

# 图 象 理 解

（第4版）

章毓晋

清华大学电子工程系 100084 北京

## 第3单元 场景解释

- 第10章 知识表达和推理
- 第11章 广义匹配
- 第12章 场景分析和语义解释

通过学习、推理、与模型的匹配等来解释场景的内容、特性、变化、态势或趋向

知识指导对客观世界认识和理解

将从图象中获得的信息与已有的解释场景的模型进行匹配

对场景的高层次解释和语义描述

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第2页

## 第12章 场景分析和语义解释

- 12.1 场景理解概述
- 12.2 模糊推理
- 12.3 遗传算法图象解释
- 12.4 场景目标标记
- 12.5 场景分类

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第3页

## 12.1 场景理解概述

### 1. 场景分析

获取场景中景物的信息，以进行场景解释

目标识别是场景分析的重要步骤和基础

要考虑目标本身的内部（结构）关系，还要关注目标之间的分布和相对位置

从认知的角度看，场景分析（比目标识别）更关注人对场景（整体）的感知和理解

常结合生物学、生理学和心理学研究成果

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第4页

## 12.1 场景理解概述

### 1. 场景分析

场景的视觉内容（景物及分布）多样，且具有很大的不确定性：

- (1) 不同的光照条件，影响景物检测和跟踪
- (2) 不同的景物外在表现（有时尽管结构元素类似），影响对景物的识别（歧义）
- (3) 不同的观察尺度，影响对景物的辨识
- (4) 不同的景物位置、朝向以及互相遮挡因素等，增加了认知景物的复杂性

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第5页

## 12.1 场景理解概述

### 2. 场景感知层次

分三个层次进行对场景的分析和语义解释

- (1) 局部层：该层主要强调对场景的局部或单个景物进行分析、识别或对图象区域进行标记
- (2) 全局层：该层考虑整个场景的全局，关注具有相似外形和类似功能的景物之间的相互关系
- (3) 抽象层：该层对应场景的概念含义，给出场景抽象的描述

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第6页

## 12.1 场景理解概述

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第7页

### 2. 场景感知层次

人类对场景的感知有很强的能力

对中层（目标）的感知有一定的优先级

人类对中层的大部分目标（如教室/运动场）具有很快和很强的辨识能力，对它们的识别、命名比对低层和高层都要快

对中层感知更加优先是由于它能同时最大化类内相似度和最大化类间的差异（同在中层的景物常有相似的空间结构及相似的行为）

## 12.1 场景理解概述

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第8页

### 2. 场景感知层次

建立高层概念与低层视觉特征和中层目标特性间的联系，并识别景物及它们之间的相对关系

(1) 低层场景建模

先直接对景物的低层属性进行表达和描述，然后借助分类识别再对场景的高层信息进行推理

(2) 中层语义建模

借助对目标的识别来提高低层特征分类性能如：借助视觉词汇建模（12.5.1）

## 12.1 场景理解概述

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第9页

### 3. 场景语义解释

多方面技术

- (1) 视频计算技术
- (2) 视觉算法动态控制策略
- (3) 对场景信息的自学习
- (4) 快速或实时计算技术
- (5) 多传感器融合协同
- (6) 视觉注意机制
- (7) 结合认知理论的场景解释
- (8) 系统的集成与优化

## 12.2 模糊推理

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第10页

模糊是一个与清晰或精确对立的观念，表达各种各样不明确、不严格、不确定的知识和信息

12.2.1 模糊集和模糊运算

12.2.2 模糊推理方法

## 12.2.1 模糊集和模糊运算

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第11页

在模糊空间 $X$ 中的一个模糊集合 $S$ 是一个有序的集合  $S = \{[x, M_S(x)] | x \in X\}$

$M_S(x)$ 代表 $x$ 在 $S$ 中的隶属程度，总取非负实数

模糊集合常可以用其**隶属度函数**唯一地描述

精确集

模糊集

模糊集

图 12.2.1 精确集和模糊集表示示意图

## 12.2.1 模糊集和模糊运算

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第12页

### 模糊逻辑运算（对模糊集合的运算）

类似一般逻辑运算的名称但定义不同的运算

交 Intersection  $A \cap B: M_{A \cap B}(x, y) = \min[M_A(x), M_B(y)]$

并 Union  $A \cup B: M_{A \cup B}(x, y) = \max[M_A(x), M_B(y)]$

补 Complement  $A^c: M_{A^c}(x) = 1 - M_A(x)$

代数运算

(a) 加强模糊集VD (very dark)

