

人工神经网络&深度学习的 跨世纪探索（上）

刘昕 PhD Candidate @ ICT
2015年7月27日

大纲

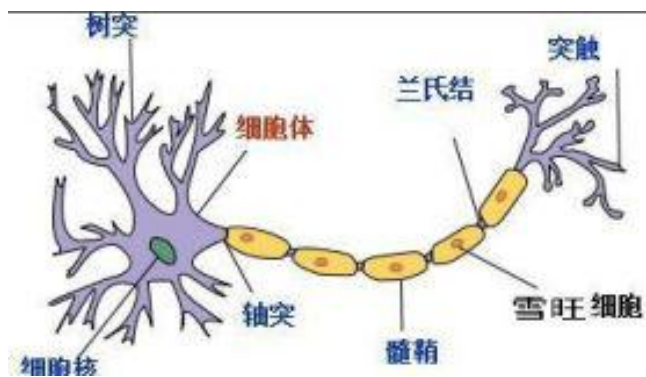
- 视觉神经生物学基础
- 计算机视觉的理论计算模型
- 人工神经网络：从感知机到多层感知机
- Lecun的成就

大纲

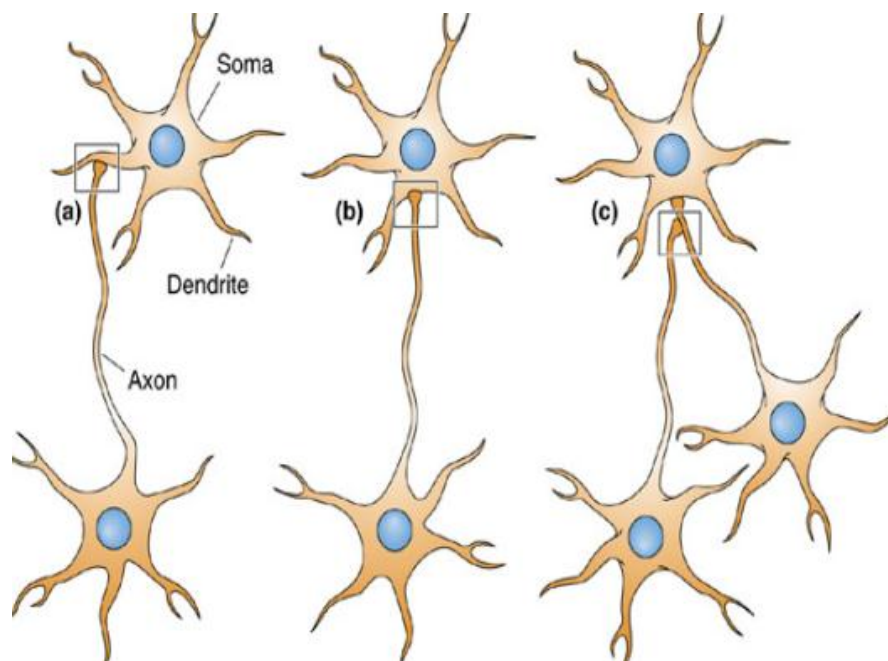
- —>视觉神经科学基础
- 计算机视觉的理论计算模型
- 人工神经网络：从感知机到多层感知机
- Lecun的成就

