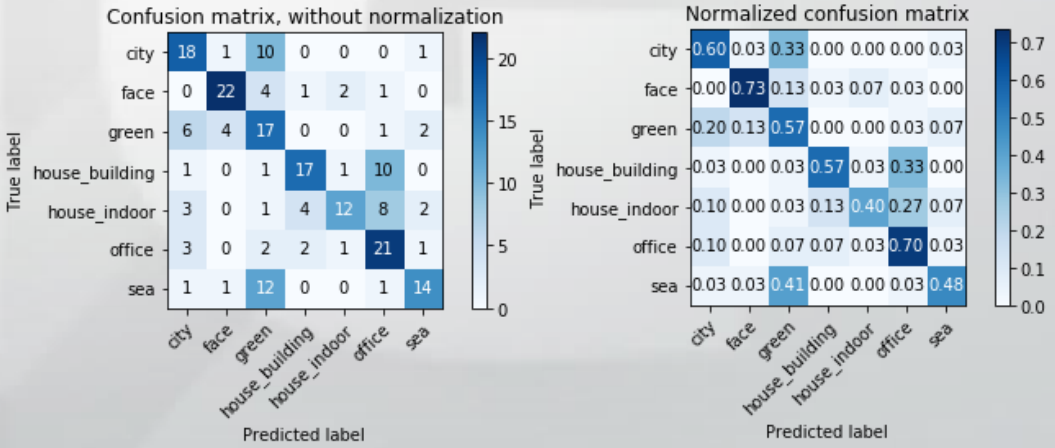
1. 计算特征




2. 大小为200的词典

accuracy score: 0.579

3. 识别结果




4. 混淆矩阵

基于词袋模型的表示方法的结果显示