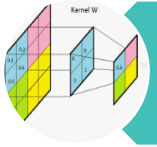
Deep Learning for Image Understanding

Convolutional Neural Networks

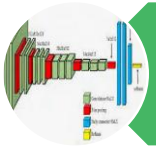
Xuejin Chen
xjchen99@ustc.edu.cn

Outline

Introduction



CNN Architectures

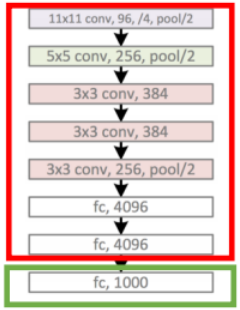


Classical CNN Models

Discussion

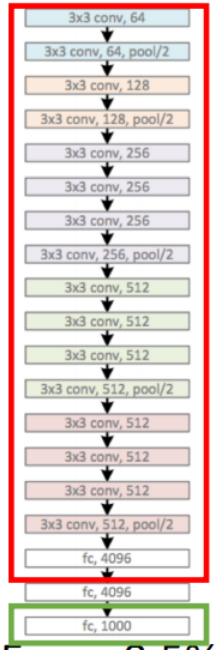
Deep ConvNet for Visual Recognition

2012: AlexNet
5 conv. layers



Error: 15.3%

2014: VGG
16 conv. layers



Error: 8.5%

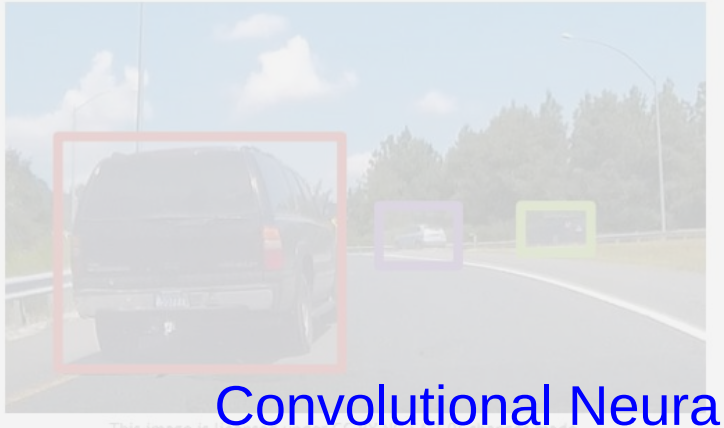
2015: GoogLeNet
22 conv. layers

Error: 7.8%

2016: ResNet
>100 conv. layers

Error: 4.4%

Slide from Bolei Zhou, MIT



This image is licensed under CC BY-NC-SA 2.0; changes made

- Object detection
- Action classification
- Image captioning
- ...

Convolutional Neural Networks (CNN) have become an important tool for object recognition

This image is licensed under CC BY-SA 2.0; changes made

This image is licensed under CC BY-SA 3.0; changes made

Recap: Image Recognition

Image

Array of 32x32x3 numbers
(3072 numbers total)

$f(\mathbf{x}, \mathbf{W})$

Parameters /weights

10 numbers giving class scores

Recap: Traditional Model

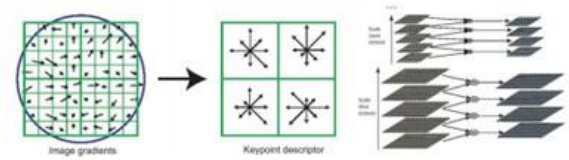
- Curse of dimensionality
- Non-linear system

hand-crafted Feature Extractor → “Simple” Trainable Classifier

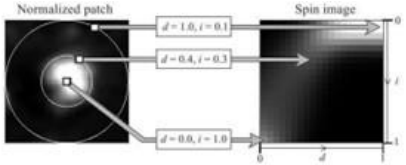
hand-crafted Feature Extractor	“Simple” Trainable Classifier
Sift	SVM
BOW	Bayes
HOG	LDA
LBP	…
…	

- The traditional model of pattern recognition (since the late 50's)
  Fixed/engineered features (or fixed kernel) + Trainable classifier

