

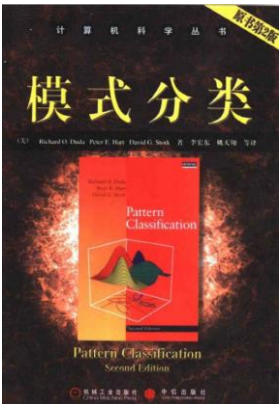
认知心理学基础

第四讲 模式识别

主讲人：刘剑毅
人工智能学院

模式识别的基本概念

- 认知心理学的知觉研究主要涉及模式识别，特别是视觉的模式识别
- 模式 (pattern)：有若干元素或者成分按照一定关系形成的某种刺激结构，也可以说模式是刺激的组合
- 视觉模式：图形、字母、汉字、头像
- 听觉模式：音节、单词、音乐旋律
- 当人能够确认他所知觉的某个模式是什么，并能够将它与其他模式区分开来，这就是模式识别



模式识别的基本过程

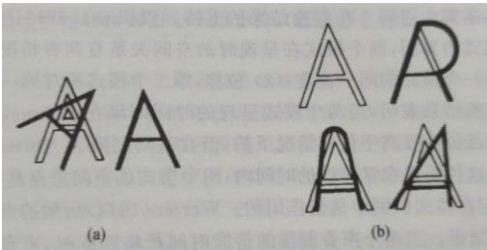
- 通常包含四个过程成分：
 - 对当前知觉对象特征的**知觉表征** ↔ **特征提取**
 - 记忆中与对象有关的表征在视觉系统中的激活
 - 当前对象的知觉表征与激活的记忆**表征的比较** } ↔ **哈希检索**
 - 决策过程——决定哪个长时记忆中的项目与当前对象有着**最佳匹配**
- ↕
距离度量空间

模式识别的理论模型一：模板说

- 模板匹配模型最早是针对机器的模式识别而提出来的，在计算机模式识别中有广泛应用，后来被用来解释人的模式识别
- **模板 (Template)**：在人的长时记忆中，贮存着许多各式各样的过去在生活中形成的外部模式的袖珍复本，称作模板
- **模板匹配 (Template Matching)**：物体识别是基于对物体的知觉输入与记忆中的有关模版进行比较匹配的结果，这种匹配要求**两者有最大程度的重叠** ↔ **例：均方误差最小化** $\min_w \sum_i (f(x_i) - y_i)^2 + \lambda \|w\|^2$
- Visual templates are specified by the conjunction of the set of receptors on which the image of the target shape would fall

模版匹配理论的困难

- 具体性问题：如果模式在外形、大小、方位等某一方面有所变化，即识别不了
- 预处理必不可少，如Face alignment
- 需要在人的记忆中**贮存不可计数的模板**，增加了记忆负担，且与人的模式识别的高度灵活性不一致



初值敏感性

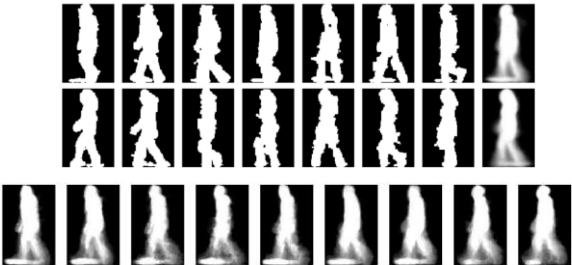
陷入局部极小

模版匹配方法的直接应用

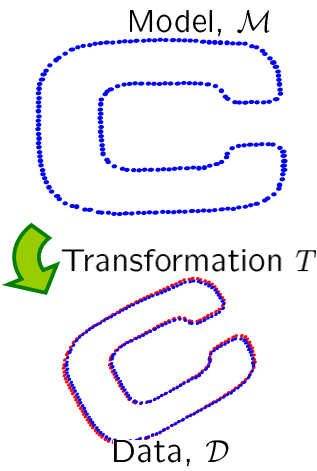
条形码、二维码识别 (QR)



行人步态识别



模板匹配之难：以点云配准为例



Input:

Two point sets
 $\mathcal{M} = \{\mathbf{M}_i\}$ and $\mathcal{D} = \{\mathbf{D}_j\}$

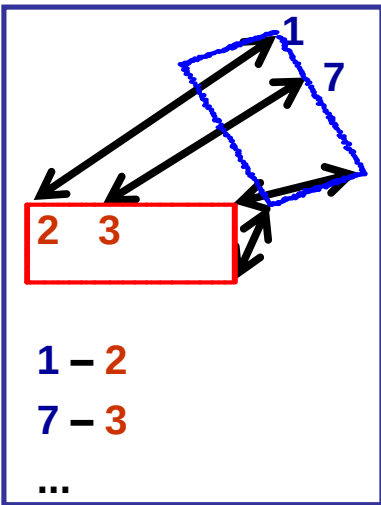
Assumption:

$\mathcal{D}$ is obtained by subjecting $\mathcal{M}$ to a transformation T , and measuring with error

Task:

