视觉智能：进展与问题

山世光（sgshan@ict.ac.cn ）

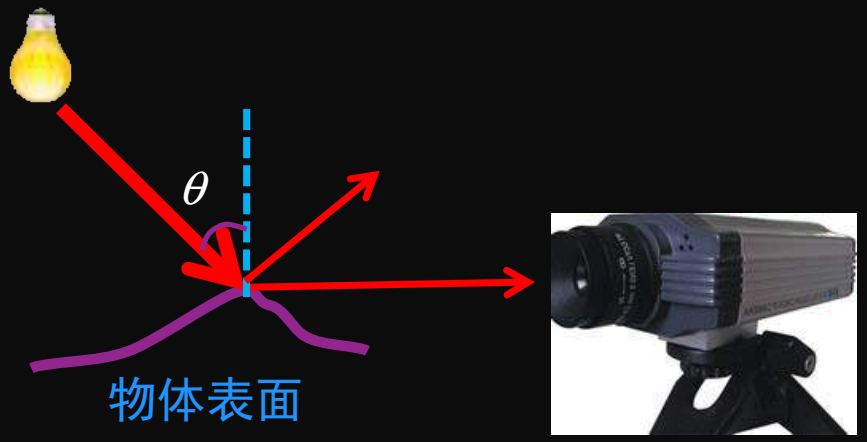
中国科学院计算技术研究所 研究员
中科院智能信息处理重点实验室 常务副主任
中科视拓（北京）科技有限公司 董事长/CTO

计算机视觉是什么?

SeetaTech 中科视拓

口相机或摄像机输出的是什么?

- ■图像/视频：空间中某物体表面点反射或发射的不同波长的光强
- ■物体表面不同材质对不同波长的光有不同的反射率（/吸收率）



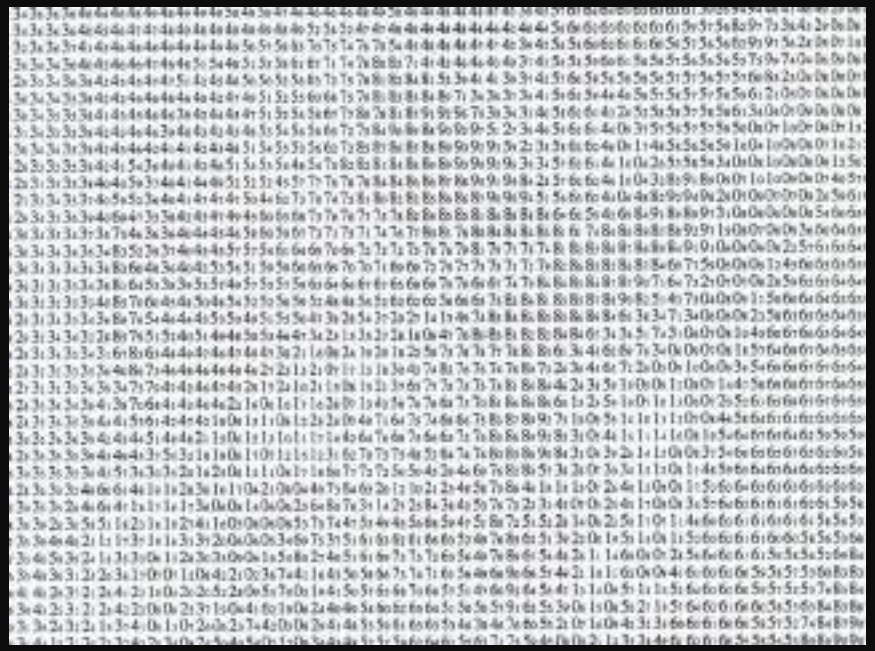
R（红）	G（绿）	B蓝
255	0	0
0	255	0
0	0	255
0	0	0
255	255	255
100	100	100
255	128	64

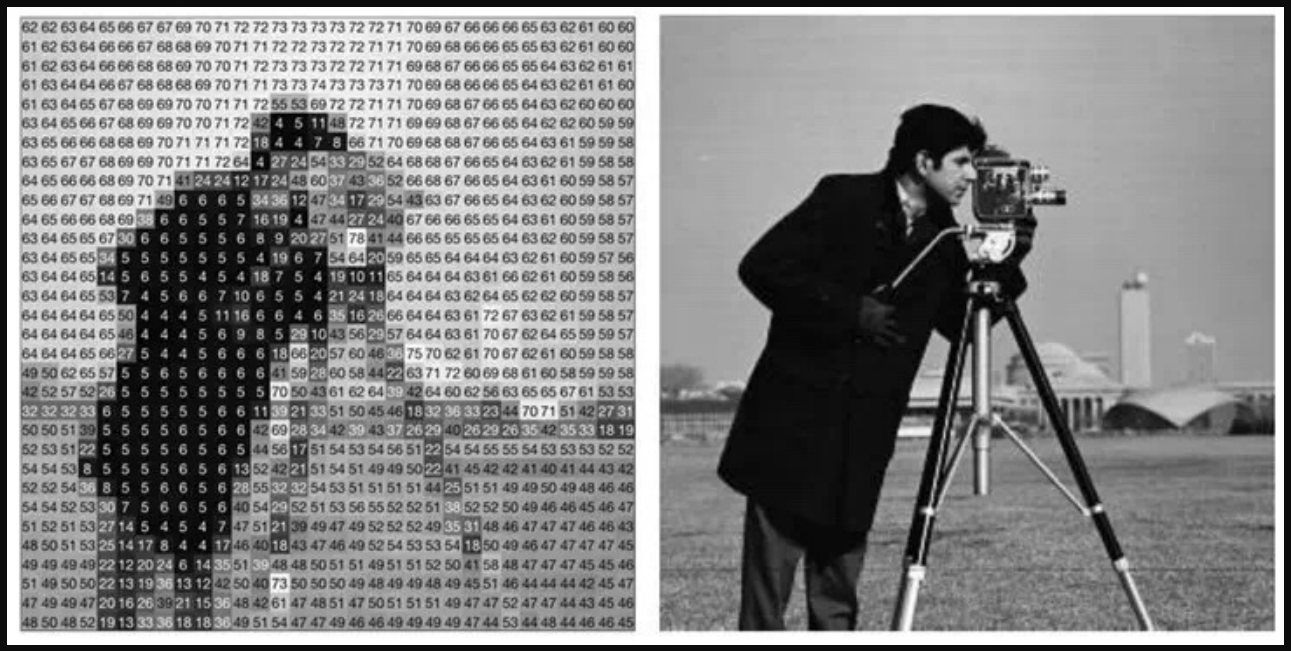
计算机视觉是什么?

SeetaTech 中科视拓

口解读w*h*3个0~255之间的数字中蕴藏的、人类可理解的内容

（边界，区域，物体，事件，意义）



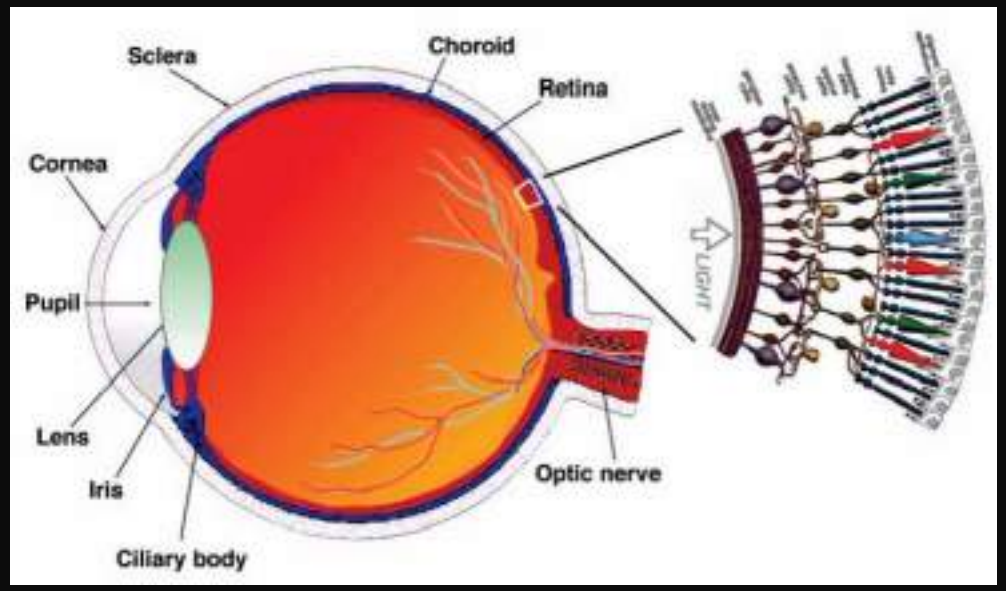


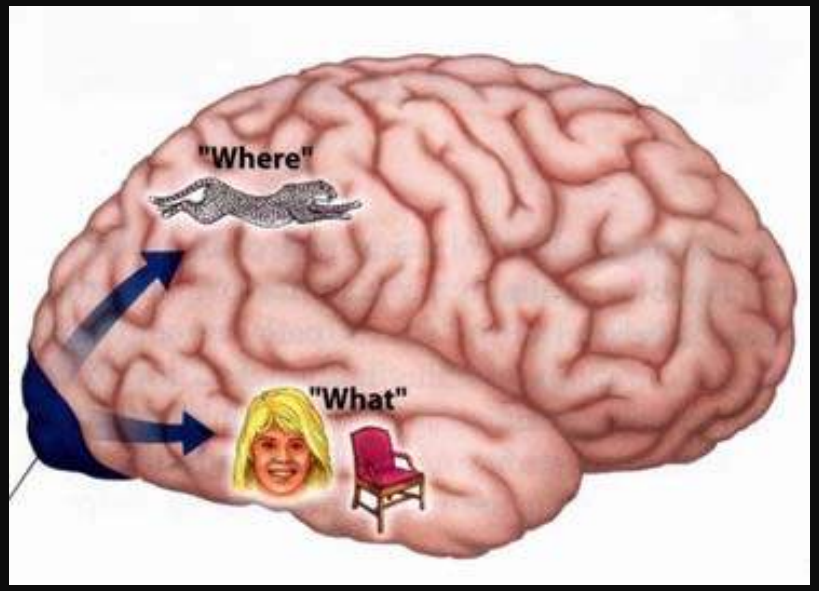
来自生物视觉系统的启示

SeetaTech 中科视拓

口我们如何看见?