$$M_{VD}(x) = M_D^2(x) \cdot M_D(x) = M_D^3(x)$$

(b) 加强加强模糊集VVD (very very dark)

$$M_{VVD}(x) = M_D^2(x) \cdot M_D^2(x) = M_D^4(x)$$

12.2.1 模糊集和模糊运算

代数运算

(c) 减弱模糊集SD (somewhat dark)

$M_{SD}(x) = \sqrt{M_D(x)}$

(d) 加强模糊集VD的非, 即NVD (not very dark)

$M_{NVD}(x) = 1 - M_D^2(x)$

图 12.2.2 对图 12.2.1(b) 的原始模糊集 D 的一些运算结果

第12讲 章毓晋 (TH-EE-IE) ZHANG YU JIN 第13页

12.2.2 模糊推理方法

将模糊集中的信息以一定规则结合起来做出决策

1. 基本模型

模糊规则: 一系列无条件的和有条件的命题

无条件模糊命题:  $x$  is  $A$

有条件模糊命题: If  $x$  is  $A$  then  $y$  is  $B$

图 12.2.3 模糊推理模型和步骤

第12讲 章毓晋 (TH-EE-IE) ZHANG YU JIN 第14页

12.2.2 模糊推理方法

1. 基本模型

与无条件模糊规则对应的隶属度就是  $M(x)$

有条件模糊规则: 最简单的是单调模糊推理

模糊规则: if  $x$  is DARK then  $y$  is LOW

图 12.2.4 基于单个模糊规则的单调模糊推理

第12讲 章毓晋 (TH-EE-IE) ZHANG YU JIN 第15页

12.2.2 模糊推理方法

2. 模糊结合 (确定隶属度)

(1) 最小-最大规则 (最小化和最大化)

用最小相关  $M_A(x)$  来限定模糊结果的  $M_B(y)$

图 12.2.5 使用最小相关的模糊最小-最大结合

第12讲 章毓晋 (TH-EE-IE) ZHANG YU JIN 第16页

12.2.2 模糊推理方法

2. 模糊结合

(2) 相关积: 对原始的结果模糊隶属度函数进行放缩而不是截去, 保持原始模糊集的形状

图 12.2.6 使用相关积的模糊最小-最大结合

第12讲 章毓晋 (TH-EE-IE) ZHANG YU JIN 第17页

12.2.2 模糊推理方法

3. 去模糊化

确定一个能最好地表达模糊解集中信息的、包含多个标量 (每个标量对应一个解变量) 的矢量

先确定模糊解的隶属度函数的重心  $c$ , 并将模糊解转化为一个清晰解  $c$

在模糊解的隶属度函数中确定具有最大隶属度值的域点给出一个清晰解  $d$

图 12.2.7 去模糊化的两种方法

第12讲 章毓晋 (TH-EE-IE) ZHANG YU JIN 第18页

3

## 12.3 遗传算法图象解释

使用自然进化机制搜索目标函数的极值

### 12.3.1 遗传算法原理

### 12.3.2 语义分割和解释

第12讲
章毓晋 (TH-EE-IE) ZHANG YU JIN
第19页

## 12.3.1 遗传算法原理

### 遗传算法的特点

- (1) 将优化问题的自然参数集合编码成有限长度的码串（常使用字符只有0和1的二值码串）
- (2) 在搜索空间里同时从大量群体样本点中进行搜索，发现全局最优的机会很大
- (3) 直接使用目标函数——进化函数，其值称为遗传算法的**适应度**
- (4) 使用概率转换规则而不是确定性的规则，用高适应度支持好码串而消除低适应度的差码串（最好的码串将在进化过程中有高的概率存活）