Determine T

Known Correspondences



Hard:

$$\epsilon(T) = \sum_j \min_i d(T * \mathbf{M}_i, \mathbf{D}_j)$$

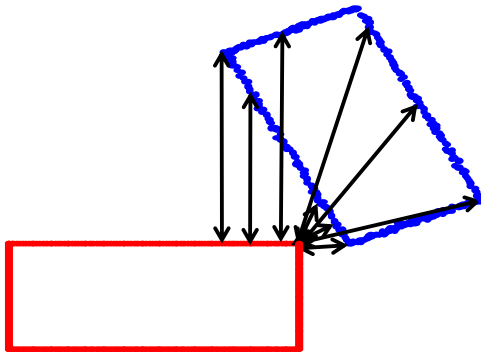
Easy:

Given *correspondences* $j \leftrightarrow \phi(j)$

Can minimize

$$\sum_j d(T * \mathbf{M}_{\phi(j)}, \mathbf{D}_j)$$

But we do not know the correspondence



- That's OK, just choose the closest point...
- *Of course* it's wrong, but it will get us closer

ICP optimization algorithm

EM-like version: task is to minimize over T and ϕ

$$-\log P(T; \phi) = \sum_j d(T * \mathbf{M}_{\phi(j)}, \mathbf{D}_j)$$

Fix T , compute $\{\phi(j)\}_{j=1}^n$:

$$\phi(j) = \operatorname{argmin}_i d(T * \mathbf{M}_i, \mathbf{D}_j)$$

Fix $\{\phi(j)\}_{j=1}^n$, compute T

$$T = \operatorname{argmin}_T \sum_j d(T * \mathbf{M}_{\phi(j)}, \mathbf{D}_j)$$

模式识别的理论模型二：原型说

■ 原型 (prototype)

不是某个特定模式的内部复本，而是一类客体的内部表征，即一个类别或范畴的所有个体的概括表征，反映一类客体具有的基本特征

■ 原型说的基本观点

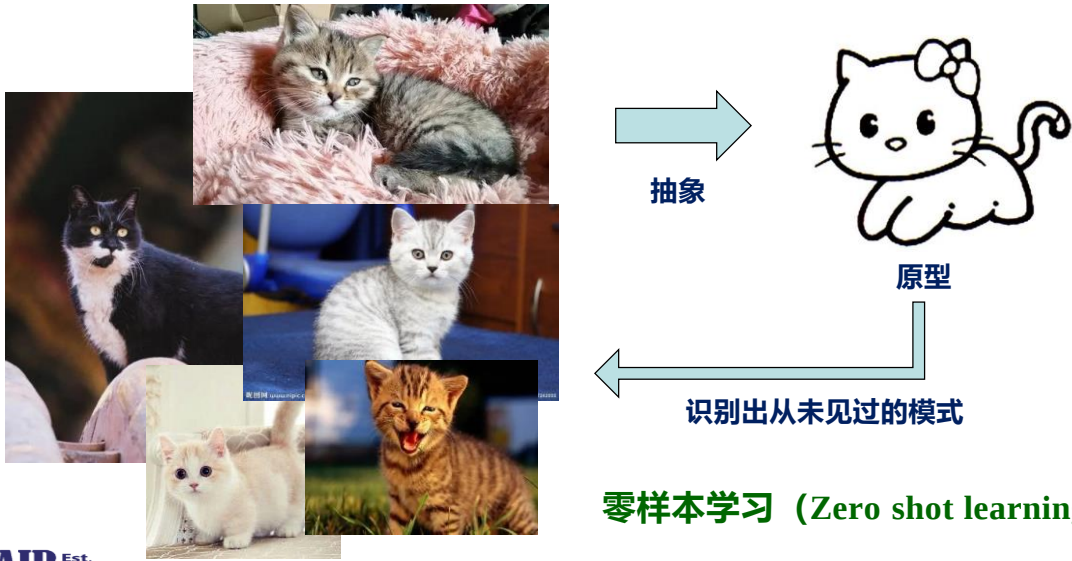
在模式识别中，外部刺激只需要与原型进行比较，由于原型是一种概括表征，因此不要求严格匹配，只需近似的匹配即可。当刺激与某一原型有最近似的匹配，即可将该刺激纳入此原型所代表的范畴，从而得到识别