视觉神经科学基础

■ 生物神经元



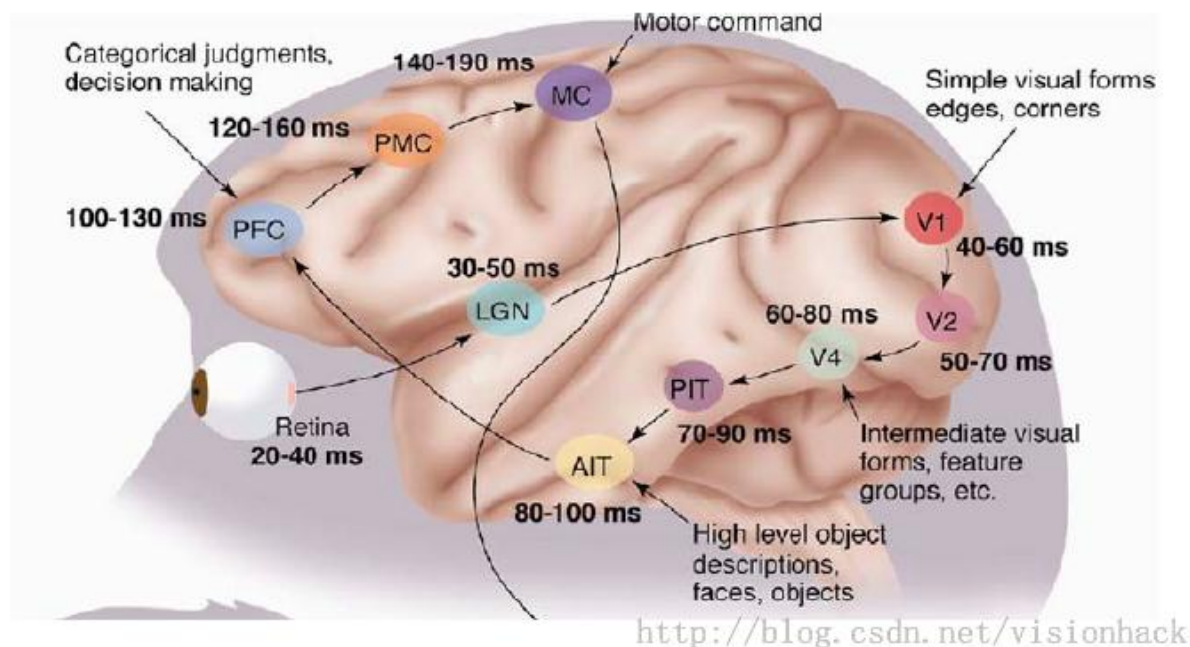
(a)生物神经元



(b)神经元之间的连接与信号传导

视觉神经科学基础

■ 生物视觉模型

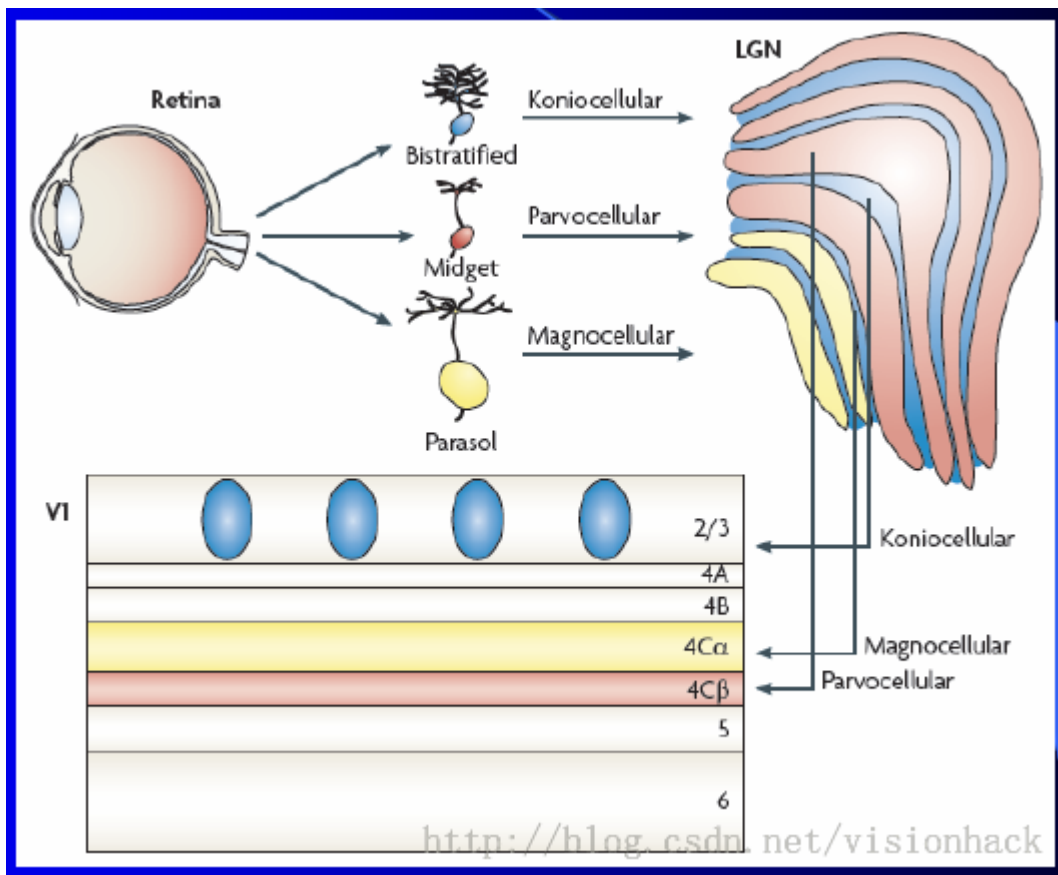


生物视觉的信号处理流程

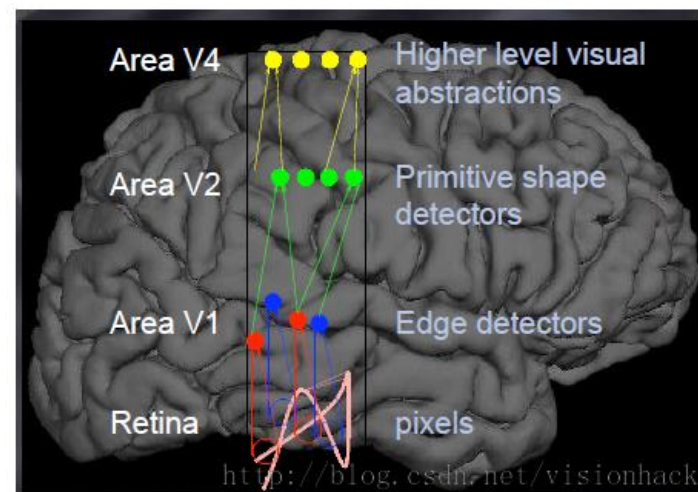
生物视觉信号的传导和解析是一个“分层”的结构！

视觉神经科学基础

■ 生物视觉模型



(a)视觉信号传导: Retina->LGN->Visual Cotex V1



(b)视觉皮层功能区划分

视觉神经科学基础

■ 生物视觉模型

表 1 腹侧视觉皮层通路(含视网膜和 LGN)细胞特性

Table 1 Cellular properties of ventral visual pathway (retina and LGN included)

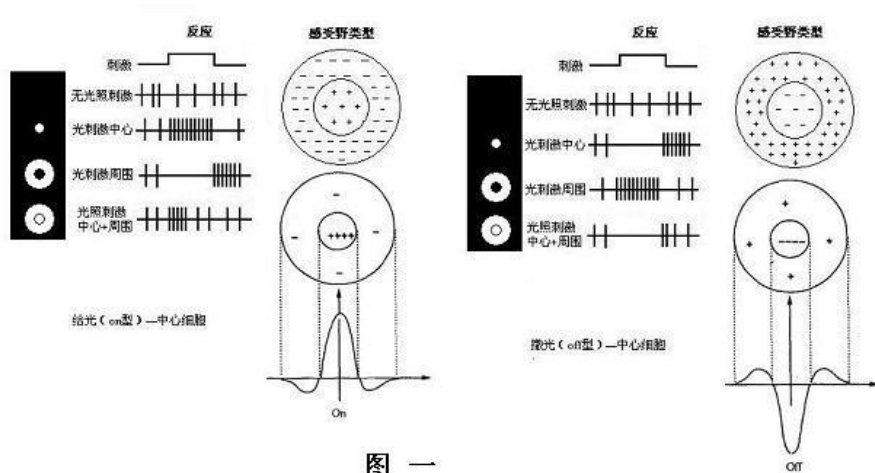
视网膜/ 视区/层	偏好的 刺激特征	感受野的 尺寸	位置严格 性(平移)	朝向选择性 (2D 旋转)	尺度不变性 (缩放)	视角变化不变 性(3D 旋转)	延时/ms	视网膜拓扑
视网膜	同心圆式	极小	严格	无	小	—	0	极好
LGN	同心圆式	极小	严格	无或弱	小	—	无实验数据	极好
VI 简单细胞	边缘/线条	1°	严格	严格	小	—	70	很好
VI 复杂细胞	边缘/线条	1.5°	在感受野 内不严格	严格	小	—	70	很好
V2	边缘/ 线条/角	3.2°	在感受野 内不严格	严格	小	无实验数据	93	好
V4	中等复杂特 征(如角,色 彩,纹理等)	8°	在感受野 内不严格	无实验 数据	较大	无实验数据	116	较好
PII	中等复杂特 征(如角/ 星形等)	20°	在感受野 内不严格	45°	较大	无实验 数据	无实验 数据	差
AIT	视调谐的 复杂特征 (如脸)	50°(位于整 个中央凹)	不严格 (约 4°)	45°	约 2 倍缩放	旋转 32°	150	没有
PFC	类调谐的 复杂特征 (如脸)	很大	不严格	无实验 数据	约 2 倍缩放	视不变	无实验 数据	没有

参考：姚行中等，基于灵长类视觉皮层的目标识别模型综述，模式识别与人工智能，22 (4)，2009

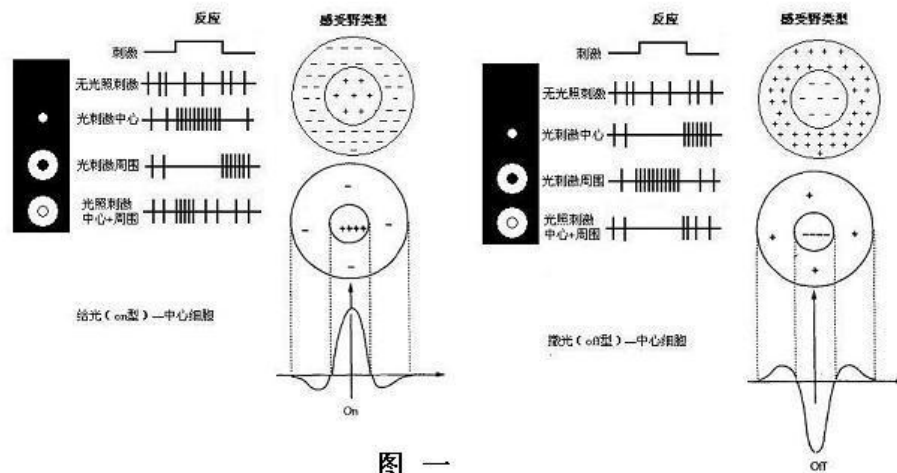
视觉神经科学基础

■ 感受野 (Receptive Field)

Receptive Field: Area in which stimulation leads to response of a particular sensory neuron.