第12讲
章毓晋 (TH-EE-IE) ZHANG YU JIN
第20页

## 12.3.1 遗传算法原理

### 遗传算法的基本运算

- (1) **复制**：根据概率让好（高适应度）的码串存活而让其他码串死亡
- (2) **交叉**：对换码串对中边界位置前/后（头和尾）码串来产生新的码串

- (3) **变异**：频繁地随机改变某些码串的某个码，以保持各种局部结构避免丢失一些优化解特性

第12讲
章毓晋 (TH-EE-IE) ZHANG YU JIN
第21页

## 12.3.1 遗传算法原理

### 遗传算法的步骤

- (1) 生成初始群体的编码字符串（码串），计算适应度，即给出目标函数（进化函数）的值
- (2) 在新群体中概率化地**复制**高适应度的码串，除去低适应度的码串
- (3) 通过**交叉**组合从旧群体复制的码串构建新码串
- (4) 不时地随机选择码串中的某个码进行**变异**
- (5) 对当前群体中的码串根据其适应度进行排序
- (6) 如果最大适应度的码串的适应度值在若干个进化过程中都没有明显增加，停止；否则返回步骤(2)继续计算

第12讲
章毓晋 (TH-EE-IE) ZHANG YU JIN
第22页

## 12.3.2 语义分割和解释

### 语义分割

指将图象根据语义信息划分为对应的区域并借助上下文等高层信息进行优化

基于**假设和验证**的原理对图象进行解释

用遗传算法所优化的目标函数评价图象语义分割和解释的质量

- 初始分割：先从一个过分割的图象（起始区域称为初始区域）开始
- 迭代合并：不断将初始区域更新为当前区域，即持续构建可行的区域划分并解释假设的新样本

第12讲
章毓晋 (TH-EE-IE) ZHANG YU JIN
第23页

## 12.3.2 语义分割和解释

### 对目标函数的优化

- (1) 根据区域本身性质 $X_i$ 对区域 $R_i$ 进行解释 $k_i$ 的置信度（与相应的概率成比例）  

$$C(k_i | X_i) \propto P(k_i | X_i)$$
- (2) 根据对其邻近区域 $R_j$ 的解释 $k_j$ 来对区域 $R_i$ 进行解释 $k_i$ 的置信度
- (3) 对整幅图象中的解释置信度的评价  

$$C_{\text{image}} = \frac{1}{N_R} \sum_{i=1}^{N_R} C(k_i) \quad C'_{\text{image}} = \sum_{i=1}^{N_R} \left[ \frac{C(k_i)}{N_R} \right]^2$$

第12讲
章毓晋 (TH-EE-IE) ZHANG YU JIN
第24页

## 12.3.2 语义分割和解释

### 基本步骤

- (1) 初始化图象为初始区域，定义各个区域与它的标记在用遗传算法生成的码串中的相对位置的对应关系
- (2) 构建一个初始区域邻接图
- (3) 随机地选取码串的起始群体
- (4) 遗传优化，计算优化分割函数的值
- (5) 如果优化准则最大值不变，转到步骤(7)
- (6) 生成一个分割解释假设的新群体，转(4)
- (7) 具有最大置信度的码串代表最终结果

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第25页

## 12.4 场景目标标记

对图象中目标区域的语义标记，即赋予目标以语义含义的符号。对场景图象中的每个目标赋一个标记（标签）以获得对场景图象的恰当解释

### 1. 标记方法和要素

- (1) 离散标记：它对每个目标只赋予1个标记，主要考虑对图象标记的一致性
- (2) 概率标记：它允许对联合存在的目标赋予多个标记。这些标记是用概率加权的，每个标记都有一个标记信任度

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第26页

## 12.4 场景目标标记

### 1. 标记方法和要素

- (1) 离散标记：能获得一致性的标记，或能检测出要赋予场景一致性标记的不可能性  
？ 分割不完善会导致不能给出一致性解释
- (2) 概率标记：总能给出标记结果，常比离散标记所给出的一致但很不可能的解释要更好