- ■看：眼睛 == 摄像机
- ■见：视皮层的what通路和where通路



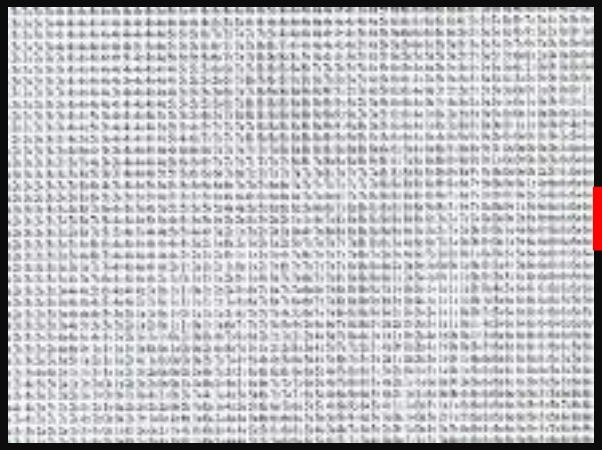


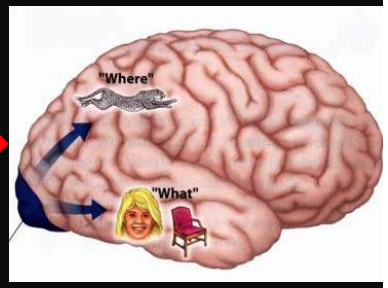
来自生物视觉系统的启示

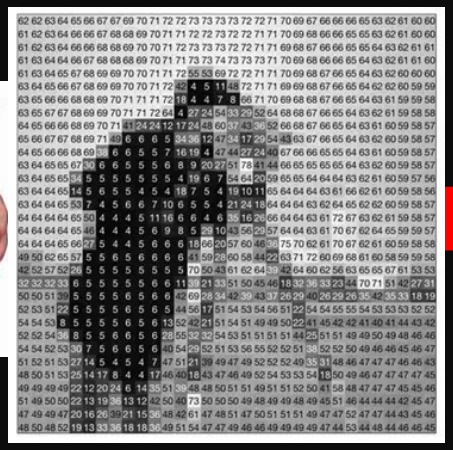
SeetaTech 中科视拓

□视皮层：从宏观到微观

- ■人类大脑共有860亿量级神经细胞，高度互联
- ■大脑→脑区→神经细胞互联→单个神经细胞→生化反应
- ■视皮层涉及多个脑区，几十亿神经细胞







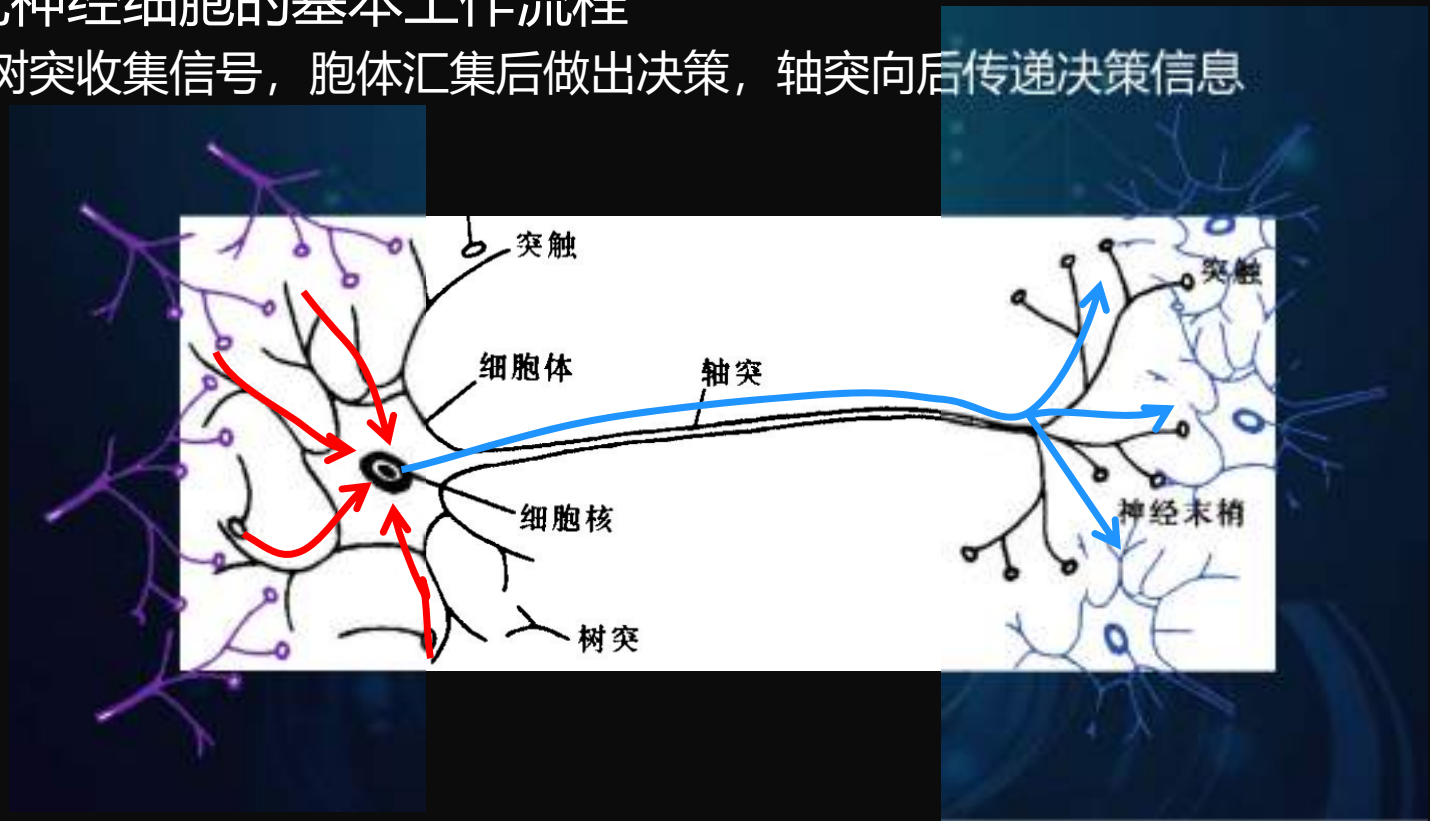
一张黑白照片，一位摄影师在用三角架上的照相机拍照，远处是...

来自生物视觉系统的启示

□神经细胞层面：高度互联，每个神经元与数千神经元连接

■感觉神经细胞的基本工作流程

□树突收集信号，胞体汇集后做出决策，轴突向后传递决策信息

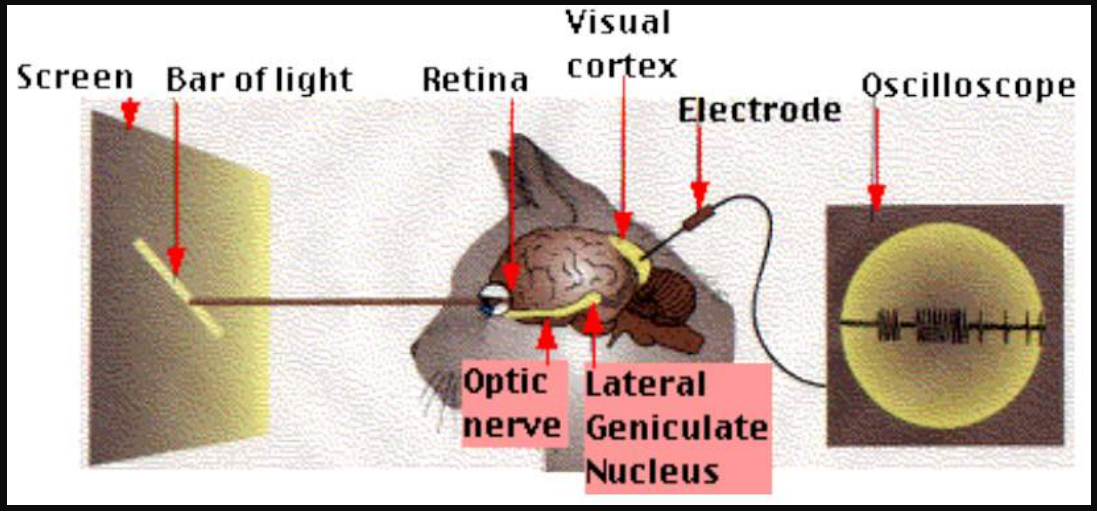


来自生物视觉系统的启示

SeetaTech 中科视拓

□每个神经元的功能

- ■高度专业化的功能分工：对某个特定决策投赞成票或者弃权票
- ■模式发现器：输入给自己的信息中，是否出现了自己感兴趣的模式

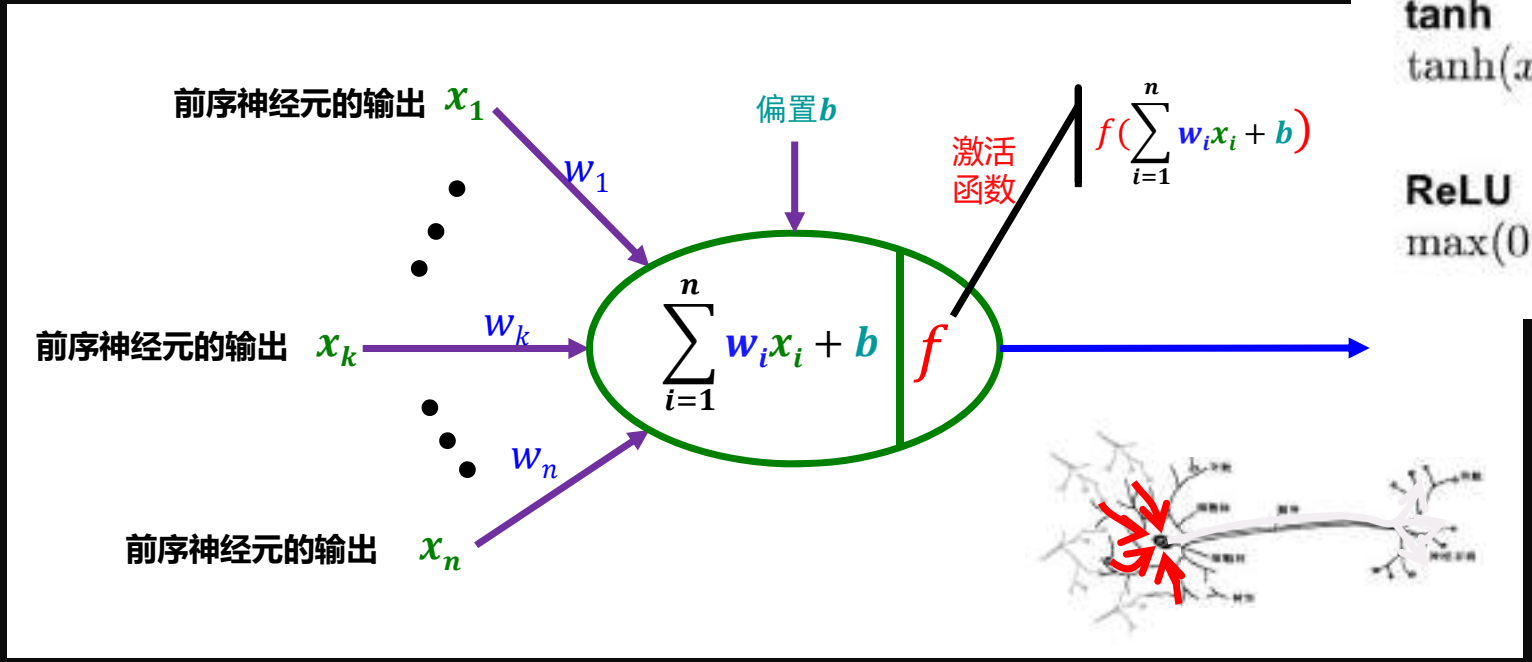


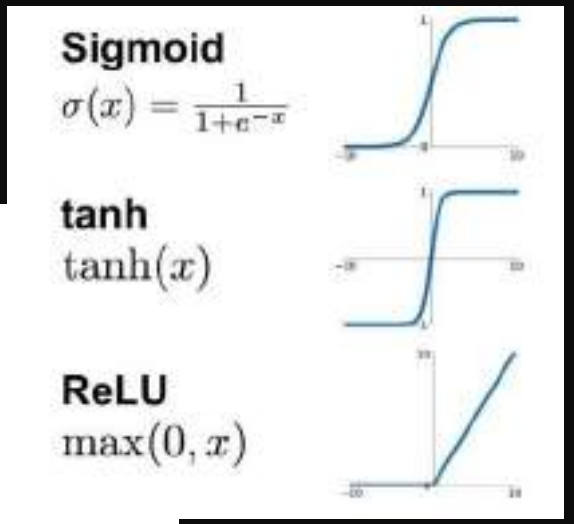
Hubel, D. H. & Wiesel, T. N. (1960s)