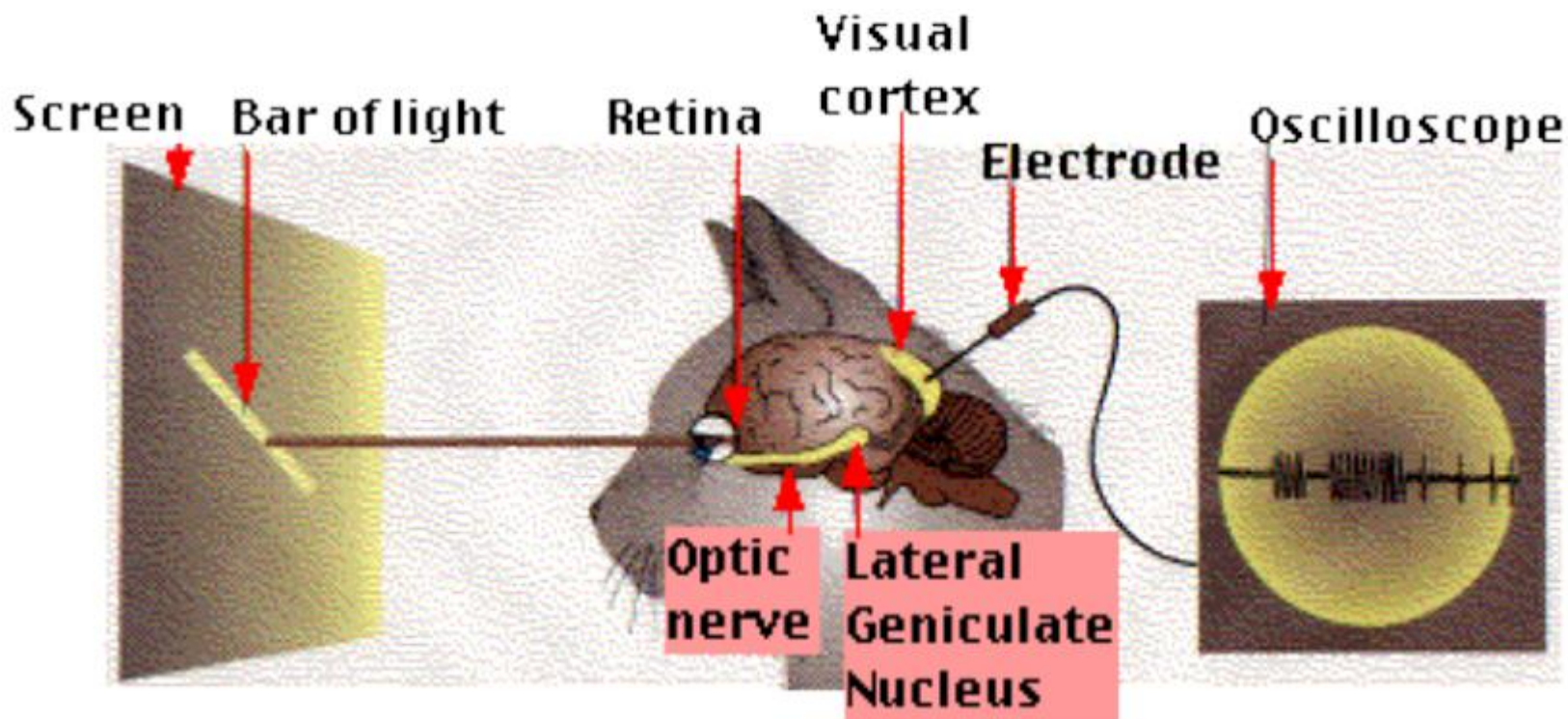
(a) On—型感受野：中心的兴奋区域和周边的抑制区域构成的同心圆结构。



(b) Off—型感受野：由中心抑制和周边兴奋区域的同心圆构成

视觉神经科学基础

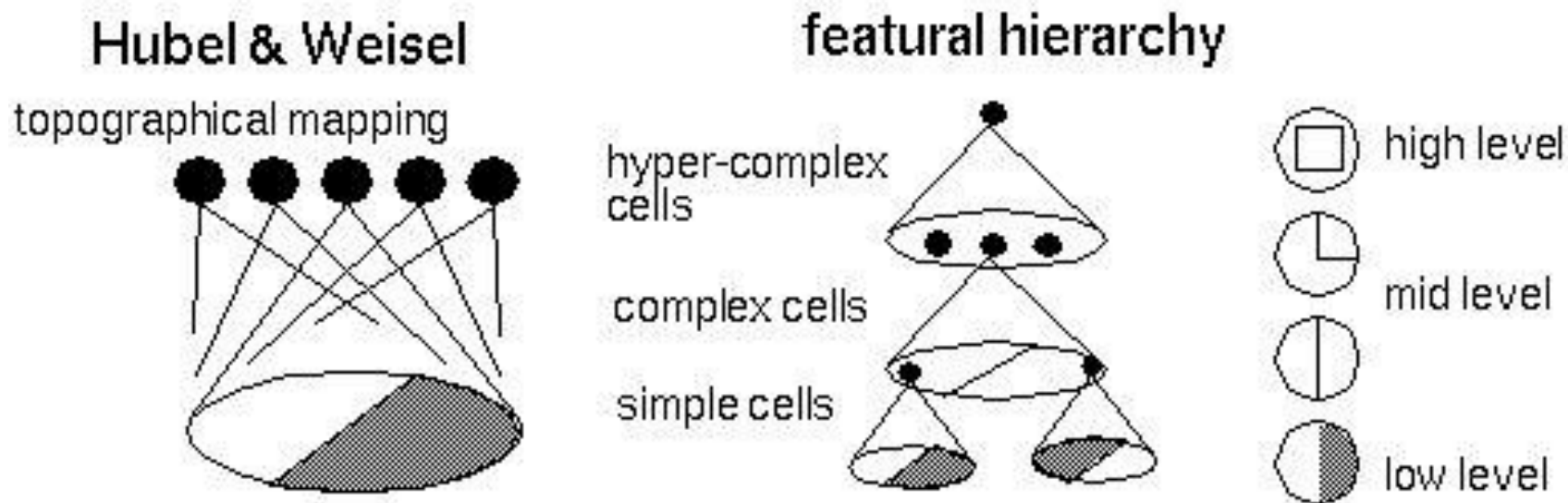
■ Hubel & Wiesel对猫感受野的研究



参考：Hubel, D. H. & Wiesel, T. N. (1962)

视觉神经科学基础

■ Hubel & Wiesel Model (1962)



Hubel & Wiesel对视觉皮层的功能划分

参考：Hubel, D. H. & Wiesel, T. N. (1962, 1965)

视觉神经科学基础

■ Colin Blakemore 1971年 感受野动物实验

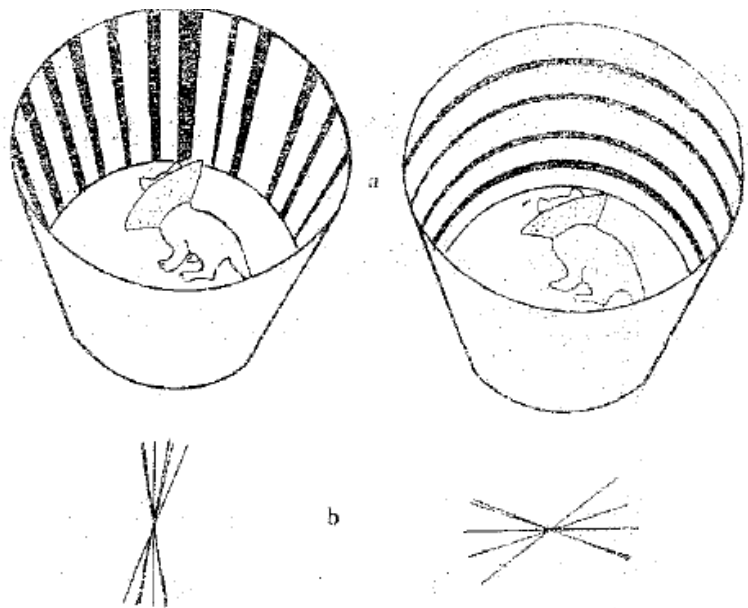


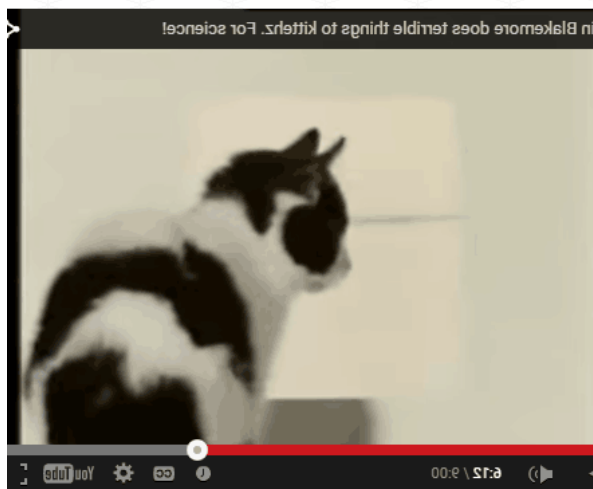
图9 小猫的视觉环境 (a) 及其皮层感受野的最优朝向分布 (b)

参考: <https://youtu.be/QzkMo45pcUo>, 汪云九(1972)

视觉神经科学基础

■ Colin Blakemore 1971年 感受野生动物实验(续)