**标记要素：**①一组目标，②一个有限的标记集合，③一个有限的目标间关系集合，④相关目标间存在的兼容性函数（约束）

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第27页

## 12.4 场景目标标记

### 2. 离散松弛标记

目标本身的性质可用一元的关系来描述

目标间的联系需用二元或多元的联系来描述

- 一 ①窗户矩形；②桌子矩形；③抽屉矩形
- 二 ①窗户位于桌子上方  
②电话放在桌子上  
③抽屉在桌子内部  
④背景与图象边缘相连

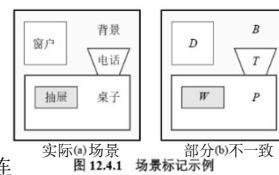


图 12.4.1 场景标记示例

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第28页

## 12.4 场景目标标记

### 2. 离散松弛标记

先把所有存在的标记都赋给每个目标  
然后迭代地对每个目标进行一致性检验，将那些有可能不满足约束的标记除去

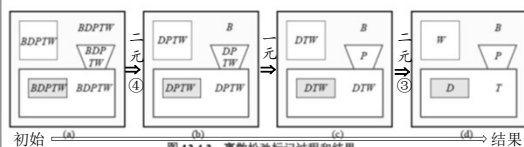


图 12.4.2 离散松弛标记过程和结果

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第29页

## 12.4 场景目标标记

### 3. 概率松弛标记

有可能克服使用离散松弛标记时由于分割中丢失目标或多出目标带来的问题

目标 $R_j$ 被标记为 $q_j$ ;  $q_j \in Q$ ,  $Q = \{w_1, w_2, \dots, w_T\}$

算法迭代地搜索在整幅图中的局部最一致性 (P.291~P.292)  
目标函数值最大化就获得最优的标记

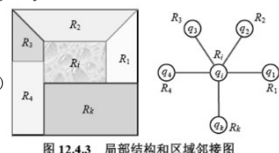


图 12.4.3 局部结构和区域邻接图

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第30页

## 12.5 场景分类

根据视觉感知组织原理，确定出图像中存在的各种特定区域，并给出场景的概念性解释

### 12.5.1 词袋/特征包模型

### 12.5.2 pLSA模型

### 12.5.3 LDA模型

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第31页

## 12.5.1 词袋/特征包模型

**词袋模型**源自对自然语言的处理，引入图像领域后也常称为**特征包模型**

特征包模型由类别特征（feature）归属于同类目标集中形成包（bag）而得名

用特征包模型表达和描述场景需要从场景中抽取局部区域描述特征，以构建视觉单词

构建**视觉词汇**集合（词典）可包括如下几个方面：①提取特征；②组成视觉单词；③量化视觉单词并构建码本；④利用视觉词汇的频率表达图像

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第32页

## 12.5.2 pLSA模型

概率隐语义分析模型源于概率隐语义索引，是为解决目标和场景分类而建立的一种图模型

### 1. 模型描述

图像集合 $T = \{t_1, \dots, t_N\}$ 所包含的视觉单词来自单词集合——词典（视觉词汇表） $S = \{s_1, \dots, s_M\}$

用尺寸为 $N \times M$ 的统计共生矩阵 $P$ 来描述图像集合 $T$ 的性质，矩阵中每个元素 $p_{ij} = p(t_i, s_j)$ 表示图像 $t_i$ 中单词 $s_j$ 出现的频率

该矩阵实际中是一个稀疏矩阵

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第33页

## 12.5.2 pLSA模型

### 1. 模型描述

隐变量（称主题变量） $z \in Z = \{z_k\}, k = 1, \dots, K$

基于主题与单词共生矩阵的条件概率模型

$$p(s_j | t_i) = \sum_{k=1}^K p(s_j | z_k) p(z_k | t_i) \quad \text{P中元素} \quad p(t_i, s_j) = p(t_i) p(s_j | t_i)$$

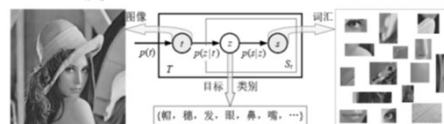


图 12.5.3 pLSA 模型示意

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第34页

## 12.5.2 pLSA模型

### 1. 模型描述

pLSA模型的目标是搜索特定主题 $z_k$ 下的词汇分布概率 $p(s_j | z_k)$ 和所对应特定图像中的混合比例 $p(z_k | t_i)$ ，从而获得特定图像中的词汇分布 $p(s_j | t_i)$