原型说

■ 人类看到各种不同外形的物体，即使从未见过，也能对它进行正确归类

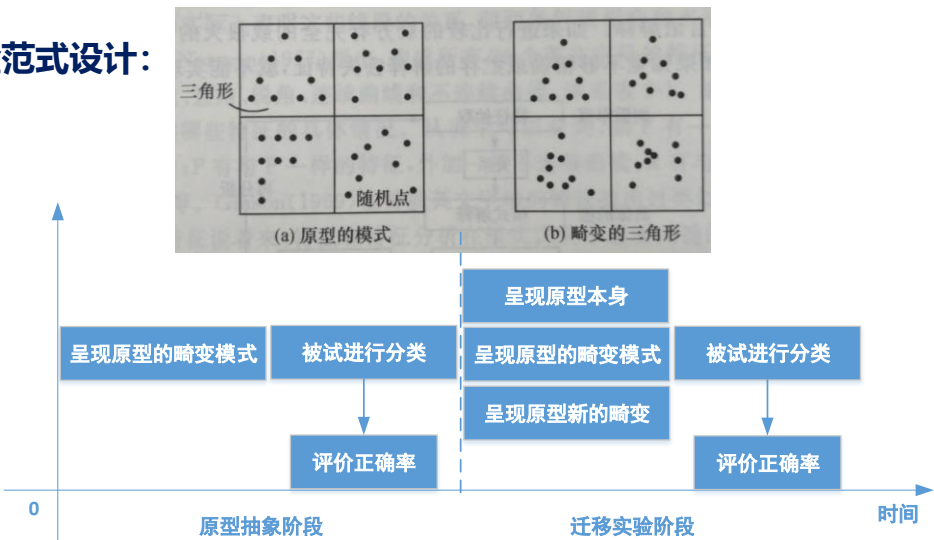


原型说



Posner (1967) 的视知觉实验

■ 实验范式设计:



Posner (1967) 的视知觉实验

■ 实验结果:

- 第一阶段实验, 分类正确率87%
- 第二阶段实验, 原型的分类正确率也为87%
- 第二阶段实验, 新畸变的分类正确率低于87%

■ 实验得到的推断:

- 人能够从一些畸变模式中有效抽象出原型
- 人对畸变模式的分类是在脑海中与原型做比较的

■ 实验结论:

- 验证了原型假说

原型说 v.s. 模板匹配说

- 原型说允许近似匹配, 即使某一范畴的个体之间存在外形、大小等方面的差异, 所有这些个体都可与原型近似匹配而得到识别
- 原型说使人的模式识别更加灵活, 更能适应环境, 新的不熟悉的模式也可以识别
- 原型说可以大大减轻记忆负担

关于原型形成的两大理论

■ Central-tendency theory(趋中理论, Posner et al., 1967)

A prototype is conceptualized as representing the average or mean of a set of exemplars

例子：平均脸假说

从上个世纪70年代起开始出现对人脸的“美”的心理学研究以来，短短几十年内，围绕着平均脸与美，先后出现了不少假说，最重要的一个就是“**平均即是美**”(Averageness is Attractiveness)。德国雷根斯堡和罗斯托克大学的研究人员通过几年的研究和实验，认为这一结论对男性和女性的人脸同样成立



Chinese

关于原型形成的两大理论

■ Attribute-frequency theory(特征-频率理论, Solso and McCarthy, 1981)

A prototype represents the mode of **most frequently experienced** combination of attributes, which is synonymous with the “best example” of a set of patterns

模式识别的理论模型三：特征说

■ 特征 (feature)

模式是由若干元素或成分按一定关系构成的，这些元素或成分可称为特征。如26个英文字母，可由7个特征组成，即垂直线、水平线、斜线、直角、锐角、连续曲线、不连续曲线

表 2-1 英文字母的 7 种特征及其分布情况

	垂直线	水平线	斜线	直角	锐角	连续曲线	非连续曲线
A	1	1	2		3		2
B	1	3		4			1
C				2			1
D	1	2		2			
E	1	3		4			
F	1	2		3			
G	1	1		1			1
H	2	1		4			
I	1	2		4			
J	1						1
K	1		2	1	2		
L	1	1		1			
M	2		2		3		
N	2		1		2		
O						1	
P	1	2		3			1
Q			1		2	1	
R	1	2	1	3			1
S							2
T	1	1		2			
U	2						1
V			2		1		
W			4		3		
X			2		2		
Y	1		2		1		
Z		2	1		2		

(引自 Lindsay & Norman, 1977)

特征说

- 外部刺激在人的长时记忆中是由特征符号的集合来表征的
- 模式识别首先要对刺激的特征进行分析，即抽取刺激的有关特征，然后将这些抽取的特征加以合并，再与长时记忆中的有关特征符号集合进行比较，一旦获得了最佳的匹配，外部刺激就被识别了
- 与模板匹配模型的关系：特征可以看做是一种局部的部件模板
- 但特征分析模型较模板匹配模型有如下优点：
 - 更强的适应性，依据刺激的特征和关系进行识别，不管刺激的大小、多少等细节
 - 同样的特征可以出现在不同的模式中，极大地减轻了记忆的负担

字母搜索实验 (Neisser, 1964)

- 实验设计:
 - 要求被试尽快搜寻到靶子字母
- 实验结果:
 - (a) 比 (b) 中的搜索快得多
- 实验结论:
 - 相似性特征会干扰搜索过程, 从而验证了特征说
 - 反驳了模板匹配说

- Stimulus for Neisser's (1964) visual search experiment
- Find the 'Z' in (a) and (b)