实验设置：猫从出生就只生活在水平黑白条纹的环境下



(a)对水平条纹运动敏感



(b)对垂直条纹运动不敏感

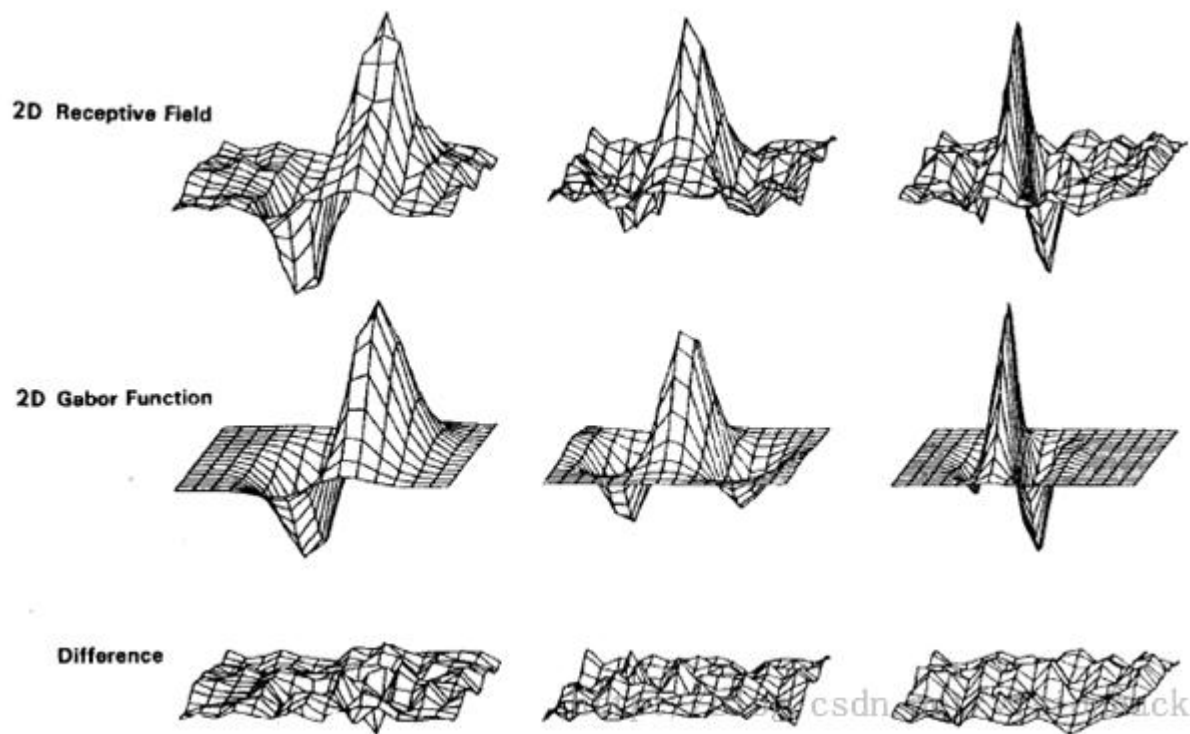


(c)无法感知复杂运动

实验启示：视觉细胞的感受野是后天形成的。

视觉神经科学基础

- 视觉皮层简单细胞（Simple Cell）感受野与Gabor小波



视觉神经科学基础

■ 对深度卷积网络单个神经元感受野的研究

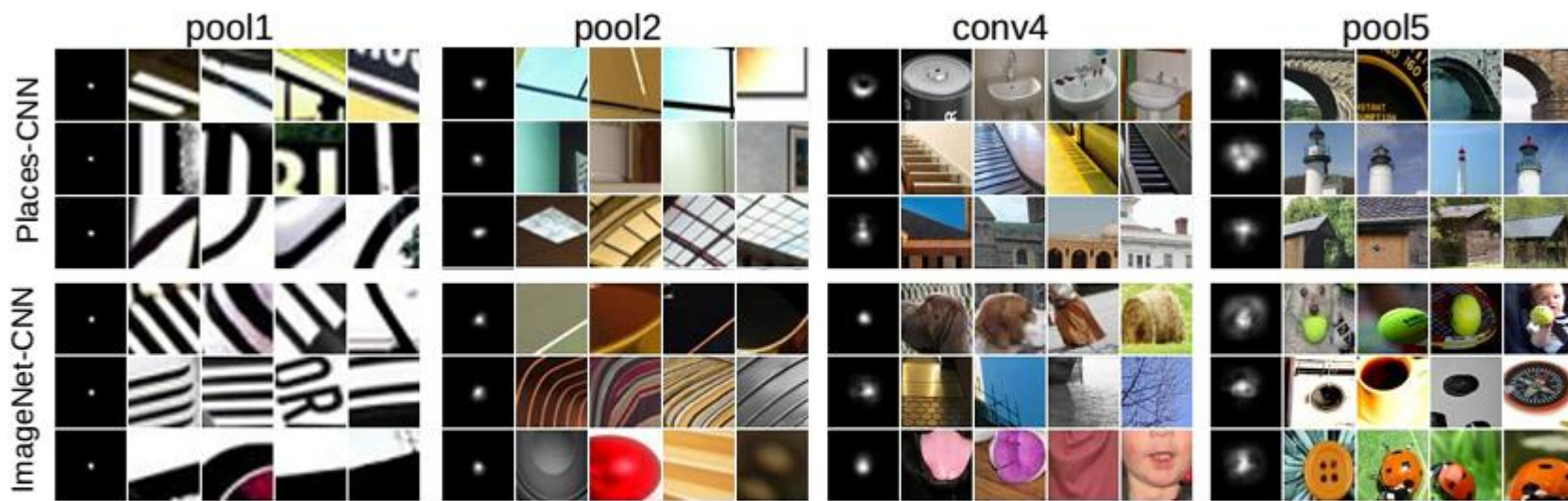


Figure 4: The RFs of 3 units of pool1, pool2, conv4, and pool5 layers respectively for ImageNet- and Places-CNNs, along with the image patches corresponding to the top activation regions inside the RFs.

参考：Bolei Zhou et al, Object Detectors Emerge in Deep Scene CNNs, ICLR 2015

视觉神经科学基础

■ 小结

生物的视觉信息处理是一个层次结构。

感受野是视觉细胞的天然属性，但感受野的结构受后天环境影响。

初级视觉皮层简单细胞的感受野类似于某些Gabor小波。

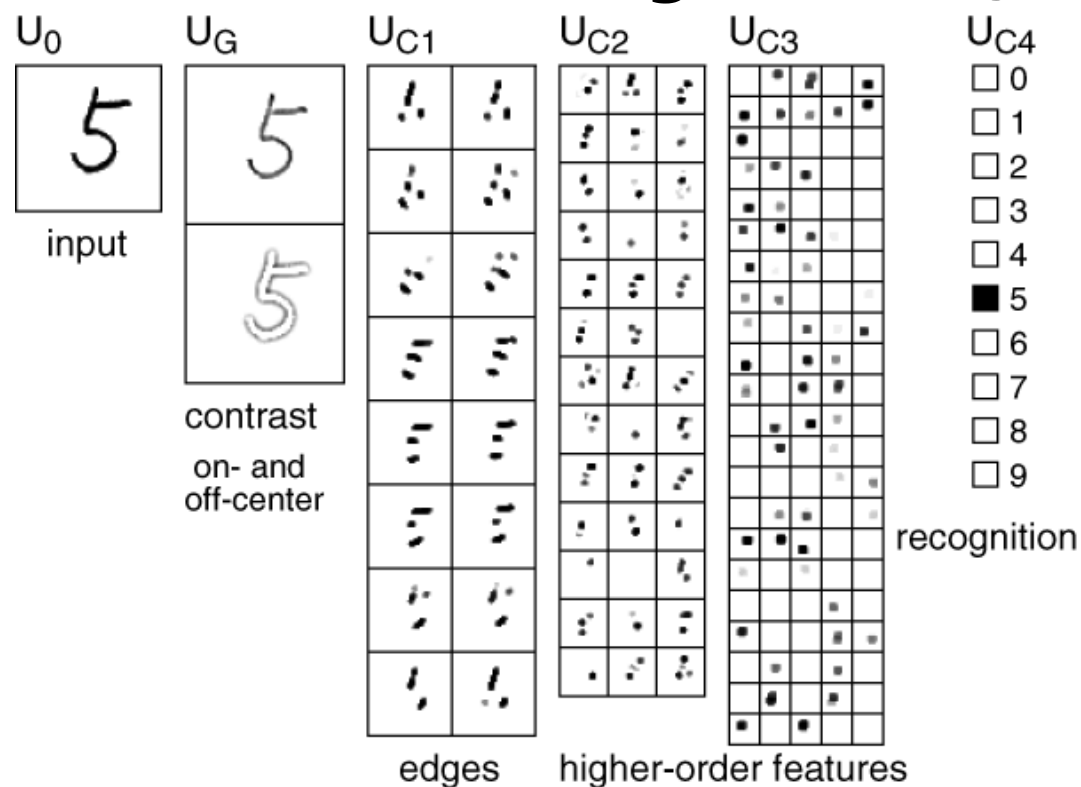
深度网络各层神经元的感受野体现了语义的层级结构。

大纲

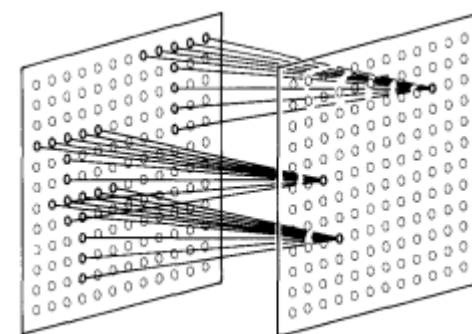
- 视觉神经科学基础
- —> 计算机视觉的理论计算模型
- 人工神经网络：从感知机到多层感知机
- Lecun的成就

计算机视觉理论模型

■ 邦彦福岛 Neocognitron 1980



(a) Handwritten digit recognition by a neocognitron.



(b) Local Connection

参考：K. Fukushima: "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", *Biological Cybernetics*, 36[4] (April 1980).