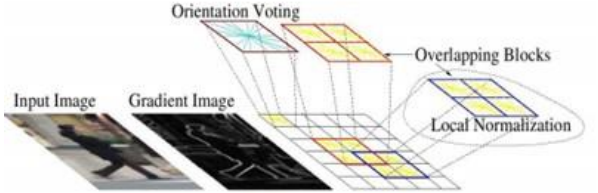
Recap: Features



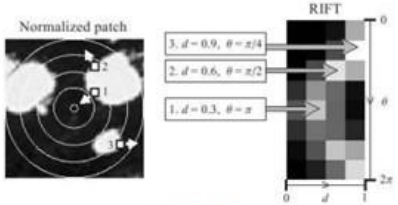
SIFT



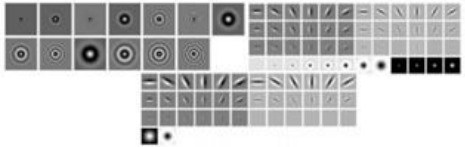
Spin image



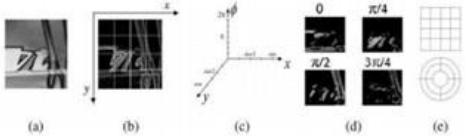
HoG



RIFT



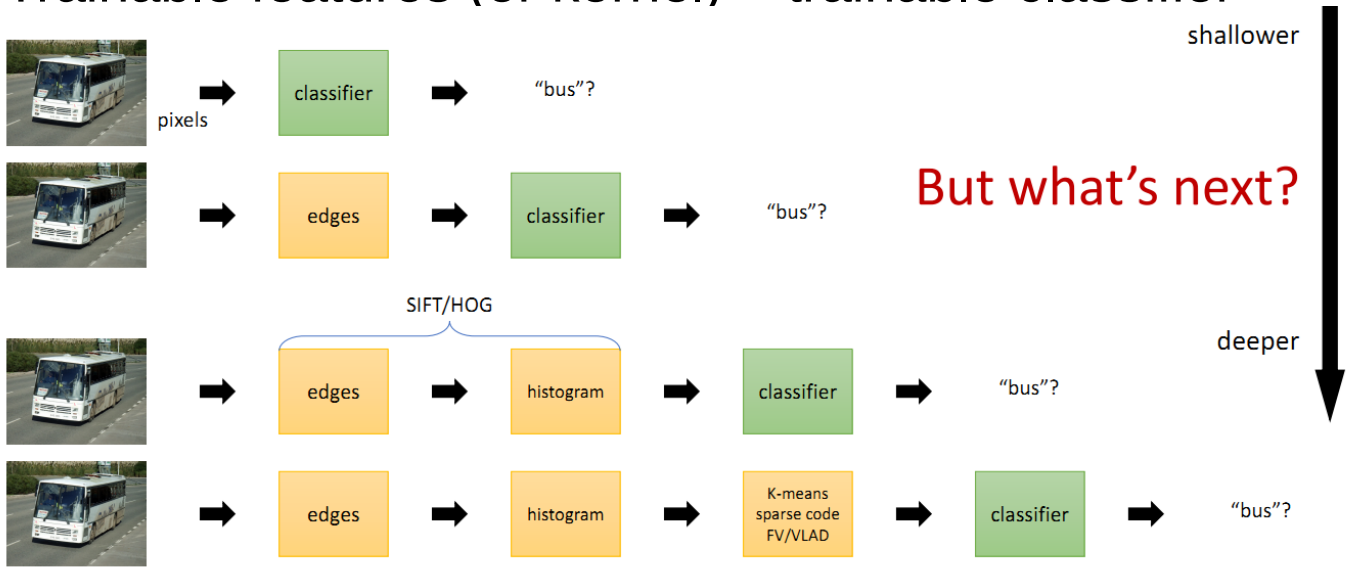
Textons



GLOH

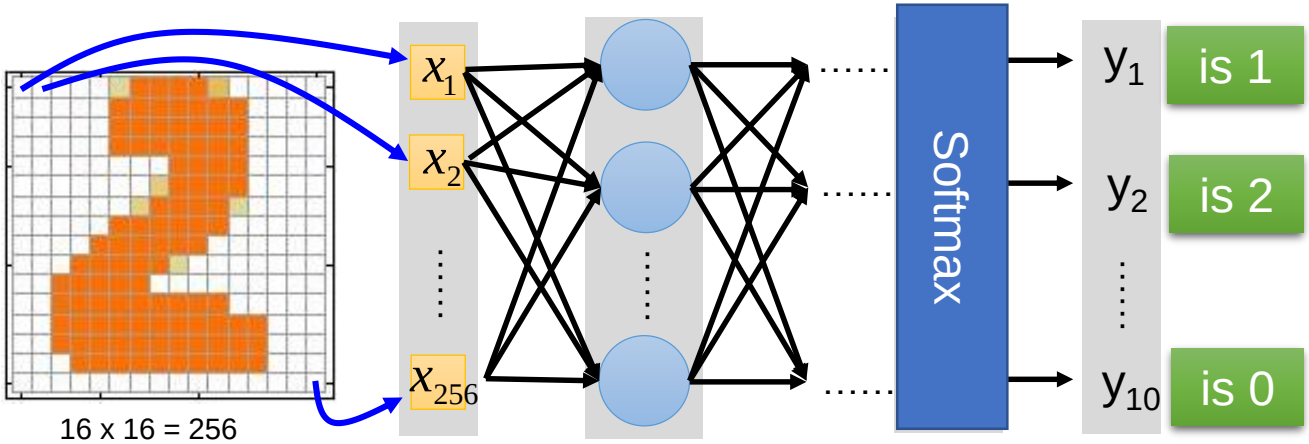
Deep Learning = Learning Representations/Features