深度学习源起——单神经元计算模型

口单神经元计算模型

■加权求和（卷积）+ 非线性激活函数



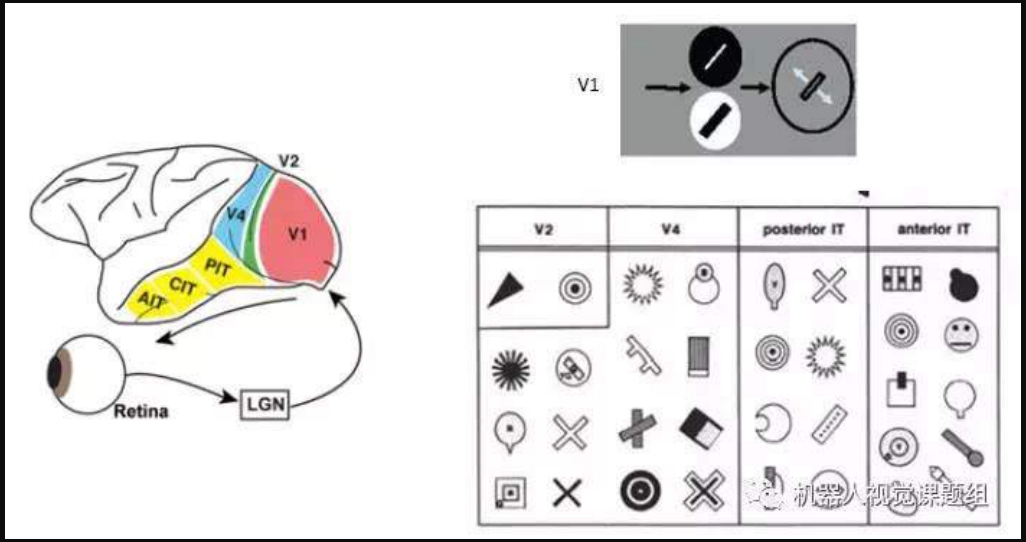


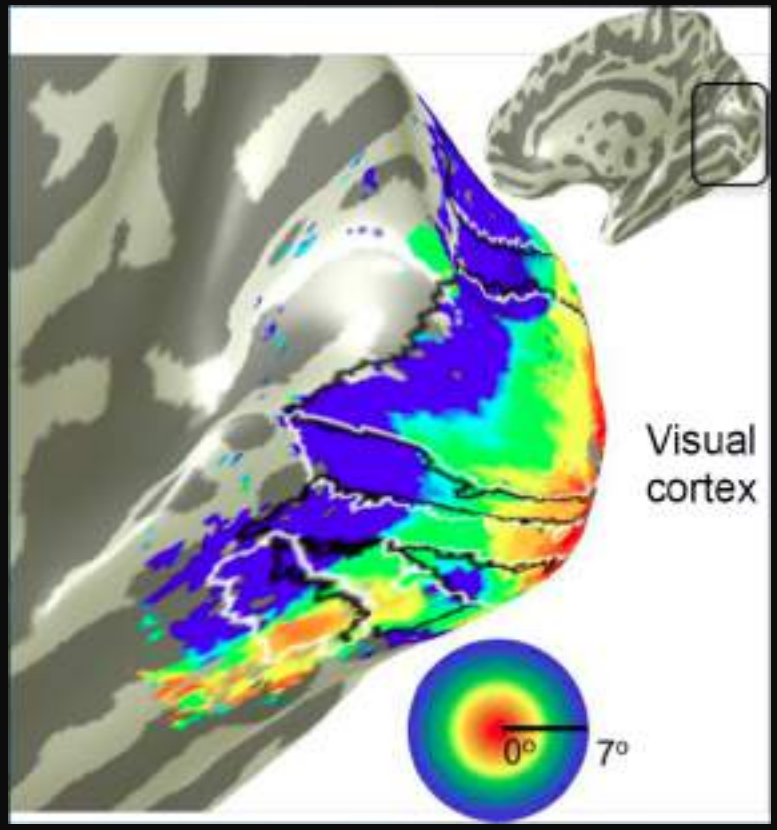
深度学习源起——层级感受野假设

SeetaTech 中科视拓

□层级感受野

- ■一个神经细胞看的更远（视野更大）、能处理更复杂的任务，是因为他站在了其他神经元的肩膀上！
- ■类比：社会结构





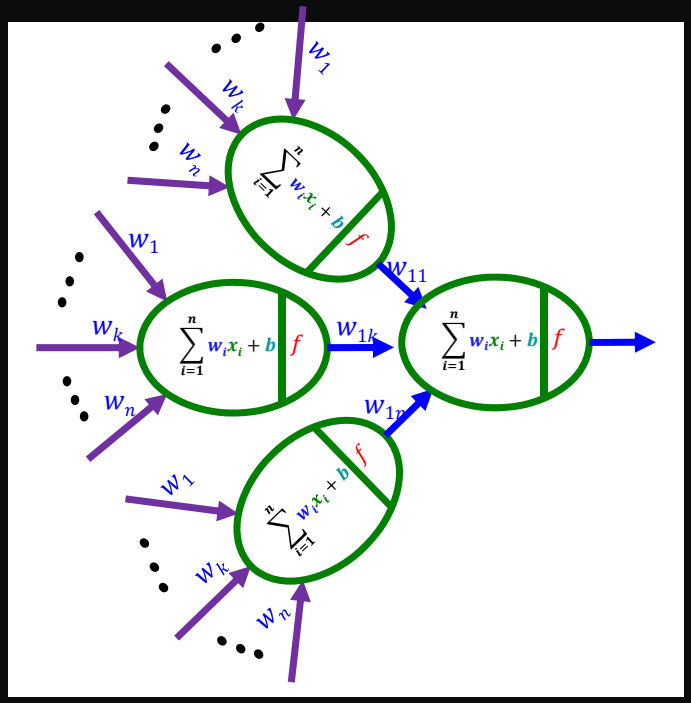
Retinotopy images courtesy of Jesse Gomez in the Stanford Vision & Perception Neuroscience Lab.