计算机视觉理论模型

■ 邦彦福岛Neocognitron 1980

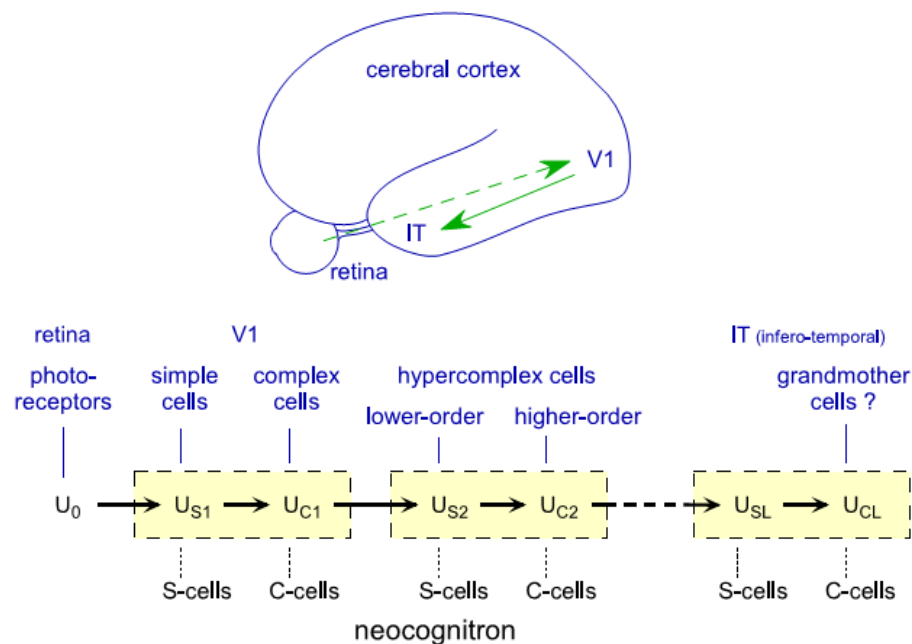
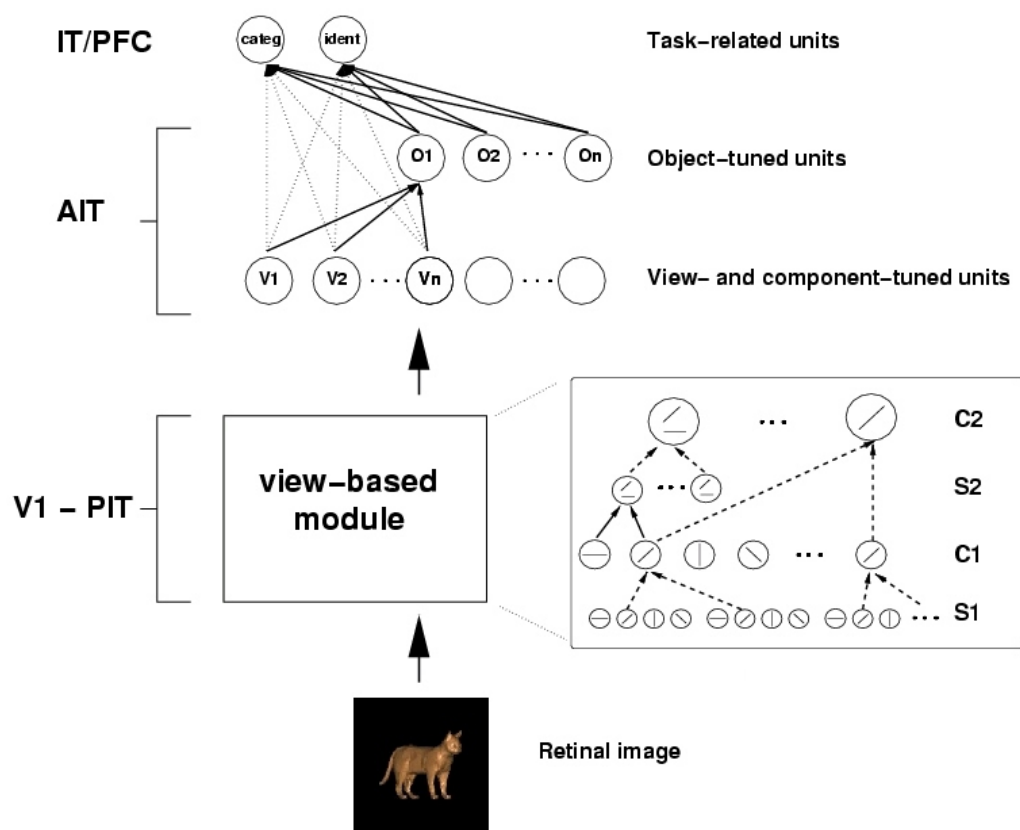


Fig. 1. Relation between the architecture of the neocognitron and the classical hypothesis of Hubel and Wiesel.
Source: (modified from Fukushima (1980)).

参考：K. Fukushima: "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", *Biological Cybernetics*, 36[4] (April 1980).

计算机视觉理论模型

■ Poggio H-max 1999

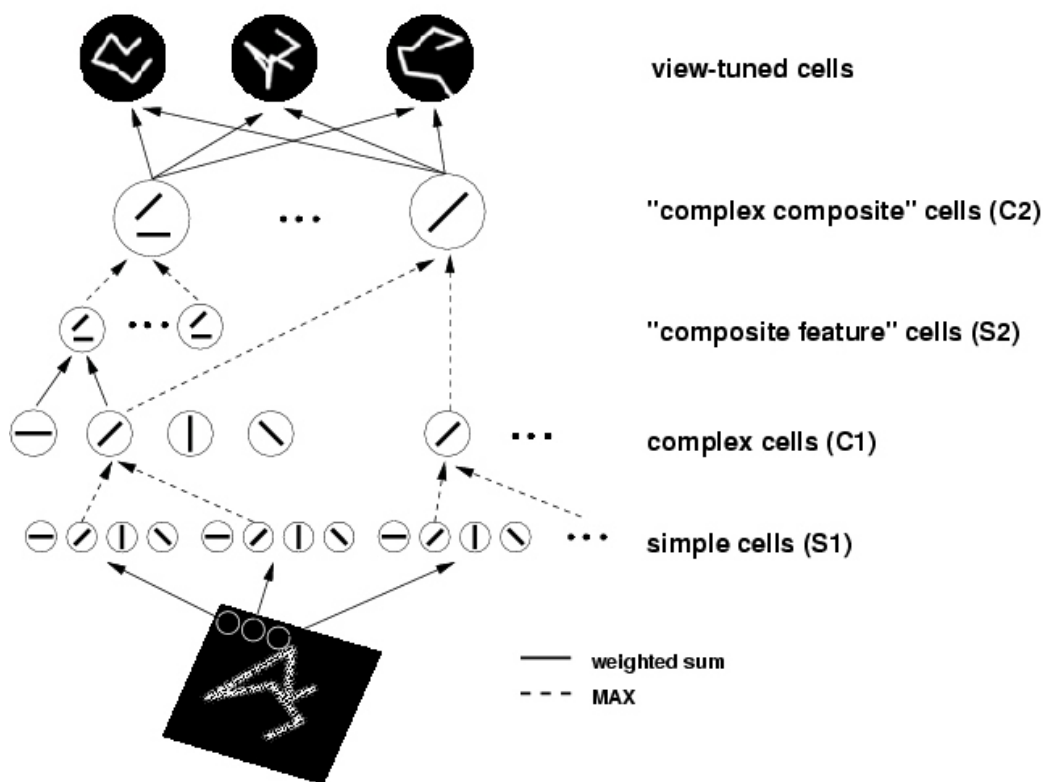


层级概念
简单到复杂
Pooling获得不变性
高层Units的任务相关性

参考：Riesenhuber, M. & Poggio, T. (1999). [Hierarchical Models of Object Recognition in Cortex](#). *Nature Neuroscience* 2: 1019-1025.