ODUGQR	IVMXEW
QCDUGO	EWVMIX
CQOGRD	EXWMVI
QUGCDR	IXEMWV
URDGGO	VXWEMI
GRUQDO	MXVEWI
DUZGRO	XVWMEI
UCGRGD	MWXVIE
DQRCGU	VIMEXW
QDOCGR	EXVWIM
CGUROQ	VWMIEX
OCURQO	VMWIEI
UOCGQD	XVWMEI
RGQCOU	WXVEMI
GRUDGO	XMEWIV
GODUCQ	MXIVEW
QCURDO	VEWMIX
DUCOQG	EMVXWI
CGRDQU	IVWMEX
UDRCOQ	IEVMWX
GQCORU	WVZMXE
GOQUCD	XEMIWV
GDQUOC	WXIMEV
URDCGO	EMWIVX
GODRQC	IVEMXW

(a)

(b)

静止网像实验 (Pritchard, 1961)

- 背景知识
 - 眼动是人类实现客体识别的重要前提 (30-70Hz) , 如果客体严格投影到视网膜同一位置, 客体知觉将会消失
 - 青蛙没有眼动, 因此只能发现动态物体, 而对静态物体视而不见



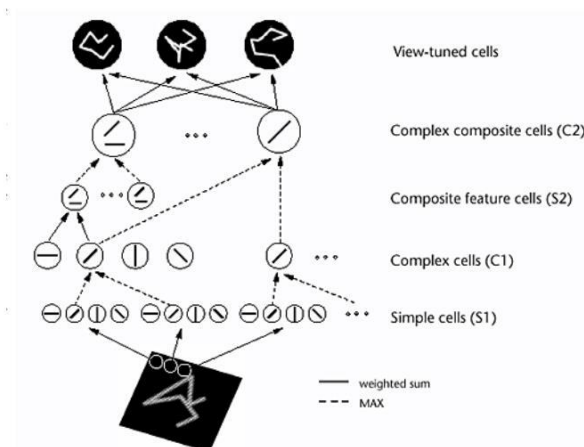
静止网像实验 (Pritchard, 1961)

- **实验手段:**
- 采用技术手段使客体在眼睛运动条件下, 在视网膜的成像位置不变, 即抵消掉眼动的影响
- 被试报告暂时保有的图像片段 (右图)
- **实验发现:**
- 客体知觉并非立即消失, 而是一部分一部分渐次消失
- 消失的均为一些完整特征
- **实验推断:**
- 特征是知觉的重要单元



Computational Model of Object Recognition

(Riesenhuber and Poggio, 1999)



幼年猫的视觉皮层实验 (Hubel, Wiesel, 1963)

■ 实验手段:

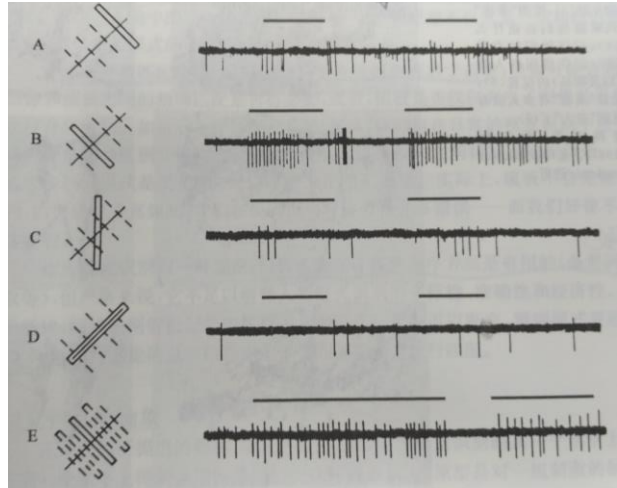
- 采用微电极植入猫的视觉皮层，把不同的光条刺激呈现在动物眼睛正前方屏幕上，记录单个神经细胞的放电

■ 实验发现:

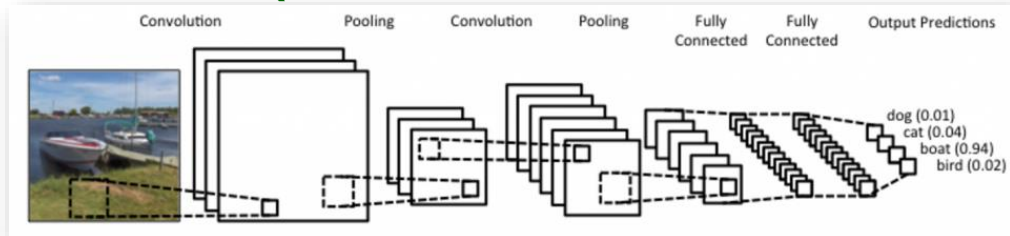
- 某些细胞仅对水平形状有反应，而另一些则仅对垂直形状有反应

■ 实验结论:

- 因幼年猫缺乏视觉经验，因此知觉形状的大脑编码是先天的



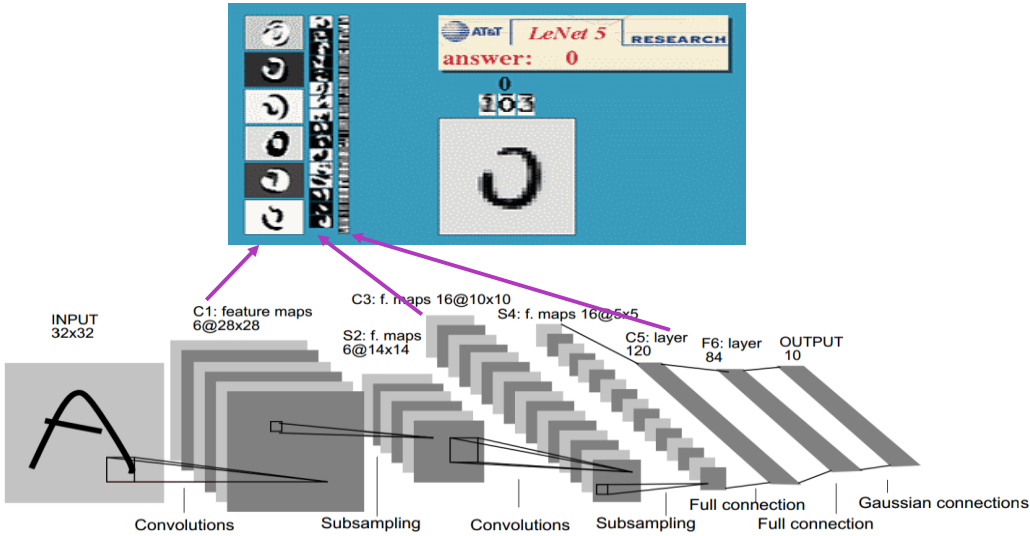
卷积神经网络 (Convolutional Neural Network)



- 20世纪60年代，Hubel和Wiesel在研究猫脑皮层中用于局部敏感和方向选择的神经元时发现其独特的网络结构可以有效地降低反馈神经网络的复杂性，提出了卷积神经网络

- Yann LeCun等于1995年受到生物过程启发、改进深度前馈人工神经网络，由输入和输出层以及多个隐藏层组成，其中隐藏层通常包括卷积层、合并层，完全连接层和标准化层

LeNet-5



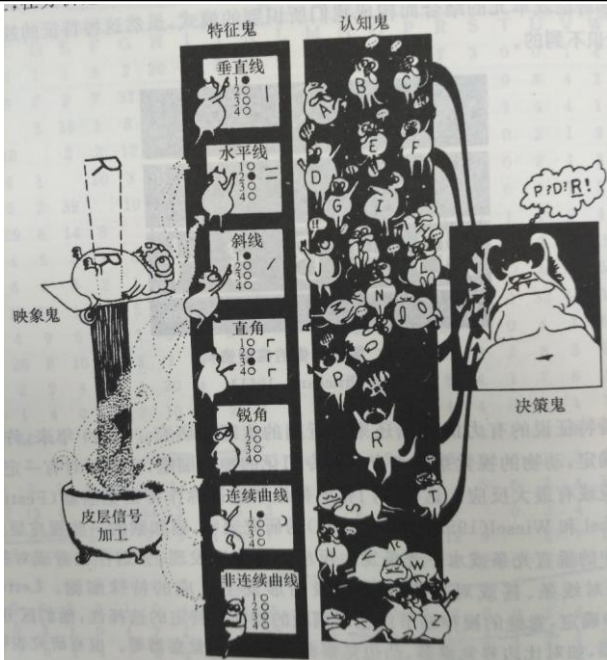
IAIR Est. 1986

27

类比一个著名的特征分析模型

—— Selfridge (1959) 的“鬼城”模型 (Pandemonium)

- 第一层：映象鬼——负责对刺激进行编码
- 第二层：特征鬼——将编码的映象分解为不同特征
- 第三层：认知鬼——统计特征数量决定喊叫声大小
- 第四层：决策鬼——选择喊声最大的认知鬼作为识别的模式

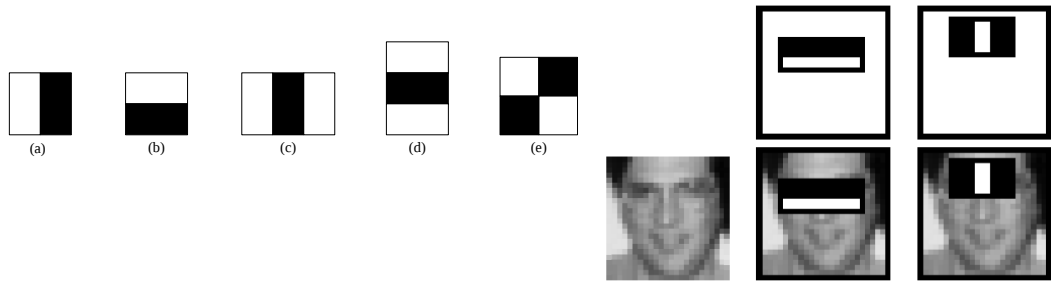


IAIR Est. 1986

2024/5/27

28

基于Haar-like特征的人脸检测算法



Haar-like特征

The first and second features selected by AdaBoost

- Given example images $(x_1, y_1), \dots, (x_n, y_n)$ where $y_i = 0, 1$ for negative and positive examples respectively.
- Initialize weights $w_{1,i} = \frac{1}{2m}, \frac{1}{2l}$ for $y_i = 0, 1$ respectively, where m and l are the number of negatives and positives respectively.
- For $t = 1, \dots, T$:

1. Normalize the weights,

$$w_{t,i} \leftarrow \frac{w_{t,i}}{\sum_{j=1}^n w_{t,j}}$$

so that w_t is a probability distribution.