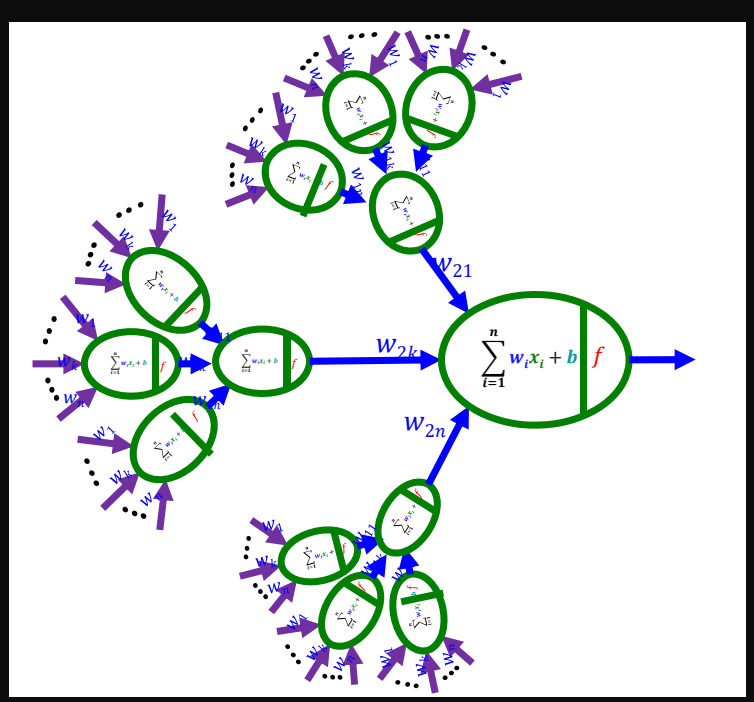
深度学习源起——多神经元互联模型

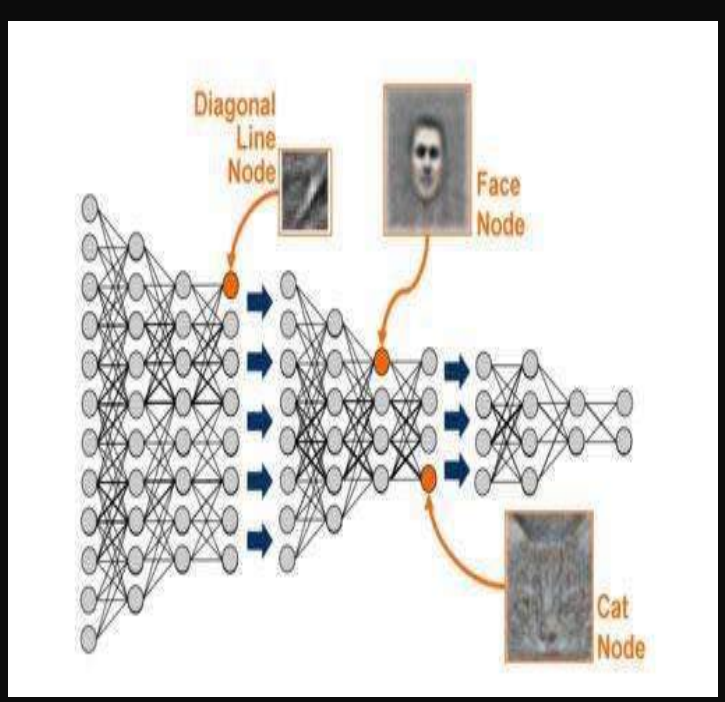
SeetaTech 中科视拓

□多个单神经元互联形成多层神经网络

■从输入到输出，中间多个隐含层（**隐层级多即所谓深度学习之深**）





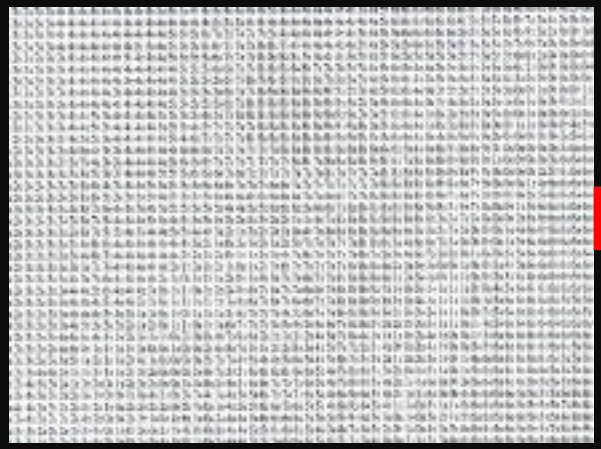


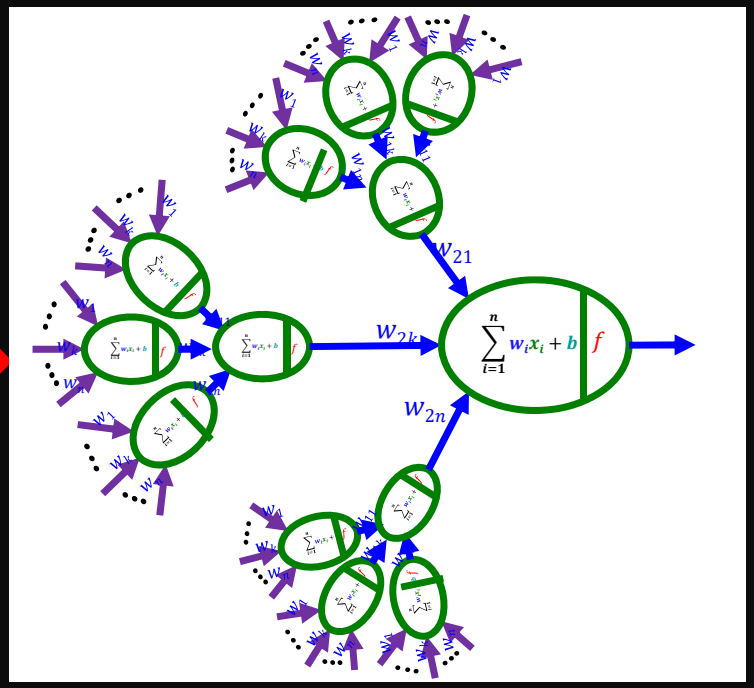
深度学习源起——多神经元互联模型

SeetaTech 中科视拓

□多个单神经元互联形成多层神经网络

■从输入到输出，中间多个隐含层（**隐层级多即所谓深度学习之深**）





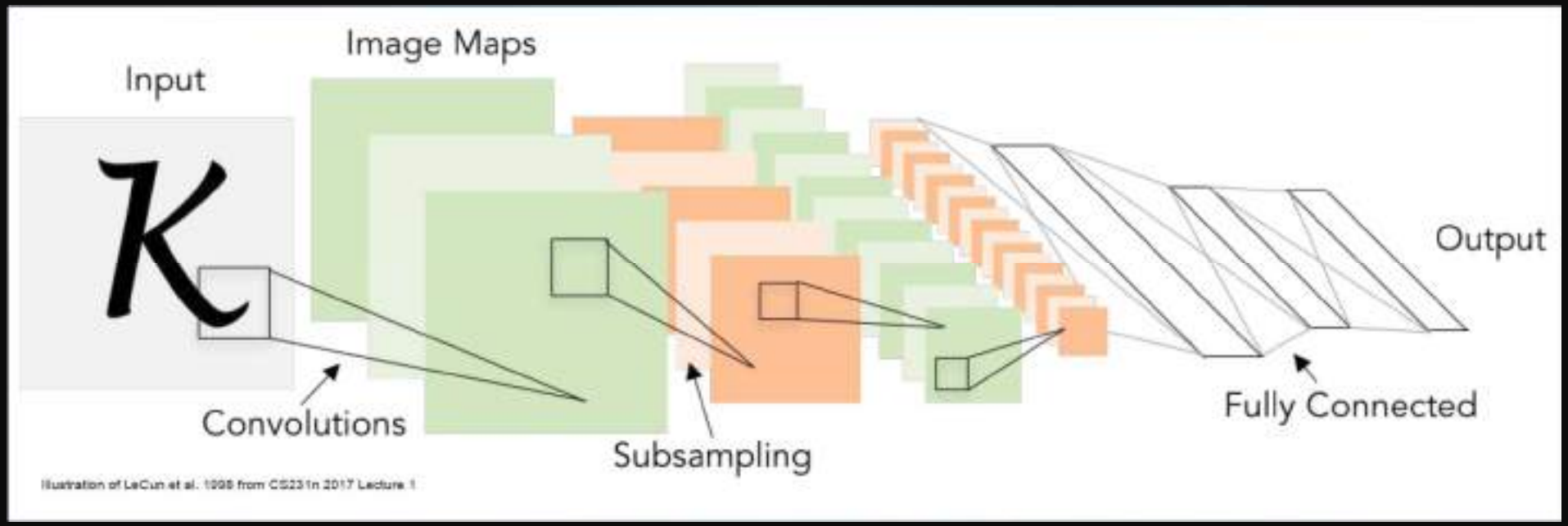
一张黑白照片，一位摄影师在用三角架上的照相机拍照，远处是...

卷积神经网络CNN

SeetaTech 中科视拓

口一种模拟层级感受野的前馈神经网络

- ■若干卷积层（非线性激活后），交叉若干Pooling(下采样)层
- ■若干全连接层，最后一层输出为类别标签（或目标回归值）



卷积神经网络CNN

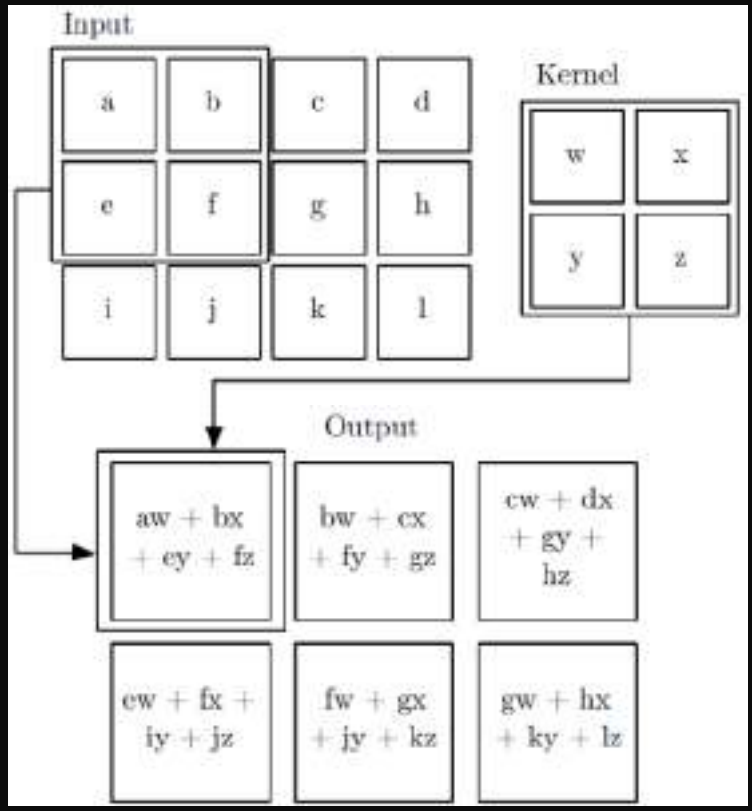
SeetaTech 中科视拓

卷积操作

- 图像与核(权重)的加权求和

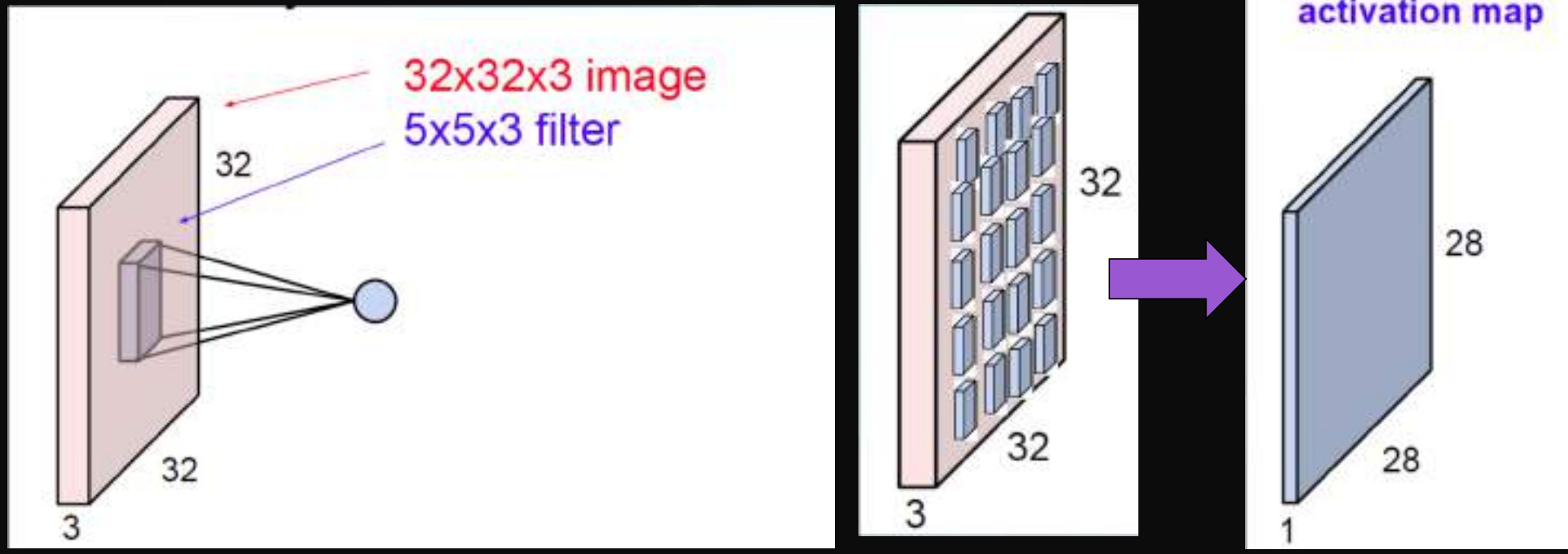
作用

- 本质上是一种滤波器，或局部特征提取器



卷积神经网络

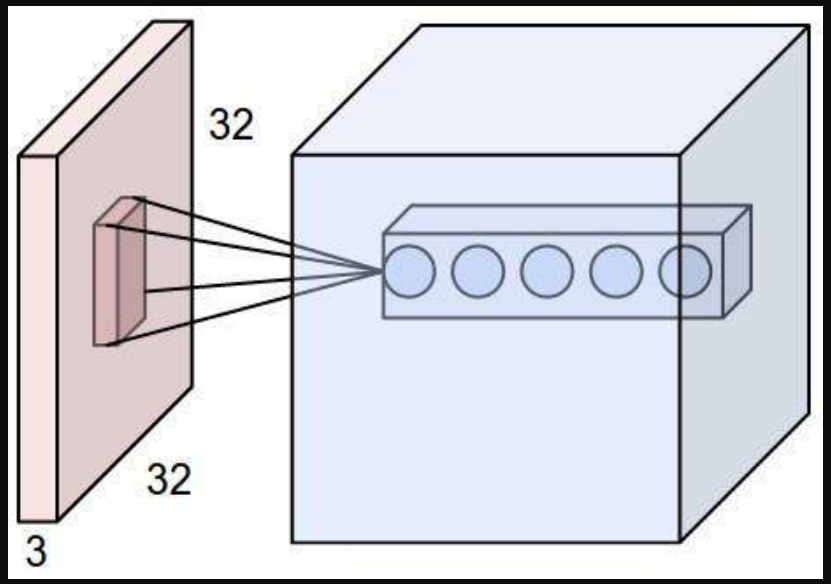
- 卷积层：局部连接，加权求和
- 功能：特征提取器，越来越复杂的特征，逐级抽象