计算机视觉理论模型

■ Poggio H-max 1999



S2就是S1特征模式的组合

C1可以用max pooling模拟

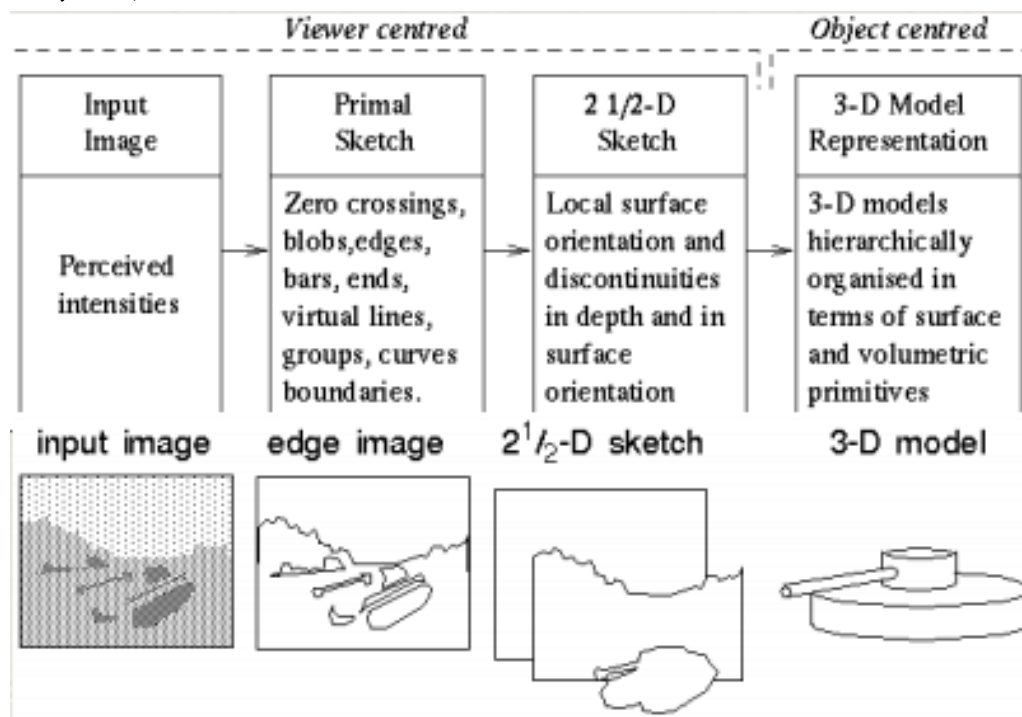
S1可以用Gabor filter模拟

参考： Riesenhuber, M. & Poggio, T. (1999). [Hierarchical Models of Object Recognition in Cortex](#). *Nature Neuroscience* 2: 1019-1025.

计算机视觉理论模型

■ Marr计算机视觉理论 1982

Marr计算机视觉理论将视觉理解为一个信息处理的任务，目的是从2D图像还原3D表示。



参考：David Marr, Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, 1982

计算机视觉理论模型

■ Marr计算机视觉理论 1982

名称	目的	基元
图象	表达场景的辉度或物体的照度	像素（值）
基素图	表达图象中亮度变化位置、物体轮廓的几何分布和组织结构	零交叉、端点、边缘段、边界等
2.5-D图	在以观察者为中心的坐标系中表达物体可见表面的取向、深度、边界等性质	局部表面朝向（“针”基元）、表面朝向的不连续点、深度、深度上不连续点
3-D图	在以物体为中心的坐标系中，用体元或面元集合描述形状和形状的空间组织形式	3-D模型，以轴线为骨架，将体元或面元附在轴线上

参考：David Marr, Vision: A Computational Investigation into the Human Representation and Processing of Visual Information, 1982

计算机视觉理论模型

■ 小结

Hubel和Wiesel模型是Neocognitron以及H-max的理论基础。

Neocognitron和H-max对视觉信号的分层处理一定程度上“类脑”。

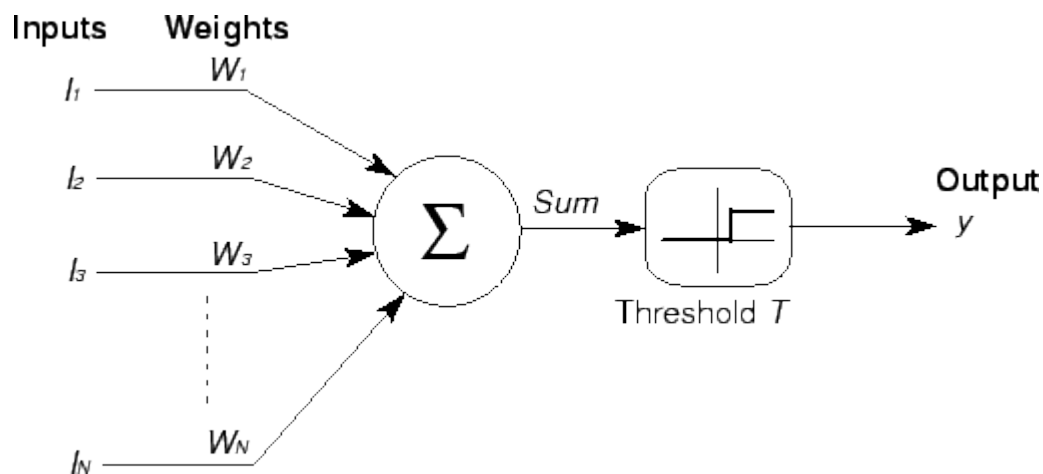
Marr计算机视觉理论强调视觉的本质任务的可计算性。

大纲

- 视觉神经科学基础
- 计算机视觉的理论计算模型
- —>人工神经网络：从感知机到多层感知机
- Lecun的成就

人工神经网络

■ McCulloch-Pitts Model of Neuron 1943



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

MCP人工神经元模型：多个input的sum操作，
激活函数，输出，神经元之间传递

参考： McCulloch, Warren; Walter Pitts (1943). "A Logical Calculus of Ideas Immanent in Nervous Activity". *Bulletin of Mathematical Biophysics* 5 (4): 115–133

人工神经网络

■ Hebb's Rules 1949

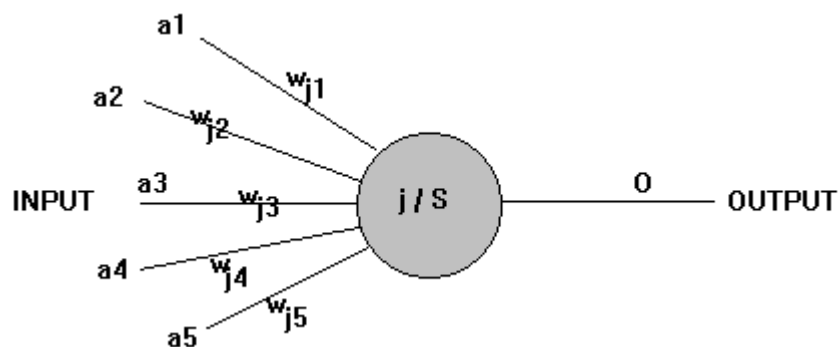
原文： When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A's efficiency, as one of the cells firing B, is increased。

对人工神经网络研究的指导意义： Hebb's principle can be described as a method of determining **how to alter the weights between model neurons**. The weight between two neurons increases if the two neurons activate simultaneously, and reduces if they activate separately. Nodes that tend to be either both positive or both negative at the same time have strong positive weights, while those that tend to be opposite have strong negative weights.

参考： Hebb, D.O. (1949). *The Organization of Behavior*. New York: Wiley & Sons.

人工神经网络

■ Rosenblatt 1958年感知机模型



$$f(x) = \begin{cases} 1 & \text{if } w \cdot x + b > 0 \\ 0 & \text{otherwise} \end{cases}$$

参考： F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386-408, 1958

人工神经网络

■ Rosenblatt 1958年感知机模型（续）

$$L(w, b) = - \sum_{x_i \in M} y_i (w \cdot x_i + b)$$

$$\nabla_w L(w, b) = - \sum_{x_i \in M} y_i x_i$$

$$\nabla_b L(w, b) = - \sum_{x_i \in M} y_i$$

则随机选取一个误分类点 (x_i, y_i) 对 w, b 进行更新式子如下：

$$w \leftarrow w + \eta y_i x_i$$

$$b \leftarrow b + \eta y_i$$

输入：训练数据集 $T = \{(x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)\}$ ， x_i 为实数向量， y_i 为 1 或 -1；
学习率 η ($0 < \eta \leq 1$)。