每列代表给定主题中的视觉词汇

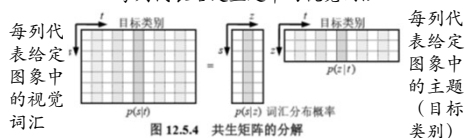


图 12.5.4 共生矩阵的分解

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第35页

## 12.5.2 pLSA模型

### 2. 模型计算

确定对所有图像公共的主题矢量和对每幅图像特殊的混合系数，其目的是确定对图像中出现的单词给以高概率的模型，从而可选取最大后验概率对应的类别作为最终的目标类别

$$\text{优化目标函数} \quad L = \prod_{j=1}^M \prod_{i=1}^N p(s_j | t_i) p(t_i, s_j)$$

对隐变量模型的最大似然估计可采用**最大期望/期望最大化（EM）**算法计算 {P.296}

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第36页

## 12.5.2 pLSA模型

### 3. 模型应用示例

基于情感语义的图像分类问题

图像来自数据库中图片，单词选自情感类别词汇，而主题为隐含情感语义因子（代表底层图像特征和高层情感类别之间的一个中间语义层概念）

利用pLSA模型来学习隐含情感语义因子，从而得到每个隐含情感语义因子在情感单词上的概率分布 $p(s_j|z_k)$ 和每张图片在隐含情感语义因子上的概率分布 $p(z_k|t_i) \Rightarrow$ 训练分类器进行分类

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第37页

## 12.5.3 LDA模型

隐含狄利克雷分配模型是一个集合概率模型，可以看作是在pLSA模型基础上加上超参层，建立隐变量 $z$ 的概率分布而形成的

### 1. 基本LDA模型

一个三层贝叶斯模型（方框表示重复/集合）

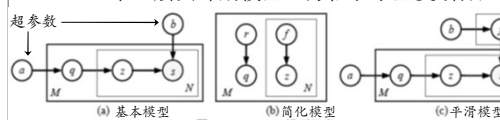


图 12.5.6 LDA模型示意

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第38页

## 12.5.3 LDA模型

### 1. 基本LDA模型

结点 $z$ 为主题结点

结点 $q$ 表示图像中的主题分布， $q$ 也称混合概率参数

图像数  
M

结点 $s$ 是唯一的观测结点（画有阴影）， $s$ 为观察变量

单词数  
N

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

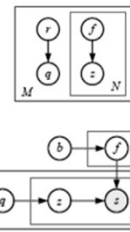
第39页

## 12.5.3 LDA模型

### 1. 基本LDA模型

为消除 $q$ 和 $b$ 之间的耦合关系，可删除图中 $q$ 和 $z$ 之间的连线以及观测结点 $s$ ，得到简化后的模型

一般总将基本的LDA模型扩展为平滑的LDA模型以获得更好的效果（克服大数据集时的稀疏问题）



第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第40页

## 12.5.3 LDA模型

### 2. SLDA模型

引入类别信息，就得到有监督的LDA模型

下部增加了一个与主题 $z$ 相关的类别标记结点 $l$ 对主题 $z$ 的推理受类别标记 $l$ 的影响

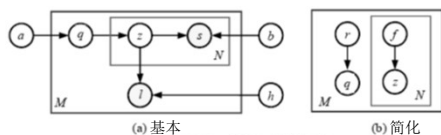


图 12.5.7 SLDA模型示意

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第41页

## 联系信息

- ✉ 通信地址：北京清华大学电子工程系
- ✉ 邮政编码：100084
- ✉ 办公地址：清华大学，罗姆楼，6层305室
- ✉ 办公电话：(010) 62798540
- ✉ 传真号码：(010) 62770317
- ✉ 电子邮件：[zhang-yj@tsinghua.edu.cn](mailto:zhang-yj@tsinghua.edu.cn)
- ✉ 个人主页：[oa.ee.tsinghua.edu.cn/~zhangyujin/](http://oa.ee.tsinghua.edu.cn/~zhangyujin/)

第12讲

章毓晋 (TH-EE-IE) ZHANG YU JIN

第42页