2. For each feature, j , train a classifier h_j which is restricted to using a single feature. The error is evaluated with respect to w_t , $\epsilon_j = \sum_i w_i |h_j(x_i) - y_i|$.
3. Choose the classifier, h_t , with the lowest error 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 $\beta_t = \frac{\epsilon_t}{1-\epsilon_t}$.

- The final strong classifier is:

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

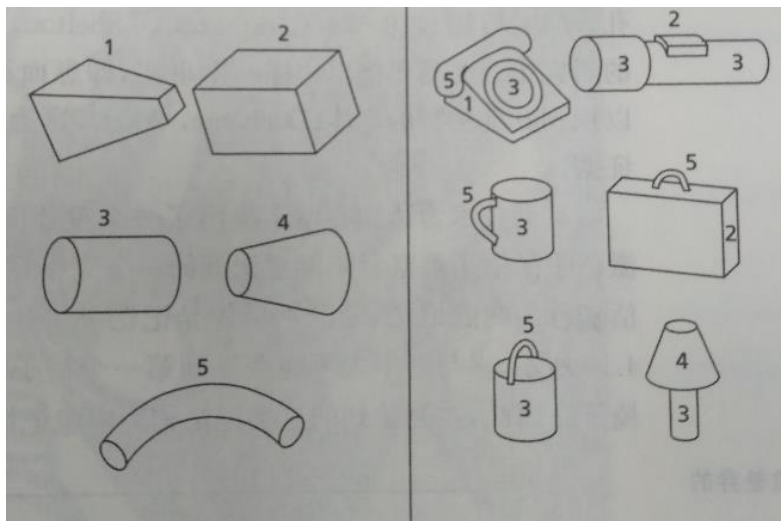
where $\alpha_t = \log \frac{1}{\beta_t}$

模式识别的理论模型四：成分说

■ 成分说又称作成分识别理论或几何子理论，Recognition by Components Theory (Biederman, 1987)

- Perception based on the detection of combinations of simple 3D “geometrical ions” (**geons** 几何子)
- Geons defined by **view invariant properties**: properties that are maintained even when viewed from different angles
- 几何子 (geons): **24 distinct forms**, like the letters of the alphabet, make up a type of systems

几何子及其客体表征的示例



儿时记忆中的“几何子”

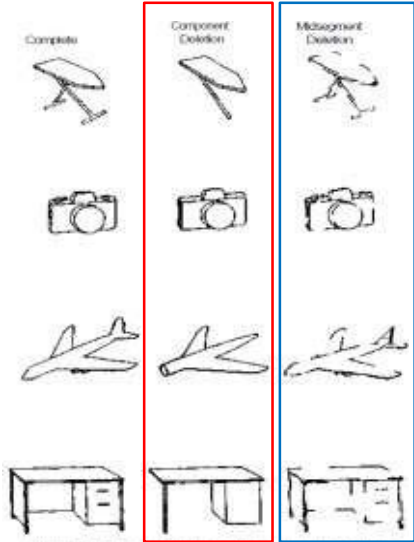


IAIR Est. 1986

2024/5/27

33

Biederman (1985) 经典实验



删除成分

删除线条

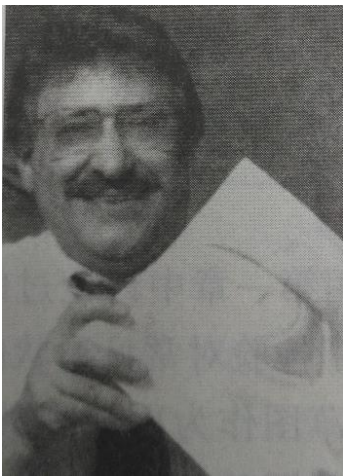


Figure 2.2. Sample stimuli, with equivalent proportion of contours removed either at subsegments or as whole components.

IAIR Est. 1986

34

Biederman (1985) 经典实验

- **实验范式:**
 - 向被试呈现65ms-200ms时长不等的视觉刺激，包括删除线条及删除成分，记录其给出识别结果的时长及准确度
- **实验发现:**
 - 当呈现时长200ms时，删除线条的识别要比删除成分的快
 - 删除线条的平均识别率约为70%，删除成分的约为50%
- **实验结论:**
 - 客体识别的基础是基本形状，验证了几何子理论