输出： w, b ，即感知机模型

(1) 选取初值 w_0, b_0 ；

(2) 在训练数据集中选取数据 (x_i, y_i) ；

(3) 如果 $y_i(w \cdot x_i + b) \leq 0$ ，则

$$w \leftarrow w + \eta y_i x_i$$

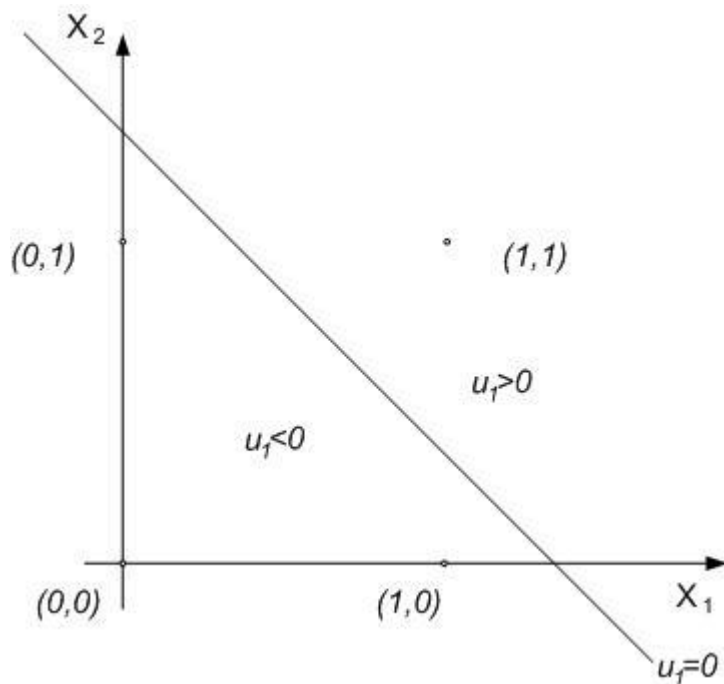
$$b \leftarrow b + \eta y_i$$

(4) 转至 (2)，直至训练数据集中没有误分类点。

参考：F. Rosenblatt. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychological Review*, 65:386-408, 1958

人工神经网络

■ 感知机的缺陷（不能处理XOR问题）



本质上是因为感知机是线性模型！

人工神经网络

■ Hinton, 1986 BP

Paul J. Werbos, his 1974 Harvard University Ph.D. thesis (**Beyond Regression: New Tools for Prediction and Analysis in the Behavioral Sciences**), which first described the process of training artificial neural networks through backpropagation of errors.

Y. LeCun: Une procédure d'apprentissage pour réseau a seuil asymmetrique (**a Learning Scheme for Asymmetric Threshold Networks**), Proceedings of Cognitiva 85, 599-604, Paris, France, 1985

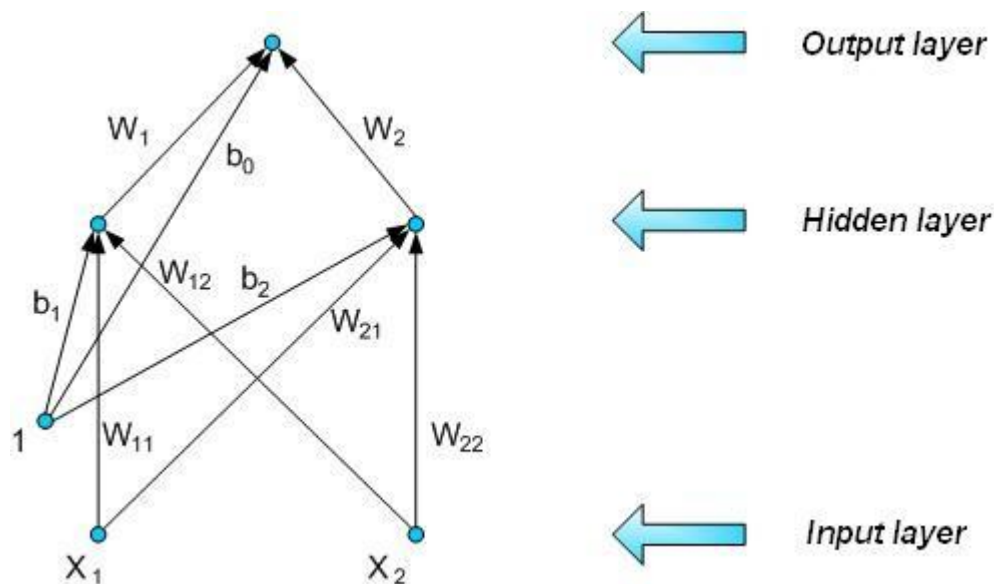
Rumelhart, David E.; Hinton, Geoffrey E.; Williams, Ronald J. (8 October 1986). "Learning representations by back-propagating errors". *Nature* **323** (6088): 533–536.

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial o_j} \frac{\partial o_j}{\partial \text{net}_j} \frac{\partial \text{net}_j}{\partial w_{ij}}$$

```
initialize network weights (often small random values)
do
  forEach training example ex
    prediction = neural-net-output(network, ex) // forward pass
    actual = teacher-output(ex)
    compute error (prediction - actual) at the output units
    compute  $\Delta w_h$  for all weights from hidden layer to output layer // backward pass
    compute  $\Delta w_i$  for all weights from input layer to hidden layer // backward pass continued
    update network weights // input layer not modified by error estimate
  until all examples classified correctly or another stopping criterion satisfied
return the network
```

人工神经网络

■ 多层感知机



理论发展：1989年Robert Hecht-Nielsen证明了对于任何闭区间内的一个连续函数都可以用一个隐含层的三层网络来逼近，这就是万能逼近定理。

人工神经网络

■ 小结

MCP模型奠定了人工神经元的理论基础。

感知机是第一个实用的人工神经网络算法，但两层感知机不能解决XOR问题。

BP算法的出现使得多层感知机可训练，解决了XOR问题，并且多层感知机具有所谓的“万能逼近”能力。

但三层以上的感知机收敛慢，容易陷入局部最优，在SVM出现之后关注者开始减少。

大纲

- 视觉神经科学基础
- 计算机视觉的理论计算模型
- 人工神经网络：从感知机到多层感知机
- —>Lecun的成就

Lecun的成就

■ 1989，卷积网络元年

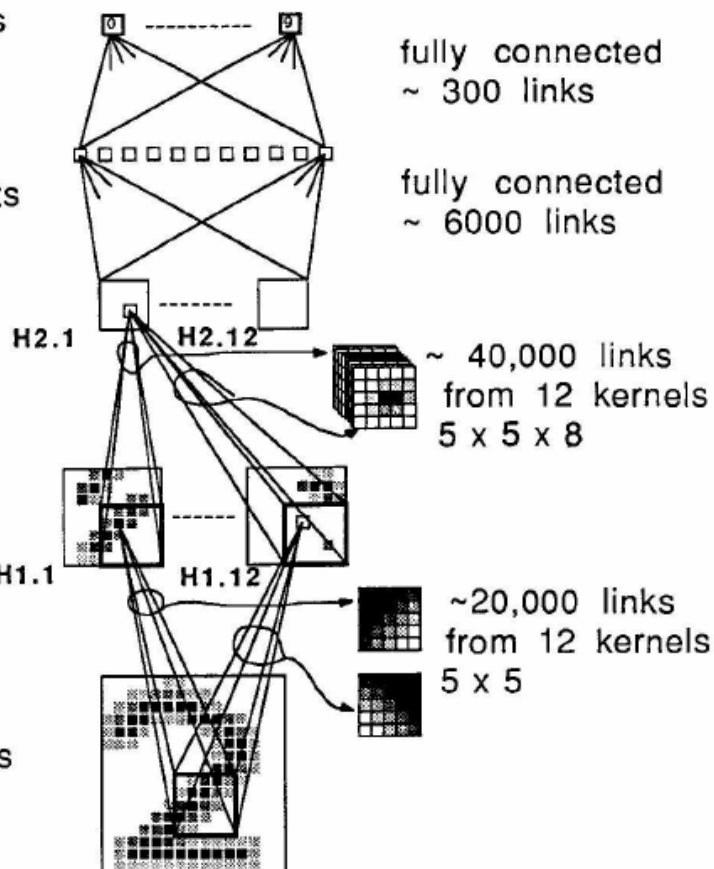
10 output units

layer H3
30 hidden units

layer H2
 $12 \times 16 = 192$
hidden units

layer H1
 $12 \times 64 = 768$
hidden units

256 input units



$$4000 \sim 192 * (5 * 5 * 8)$$

Lecun的成就

■ What is in the 1989 CNN Model?

All simulations were performed using the backpropagation simulator SN (Bottou and LeCun 1988) running on a SUN-4/260.

The nonlinear function used at each node was a scaled hyperbolic tangent. Symmetric functions of that kind are believed to yield faster convergence, although the learning can be extremely slow if some weights are too small (LeCun 1987). The target values for the output units were chosen within the quasilinear range of the sigmoid. This prevents the weights from growing indefinitely and prevents the output units from operating in the flat spot of the sigmoid. The output cost function was the mean squared error.

Before training, the weights were initialized with random values using a uniform distribution between $-2.4/F_i$ and $2.4/F_i$, where F_i is the number of inputs (fan-in) of the unit to which the connection belongs. This technique tends to keep the total inputs within the operating range of the sigmoid.