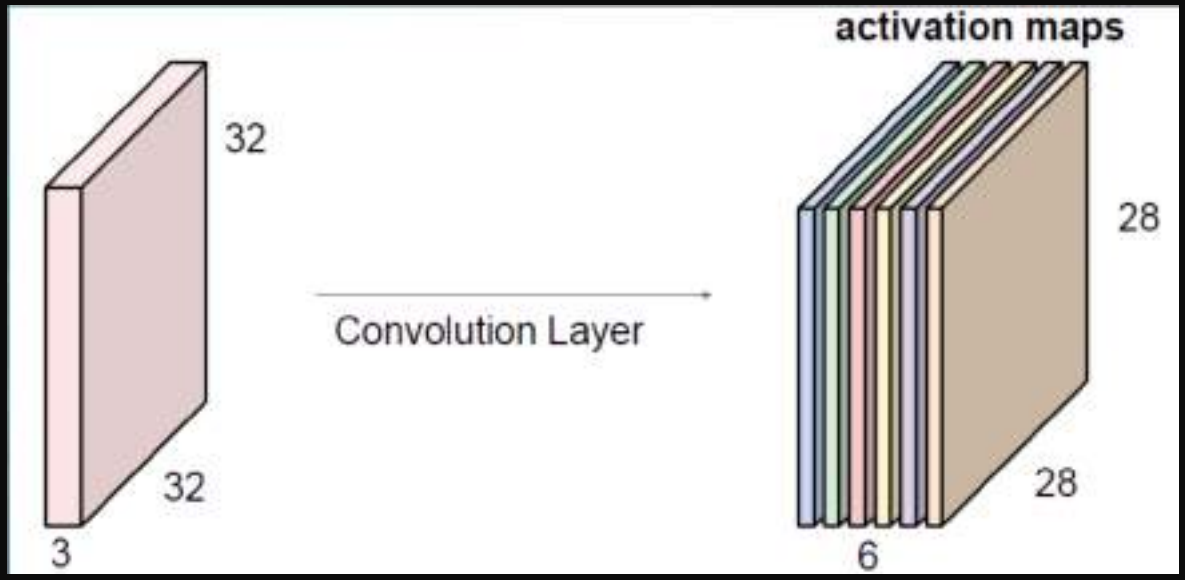


此页图片修改自李飞飞教授胶片

卷积神经网络

- 卷积层：局部连接，加权求和
- 功能：特征提取器，越来越复杂的特征，逐级抽象



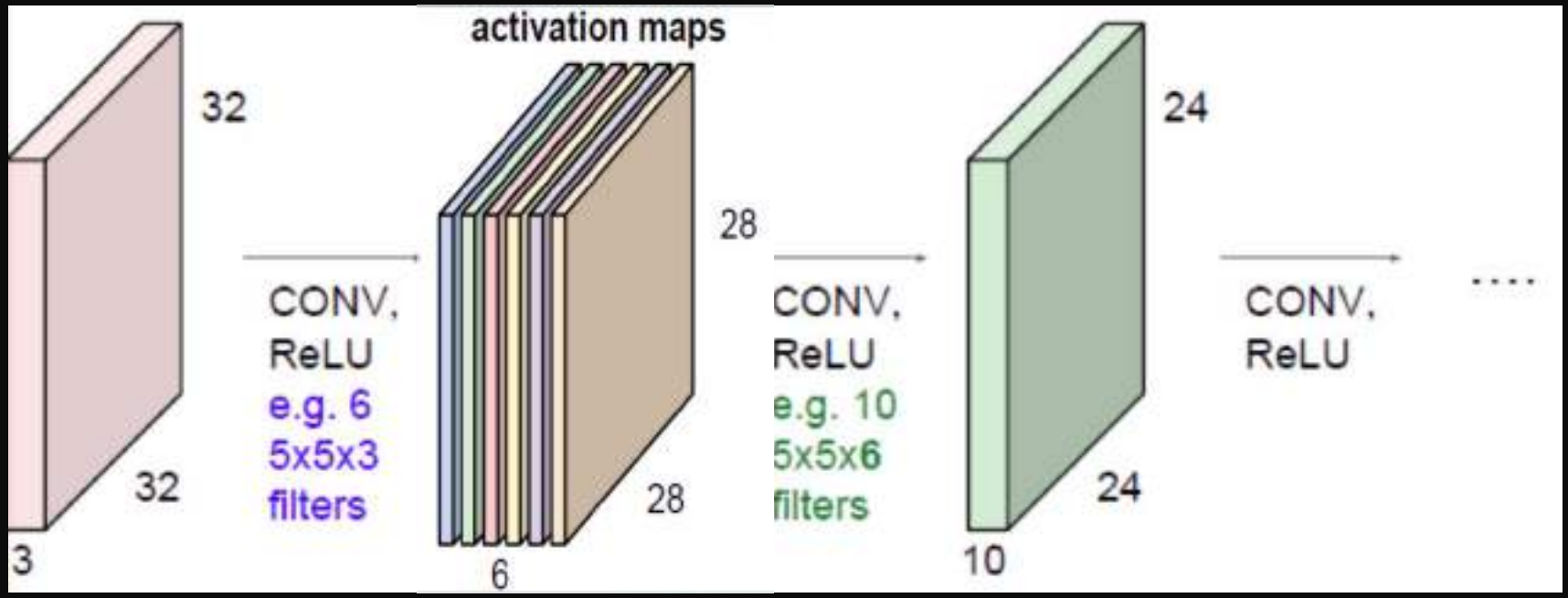


注：此页图片来自斯坦福大学李飞飞教授胶片

卷积神经网络CNN

SeetaTech 中科视拓

- 卷积层：局部连接，加权求和
- 功能：特征提取器，越来越复杂的特征，逐级抽象



注：此页图片来自斯坦福大学李飞飞教授胶片

卷积神经网络CNN

SeetaTech 中科视拓

口与普通滤波器的不同

- ■LoG，DoG，Gabor：权重系数人工设计
- ■CNN卷积：权重系数学习而来

卷积神经网络CNN

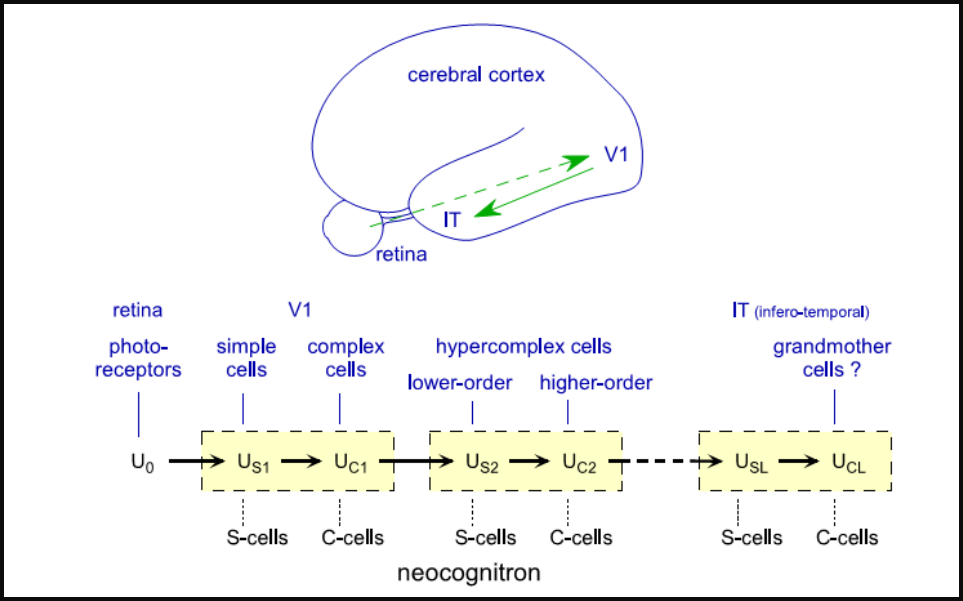
SeetaTech 中科视拓

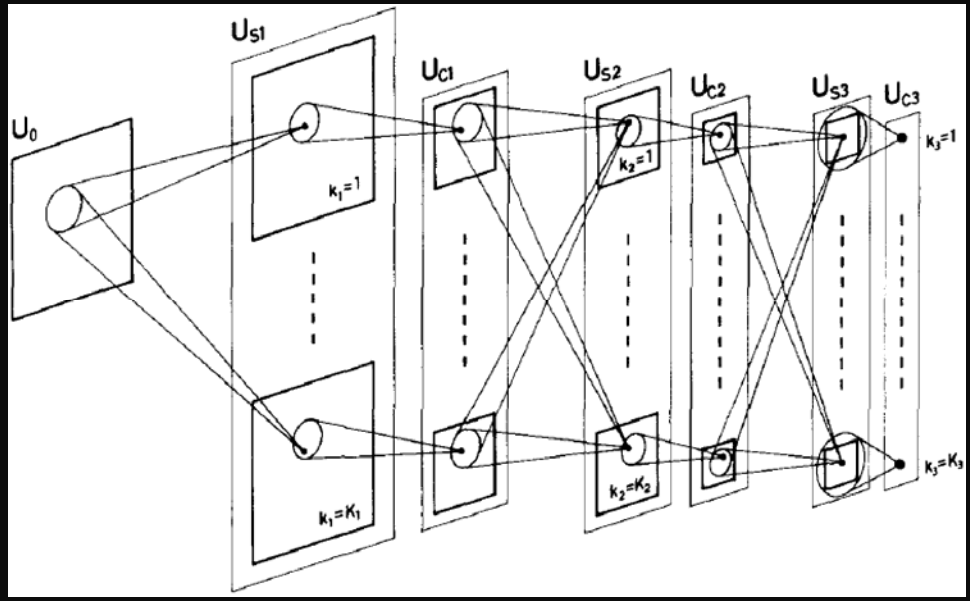
- K. Fukushima, “Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position,” Biological Cybernetics, vol. 36, pp. 193–202, 1980
- Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, “Backpropagation applied to handwritten zip code recognition,” Neural Computation, vol. 1, no. 4, pp. 541–551, 1989
- Y. Le Cun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-based learning applied to document recognition,” Proceedings of the IEEE, vol. 86, no. 11, pp. 2278–2324, 1998