模式识别的理论模型小结

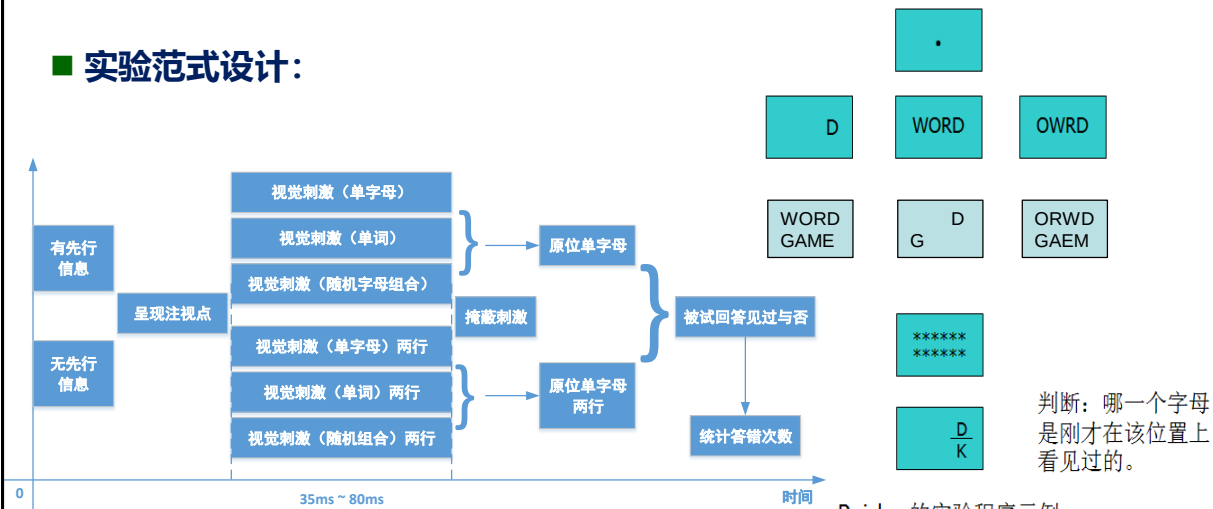
- **模板说 (模板匹配模型) ——自下而上模型**
- **原型说 (原型匹配模型) ——自下而上+自上而下模型**
- **特征说 (特征分析模型) ——自下而上模型**
- **成分说 (几何子理论) ——自下而上+自上而下模型**

结构优势效应

- **结构优势效应：**
又被称为**上下文效应**，指整体的结构在模式识别中所起到的重要作用
- 过去的知识经验对知觉具有重要影响，表现为一个模式常常不是孤立地出现，而是处于与其他模式的相互联系之中
- 结构优势效应与**建构主义**、**概念驱动**、**自上而下的加工**、**整体加工**具有密切的关联，已引起心理学家的广泛关注

字词优势效应实验（Reicher, 1969）

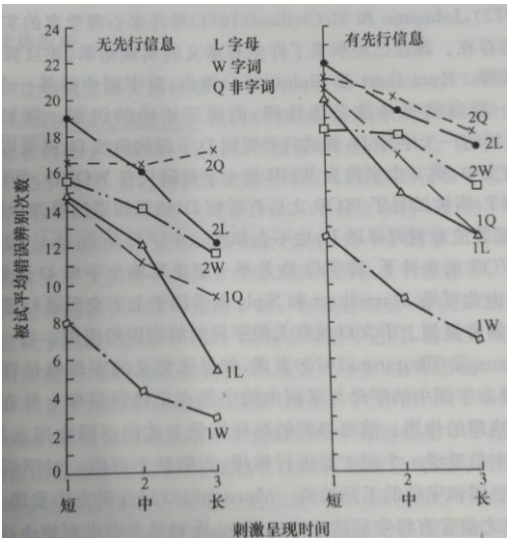
■ 实验范式设计：



字词优势效应实验

实验结果:

- 识别字词中的一个字母要优于单个字母或非字词中的字母，正确率分别高出8%
- 单行错误率低于两行错误率
- 延长视觉刺激时间可以降低错误率
- 以上差异均达到显著水平

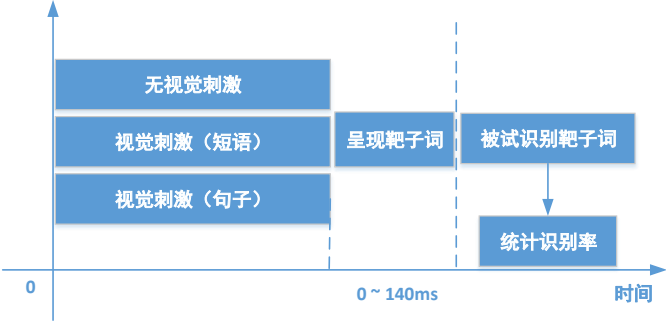


字词优势效应

- 字词优势效应可能的解释
 - 推论说：强调过去知识经验为基础的自上而下的加工，可以利用字词的缀字规则进行上下文推理
 - 编码说：以信息的内部表征为核心，强调字词的语音编码属性，而单个字母是视觉编码
 - 加工说：字词可以看作整体，字词先得到整体加工，然后再有局部加工，而单个字母只有局部加工