From empirical study (supported by theoretical arguments), the stochastic gradient was found to converge much faster than the true gradient,

正切函数收敛更快!

Sigmoid Loss!

网络参数初始化!

SGD!

Lecun的成就

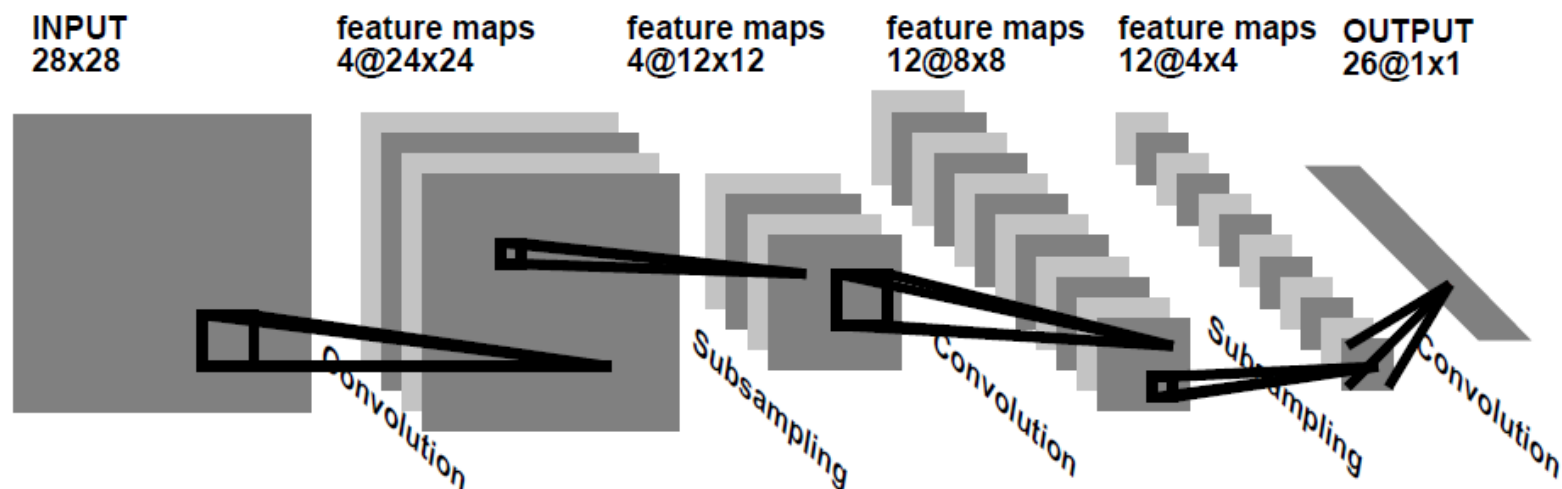
■ 98年之前的工作

1992年，就开始关注网络的泛化误差问题

H. Drucker and Y LeCun: **Improving Generalization Performance Using Double Backpropagation**, *IEEE Transaction on Neural Networks*, 3(6):991-997, **1992**,
[|cite{drucker-lecun-92}|](#).

1995，卷积网路的应用扩展

Y. LeCun and Y. Bengio: **Convolutional Networks for Images, Speech, and Time-Series**, in Arbib, M. A. (Eds), *The Handbook of Brain Theory and Neural Networks*, MIT Press, **1995**



Lecun的成就

■ 1998, LeNet

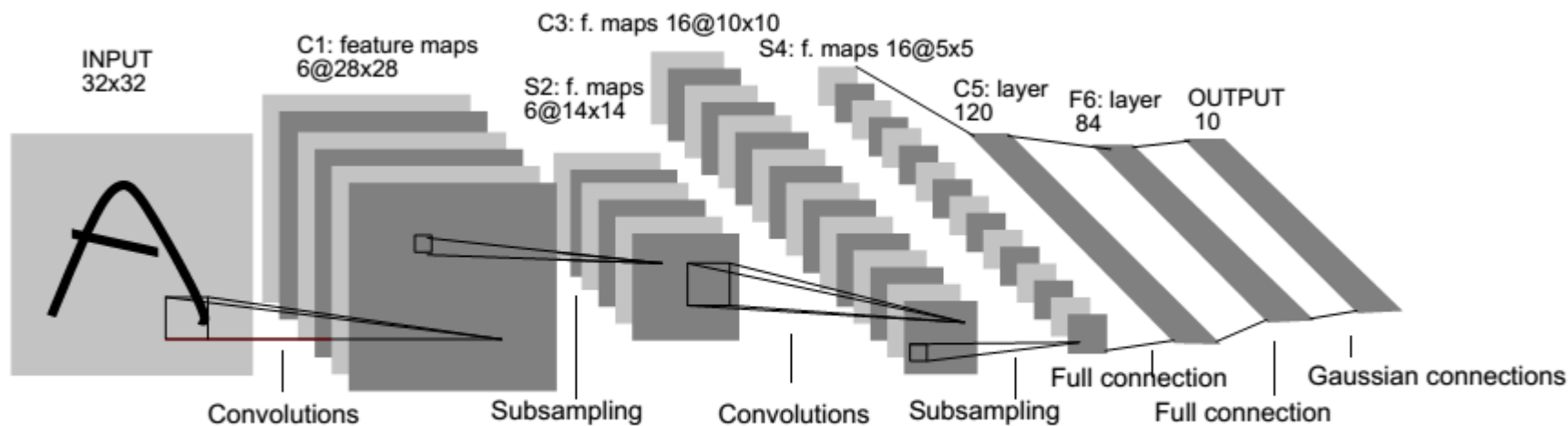


Fig. 2. Architecture of LeNet-5, a Convolutional Neural Network, here for digits recognition. Each plane is a feature map, i.e. a set of units whose weights are constrained to be identical.

Lecun的成就

■ Inside LeNet

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
0	X				X	X	X			X	X	X	X		X	X
1	X	X				X	X	X			X	X	X	X		X
2	X	X	X				X	X	X			X		X	X	X
3		X	X	X			X	X	X	X			X		X	X
4			X	X	X			X	X	X	X		X	X		X
5				X	X	X			X	X	X	X		X	X	X

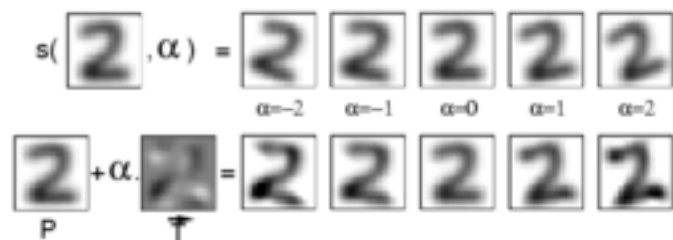
TABLE I

EACH COLUMN INDICATES WHICH FEATURE MAP IN S2 ARE COMBINED
BY THE UNITS IN A PARTICULAR FEATURE MAP OF C3.

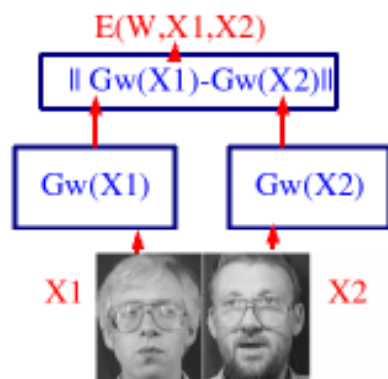
C₃层参数个数: $(3*6+4*6+4*3+6*1)*25 + 16 = 1516$

Lecun的成就

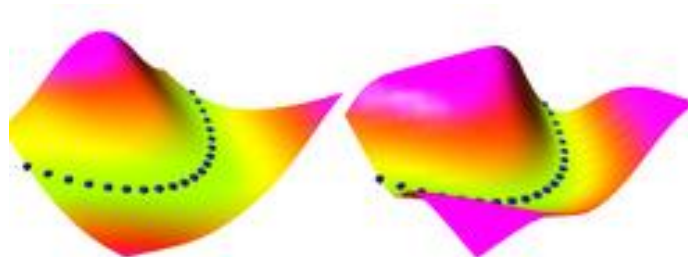
■ 1998之后



2000, Transformation Invariant Pattern Recognition



CVPR2005



2003至今, EBM

Lecun的成就

■ 小结

89年之前是读博士，搞BP。

89年到98年，主要就是卷积网络和BP算法的完善。

98年搞出了LeNet和Efficient BP。

无可奈何ConvNet在12年之前都没有玩的过hand-crafted feature，只好搞起了EBM理论模型。

下集预告

- Hinton的跨世纪探索
- 作为绿叶的Bengio
- 深度学习兵器谱
- 展望未来

谢谢支持!