卷积神经网络CNN

□Neocognitron, Fukushima 1980

- ■层级感受野：简单细胞→复杂细胞→超复杂细胞...
- ■中间层**无监督训练**（自组织），分类层有监督训练

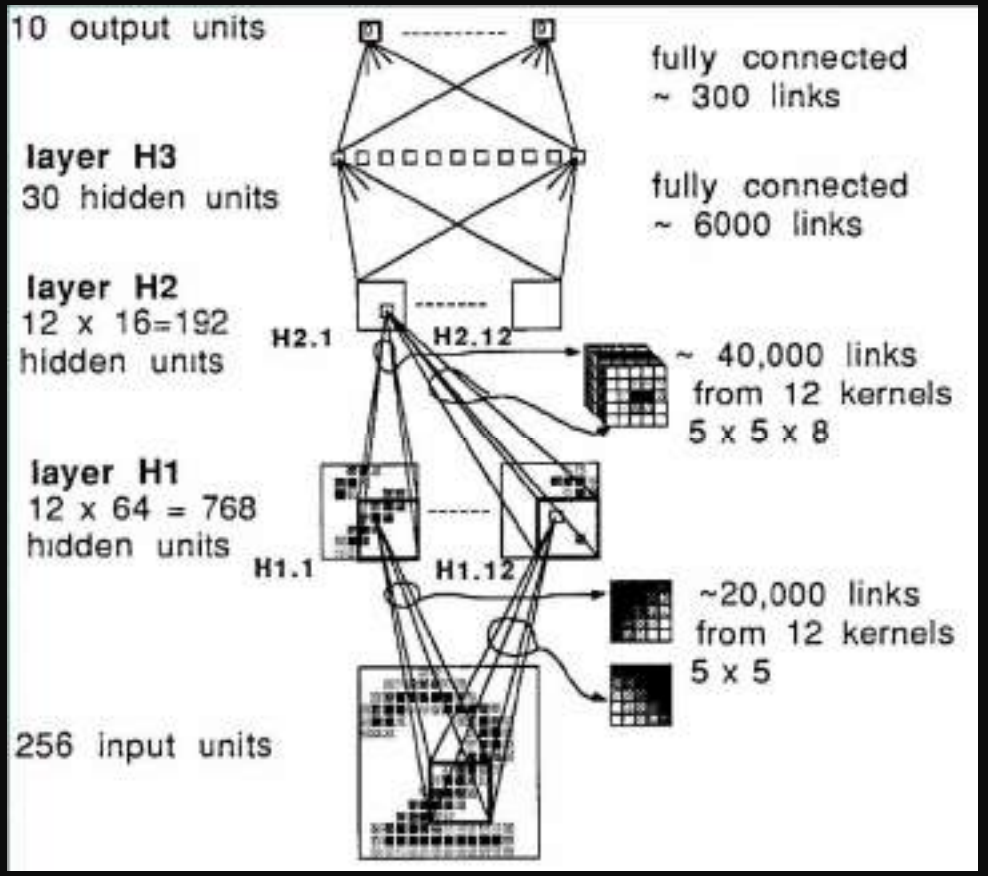




卷积神经网络CNN

SeetaTech 中科视拓

- LeCun et al. 1989
  - 简化了Neocognitron的结构
  - 训练方法
    - **监督训练**
    - BP算法
      - Sigmoid Loss
      - 随机梯度下降SGD
      - 正切函数收敛更快
  - 应用
    - 邮编识别

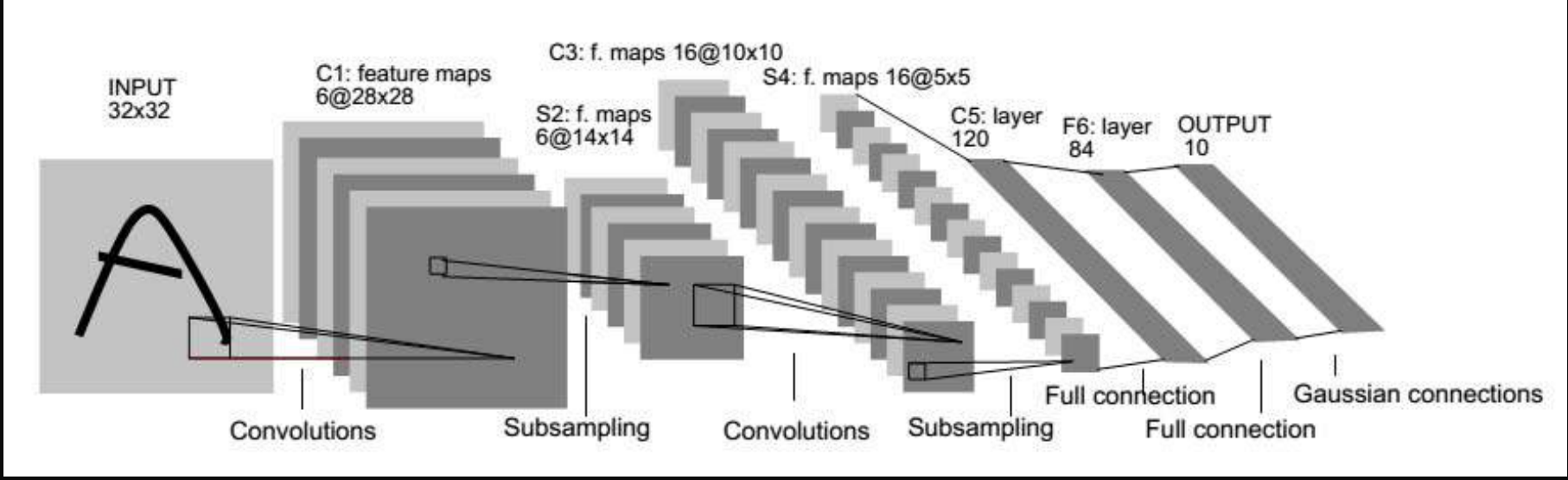


卷积神经网络CNN

SeetaTech 中科视拓

□CNN：LeCun et al 1998

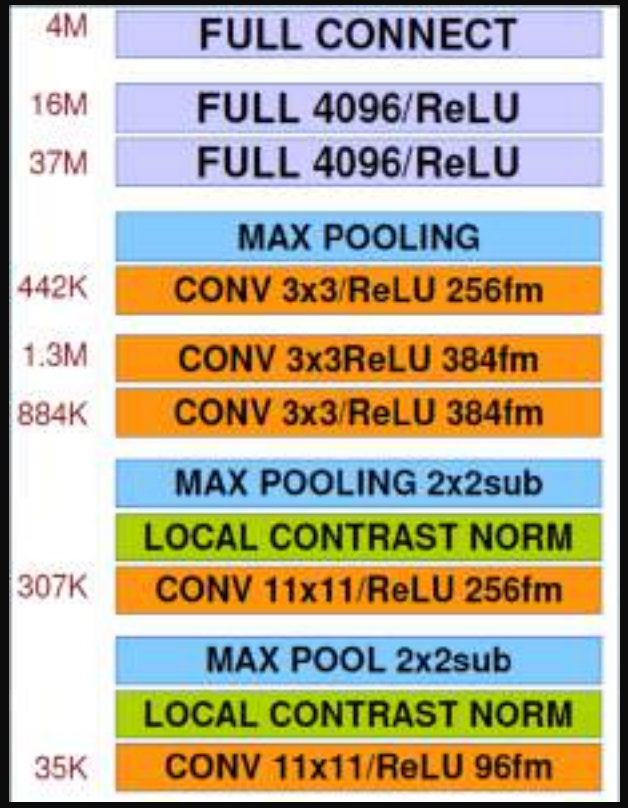
- ■2个卷积层+3个全连接层
- ■Pooling(/subsampling)

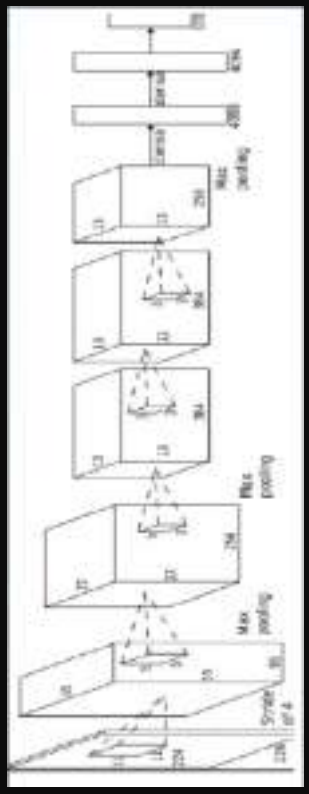


卷积神经网络CNN

SeetaTech 中科视拓

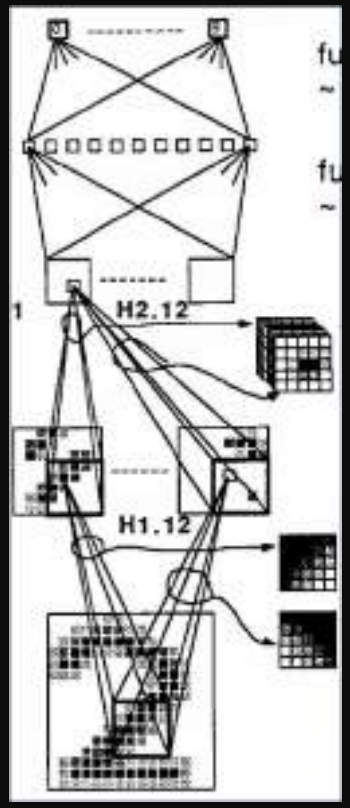
- AlexNet
  - 650K神经元
  - **6000万参数要学习**
  - 训练
    - BP on GPU
  - 各种优化技巧
    - ReLU
    - Dropout
    - 数据增广
    - ...