- End-to-end learning / Feature learning / Deep learning
  - Trainable features (or kernel) + trainable classifier



[Lowe 1999, 2004], [Sivic & Zisserman 2003], [Dalal & Triggs 2005], [Grauman & Darrell 2005]
[Lazebnik et al 2006], [Perronnin & Dance 2007], [Yang et al 2009], [Jégou et al 2010], ……

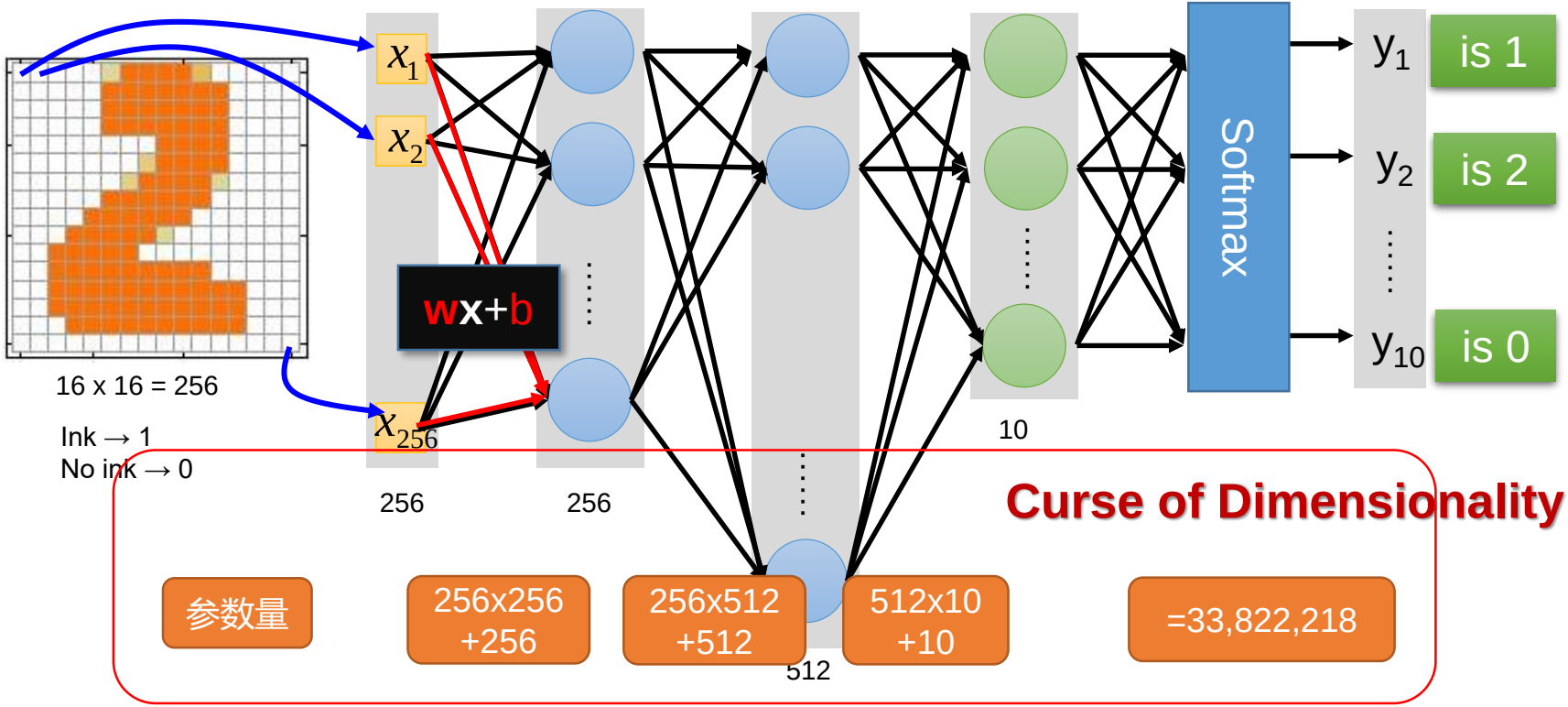
Neural Networks for Image Classification