句子优势效应实验

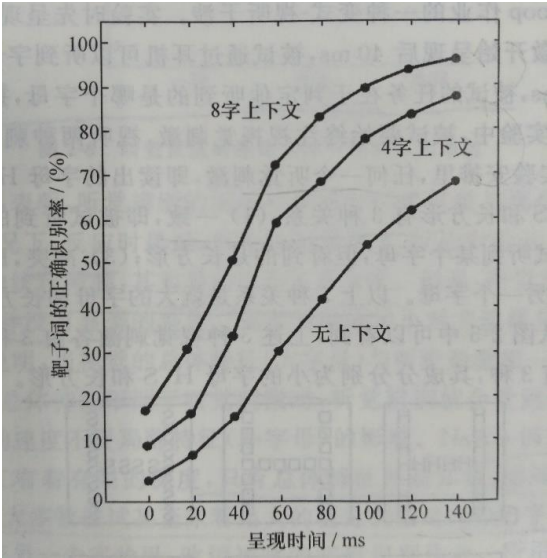
- Tulving、Mandler、Baumal (1964) 的视知觉实验
- 实验范式设计:



- 靶子词范例: disorder
- 刺激范例:
 - (a) 短语: filled with dirt and disorder
 - (b) 句子: The huge slum filled with dirt and disorder

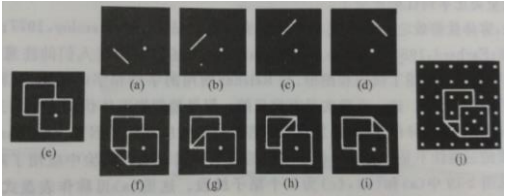
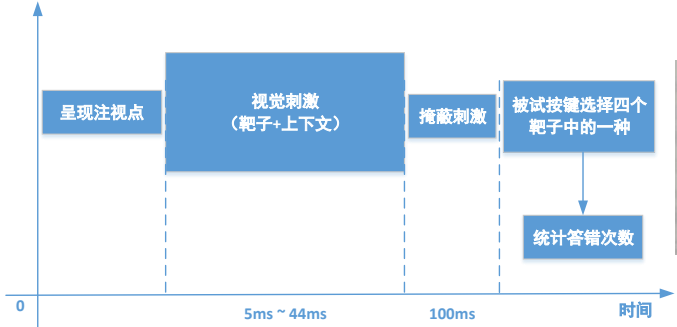
句子优势效应实验

- 实验结果
- 有上下文的靶子词识别率均高于无上下文的情况——验证了top down
- 较长的靶子词呈现时间在各种情况下均能够提高识别率——验证了bottom up
- 实验结论
- 字词识别既依赖于自下而上加工, 也依赖于自上而下加工



客体优势效应实验 (Weisstein, 1974)

- 实验目的：研究不同图形上下文中的直线线段识别能力
- 实验范式设计：



客体优势效应实验

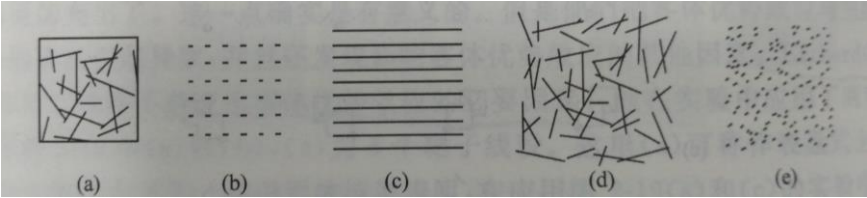
- 实验结果
- 当上下文图形更像一个三维客体时，识别率更高
- 有上下文时的正确识别率比无上下文的情况要高10%，差异达到显著水平
- 实验结论
- 在一个结构严谨的三维图形中的线段识别起来要优于组织较差的图形中的同一线段

被试数目	差别
(a) 9、6	—
(b) 6	-0.6%
(c) 6	-2.3%
(d) 9	-5.2%
(e) 6	-5.6%
(f) 9	-19.2%

客体优势效应实验

Klein (1978) 关于掩蔽刺激作用的实验

- 实验结果的解释
- 存在两种知觉系统：特征系统——自底向上；客体系统——可被结构严谨的整体刺激激活
- 某些随机掩蔽与上下文有共同成分，故破坏了客体系统的内部表征



出现客体优势效应：

未出现客体优势效应：

其他结构优势效应实验

- 构型优势效应实验 (Pomerantz, 1977)
- 实验发现：
对于 “ () ” 和 “)) ” 视觉刺激的识别速度要显著快于对 “ (” 和 “) ” 的识别速度
- 字母优势效应实验 (Schendel, Shaw, 1976)
- 实验发现：
字母 “H” 和 “N” 的辨别速度和精确性均高于 “-” 和 “\”

结构优势效应小结

- 结构优势效应包括：
 - 字词优势效应
 - 句子优势效应
 - 客体 (object) 优势效应
 - 构型优势效应
 - 字母优势效应
- 以整体加工和局部加工为核心的假设，可以为上述各类结构优势效应给出统一的解释
- 视觉以外的其他感觉通道可能也存在着类似的效应
- 模式识别任务以外的其他的知觉活动可能也存在着类似的效应