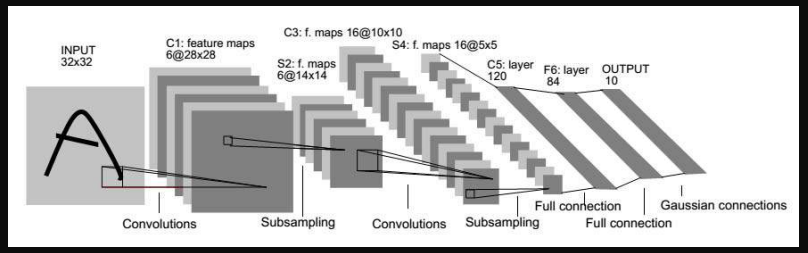


A. Krizhevsky, L. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” NIPS, 2012.

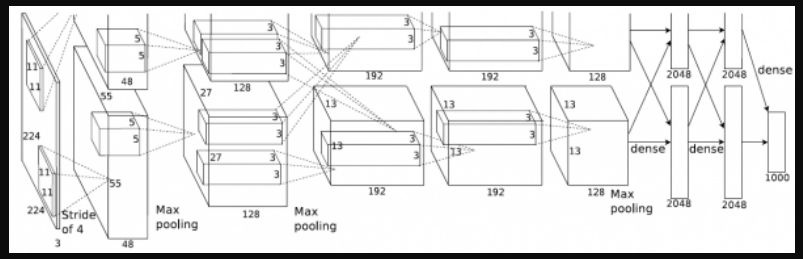
Renaissance神经网络（以CNN为例）



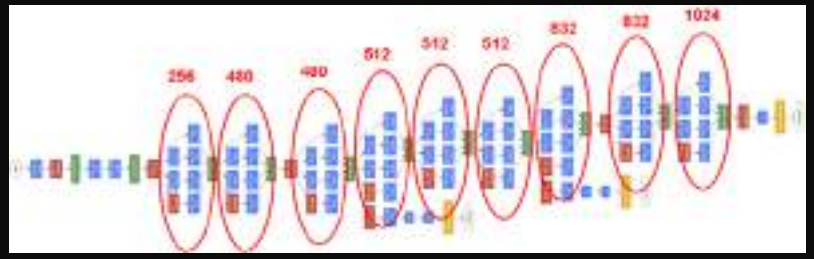
1989



1998



2012



2015

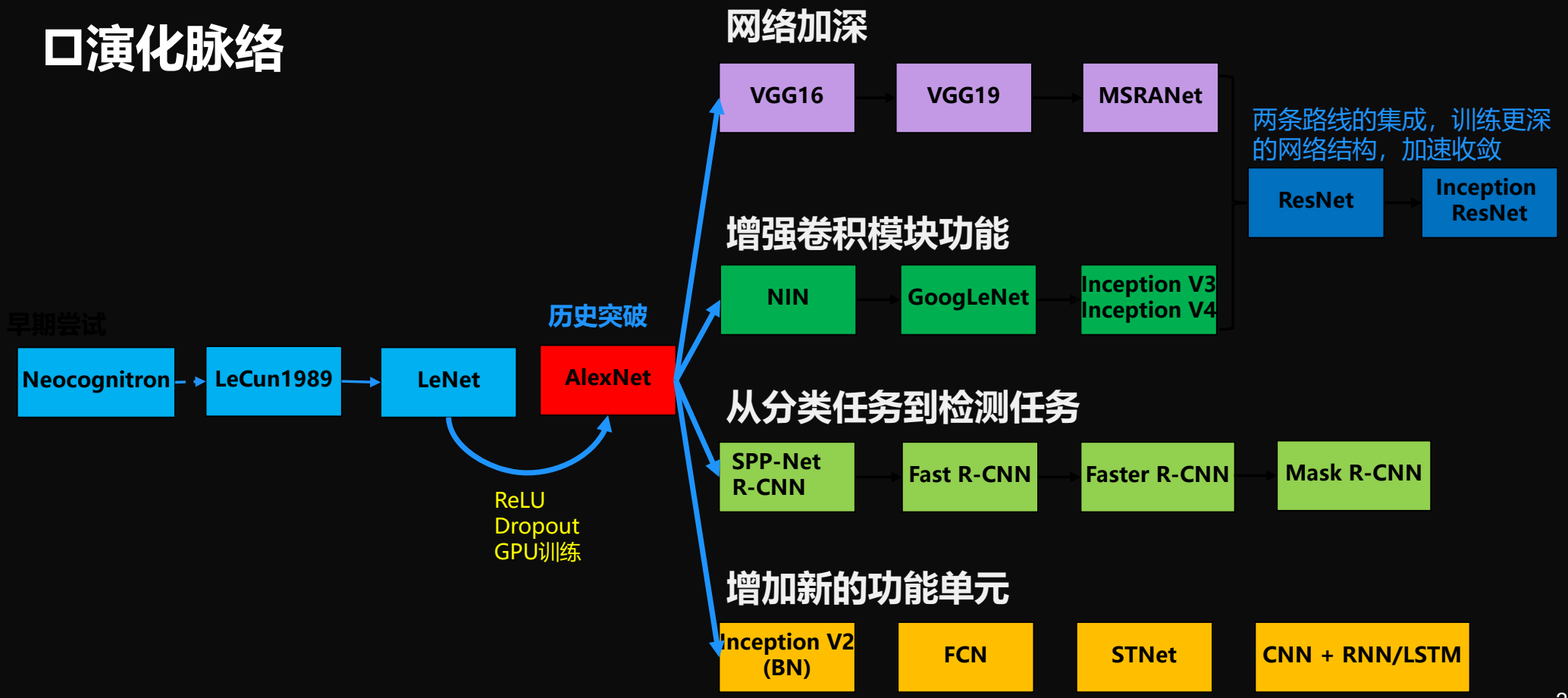
2014

2015

Renaissance神经网络（以CNN为例）

SeetaTech 中科视拓

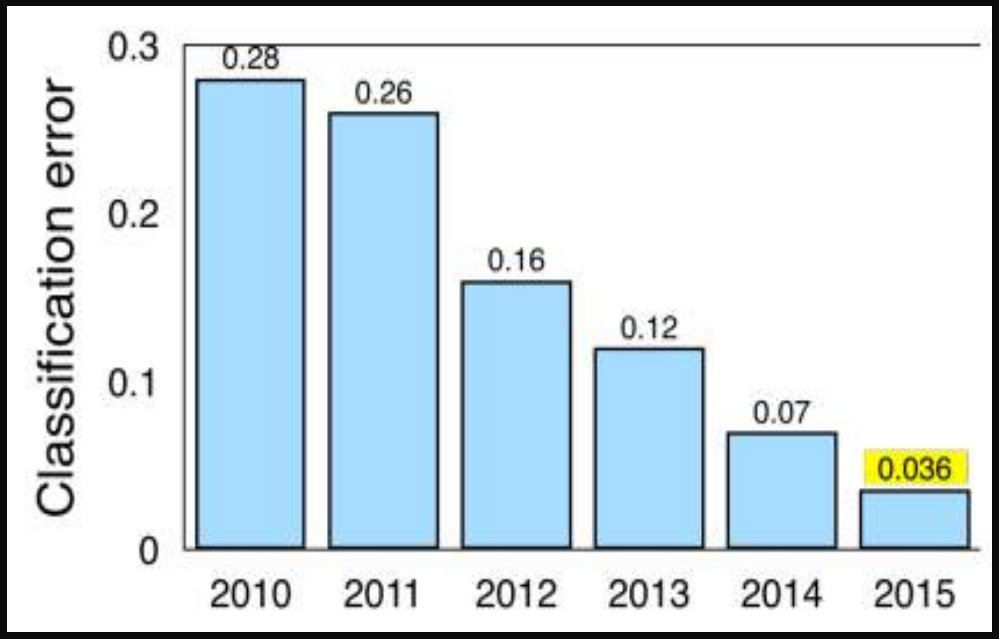
口演化脉络



深度学习为计算机视觉带来的进步

SeetaTech 中科视拓

- 图像分类上的跨越式进步【1000类Top5错误率: 26%→2.3%】

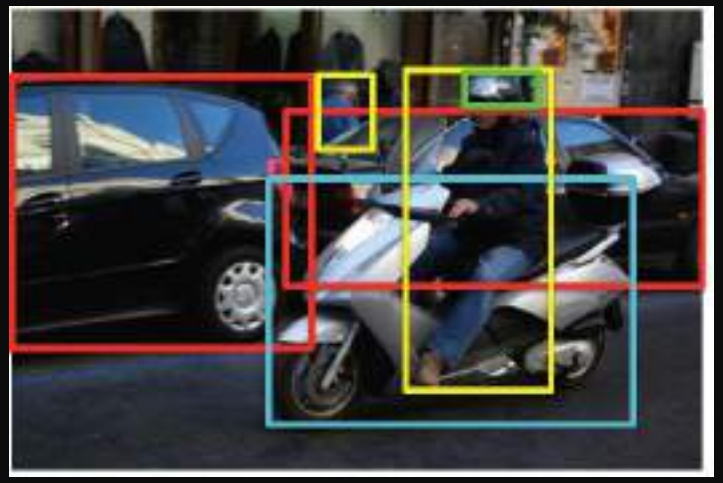


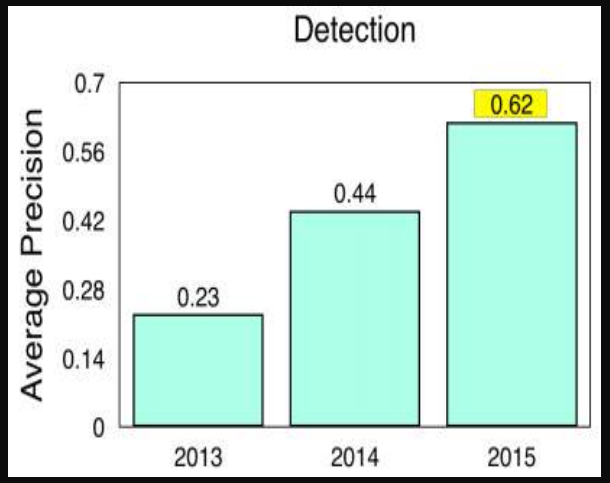
深度学习为计算机视觉带来的进步

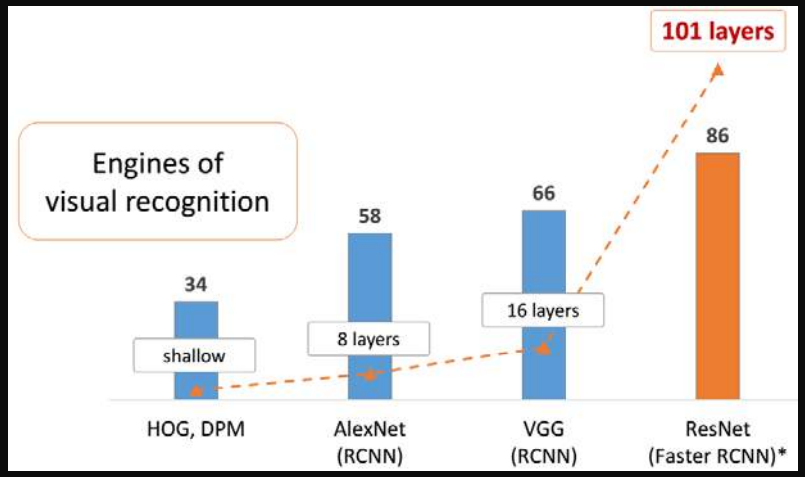
SeetaTech 中科视拓

口物体检测上的跨越(ImageNet竞赛)

- ■图像中200类物体检测mAP: 23%→**66%**(2016年)→**73%**(2017)
- ■视频中 30类物体检测mAP：68%(2015)→**81%**(2016年)→**82%**(2017)