16 x 16 = 256

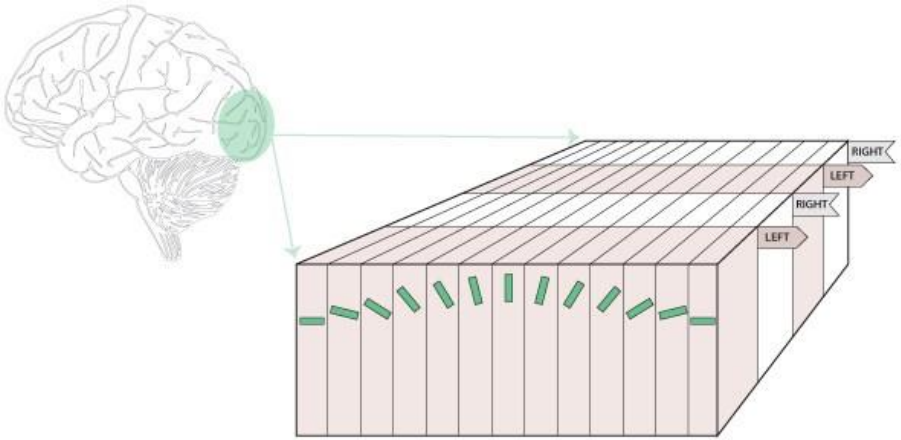
Ink → 1
No ink → 0

Fully Connected Network



The Mammalian Visual Cortex is Hierarchical

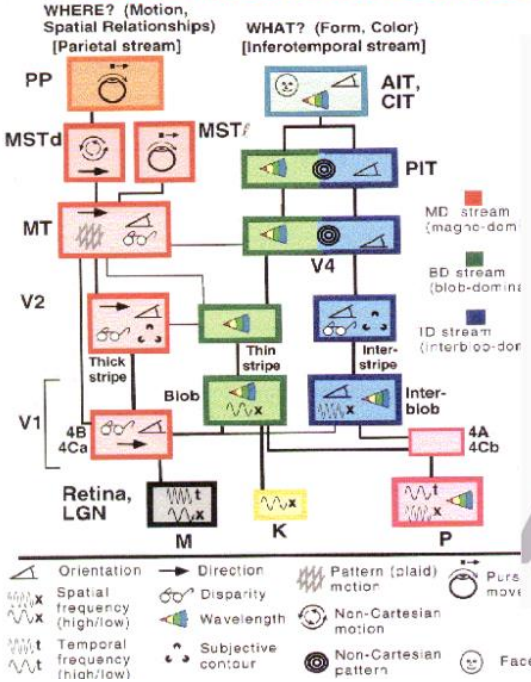
- The ventral (recognition) pathway in the visual cortex has multiple stages
- Retina - LGN - V1 - V2 - V4 - PIT - AIT ....

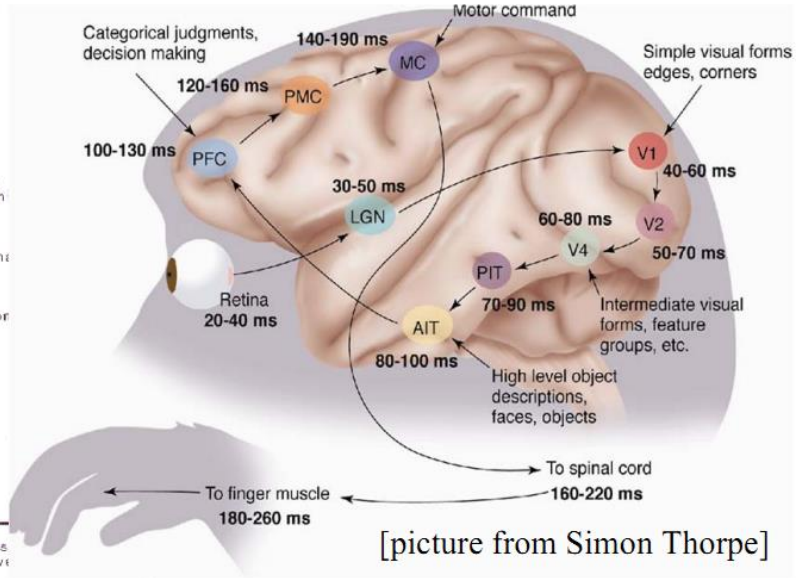


Hubel & Wiesel
Nobel Prize 1981

The Mammalian Visual Cortex is Hierarchical

- The ventral (recognition) pathway in the visual cortex has multiple stages
- Retina - LGN - V1 - V2 - V4 - PIT - AIT ....
- Lots of intermediate representations

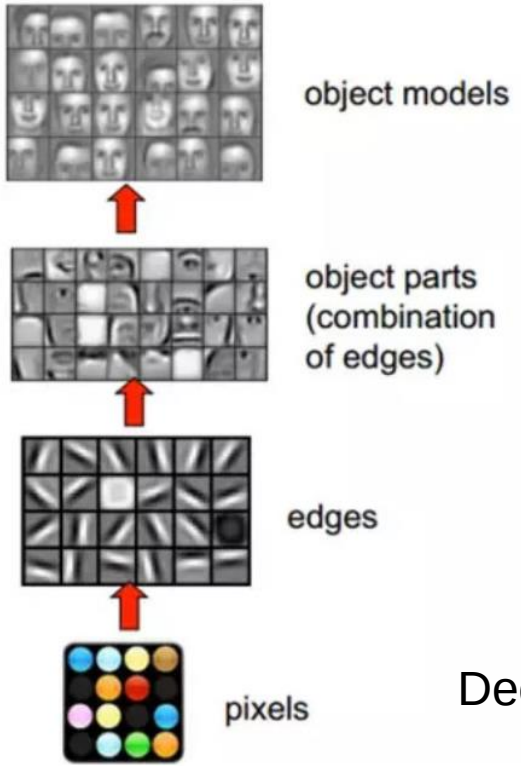


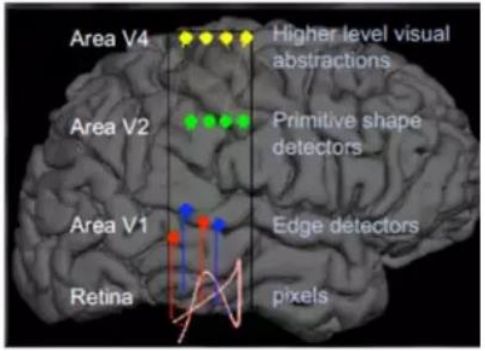


[picture from Simon Thorpe]

[Gallant & Van Essen]