Pascal VOC2007 检测性能演化

深度学习为计算机视觉带来的进步

SeetaTech 中科视拓

□视频结构化技术（中科视拓Seeta视觉引擎）

■监控场景行人车辆检测、跟踪与属性估计

深度学习为计算机视觉带来的进步

SeetaTech 中科视拓

□无人机视觉技术（中科视拓Seeta视觉引擎）

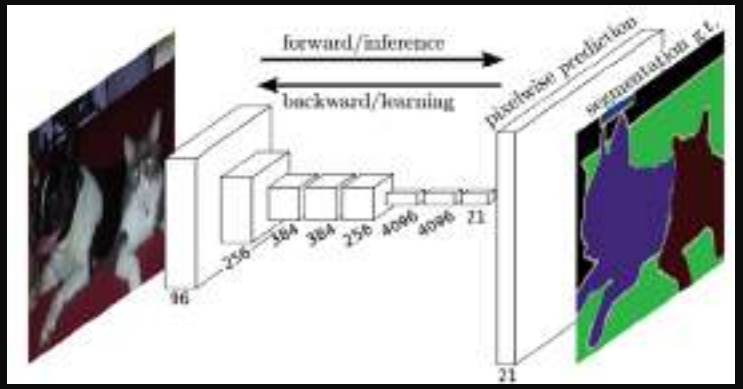
■地面目标检测与跟踪技术：检测率90%以上

深度学习为计算机视觉带来的进步

SeetaTech 中科视拓

口语义分割任务

- ■VOC2012数据集
- ■50%(2013) →75% (2015)→86.9%(2017)



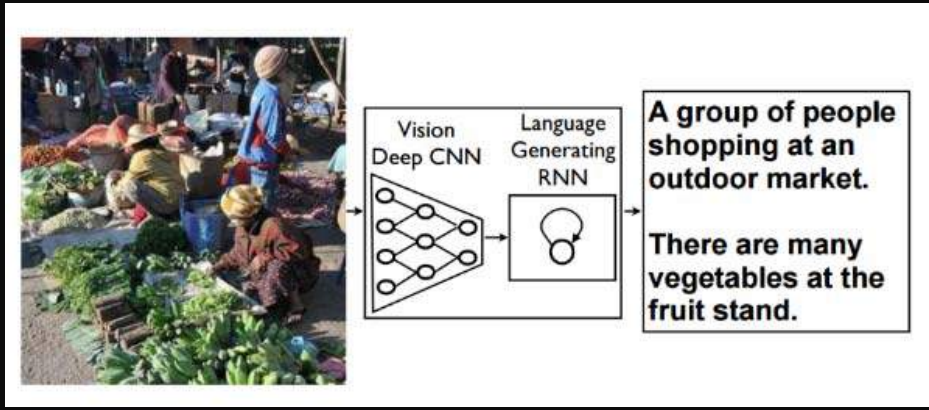
方法	VOC2012
NUS-UDS CVPR13	50.0
FCN-8s CVPR15	62.2
DeepLab ICLR15	71.6
FCN + CRF-RNN ICCV15	74.7
CMT-FCN-ResNet-CRF 2016	80.0
DeepLabv3-JFT 2017	86.9

VOC2012分割测试Mean IU Accuracy比较

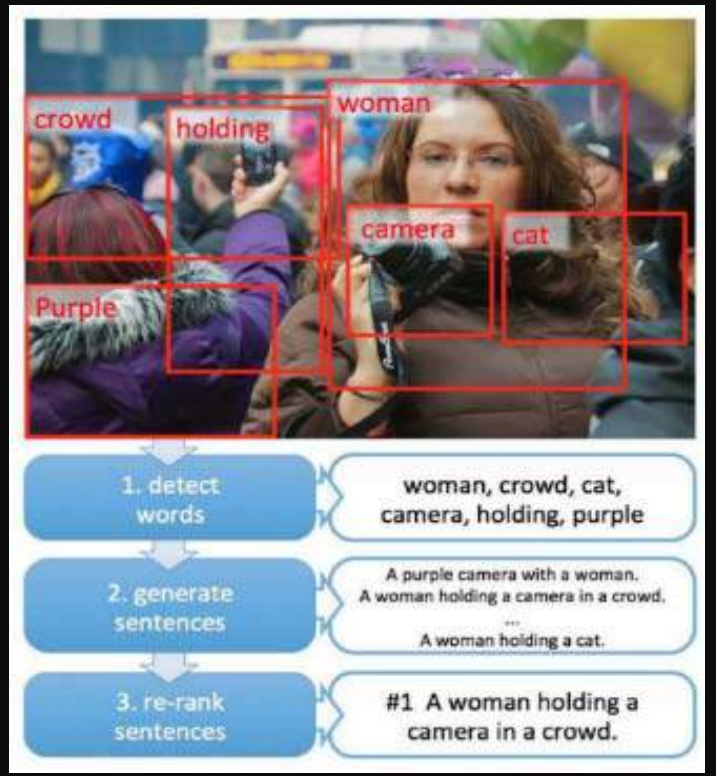
深度学习为计算机视觉带来的进步

SeetaTech 中科视拓

- Image Captioning（看图说话）
  - MSCOCO2015上（Google系统）
    - 27.3%优于人，31.7%通过图灵测试
- 视觉与语言的联姻



Show and Tell: A Neural Image Caption Generator (a work from Google)



From Captions to Visual Concepts and Back (a work from Microsoft)

深度学习为计算机视觉带来的进步

SeetaTech 中科视拓

□VQA：Visual Question Answering

□主要方法

- ■图像用CNN提取视觉特征
- ■语言(问句) 用LSTM形成特征
- ■二者互动，以分类或预测方式得到答案（非常类似于机器翻译）

□基本进展

- ■取得了有限的进步，依赖于大量问答句子进行训练（推理能力有限）

深度学习为计算机视觉带来的进步

SeetaTech 中科视拓

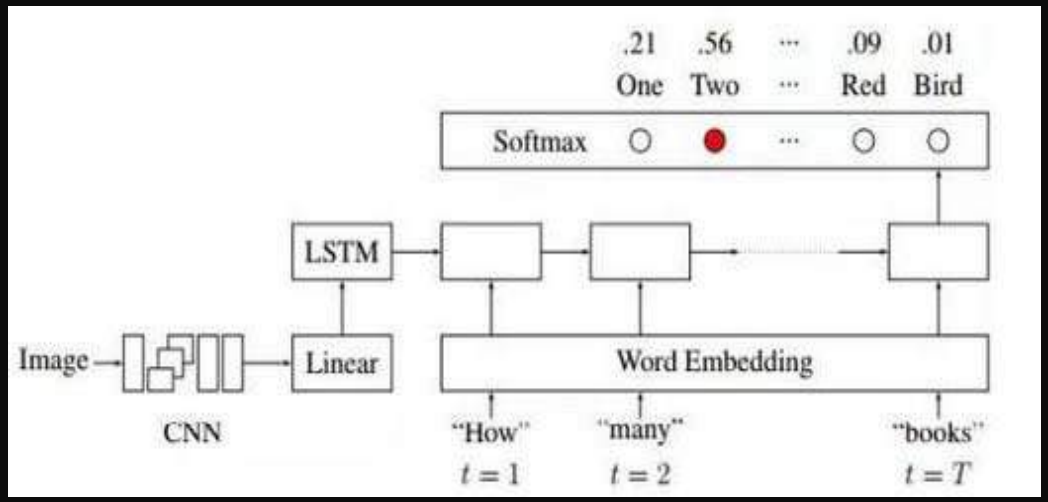
□VQA：Visual Question Answering

□主要方法

- ■图像用CNN提取视觉特征
- ■语言(问句) 用LSTM形成特征
- ■二者互动，以分类或预测方式得到答案（非常类似于机器翻译）

□基本进展

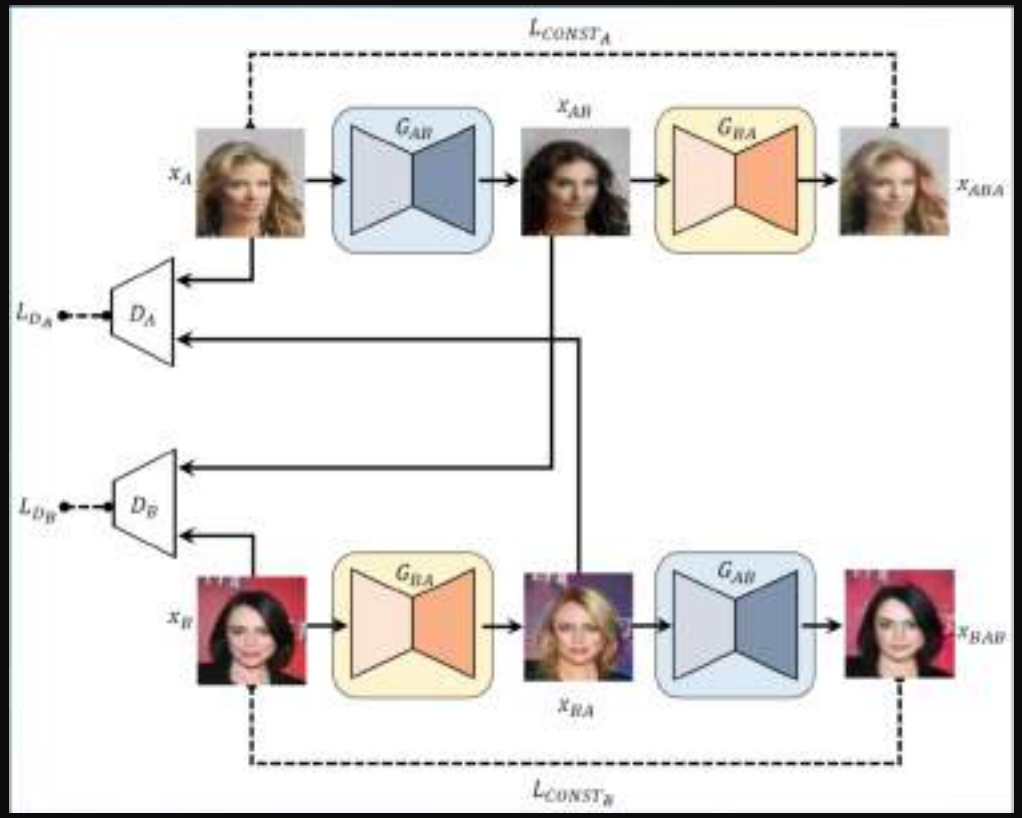
- ■取得了有限的进步，依赖于大量问答句子进行训练（推理能力有限）



深度学习为计算机视觉带来的进步

SeetaTech 中科视拓

- 图像合成及风格转换
  - 生成对抗网络GAN
  - 生成目标风格的逼真图像
  - 风格的学习
- 其意义在于“举一反三”



深度学习为计算机视觉带来的进步

SeetaTech 中科视拓

- 图像合成及风格转换
  - 生成对抗网络GAN
  - 生成目标风格的逼真图像
  - 风格的学习
- 其意义在于“**举一反三**”