Convolutional Neural Networks





Deep learning is learning hierarchical feature

Convolutional Neural Networks

- 部署于很多实际应用中
  - 图像识别、语音识别、图像标记（谷歌、百度…）

Image Classification

Image Retrieval

Object Detection

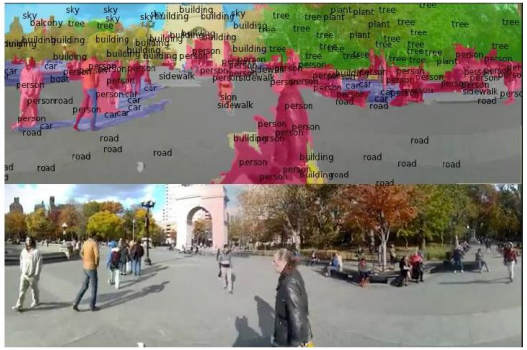
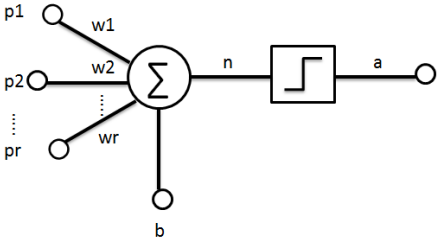


Image Segmentation

Fully Connected Network: Curse of Dimensionality

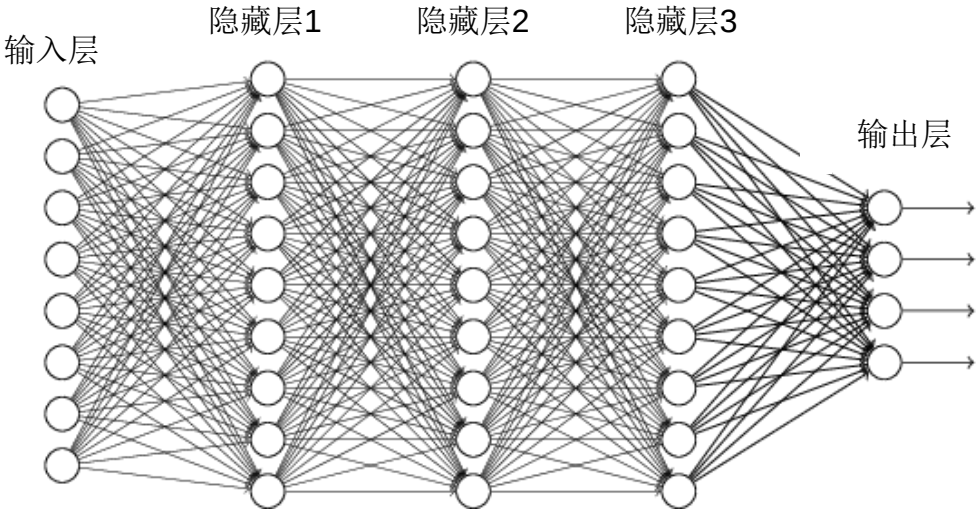
- **以200x200图像为例**
  - **全连接网络**
    - 输入层：40000维
    - 隐藏层：400000 神经元
    - 参数数目：160亿参数

内存、计算量巨大,训练困难



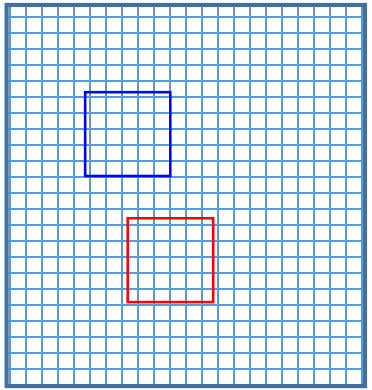
单个神经元

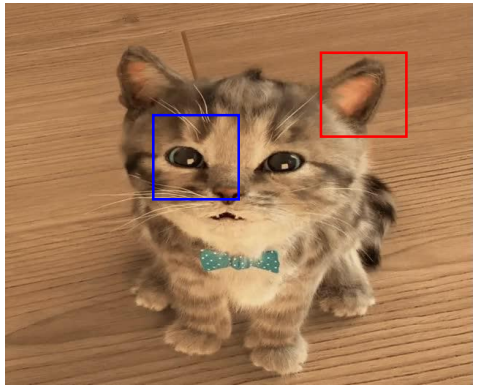
f(wx+b)



卷积神经网络CNN

- 图像数据特性
  - 矩阵数据：空间结构化
  - 相邻数据相关

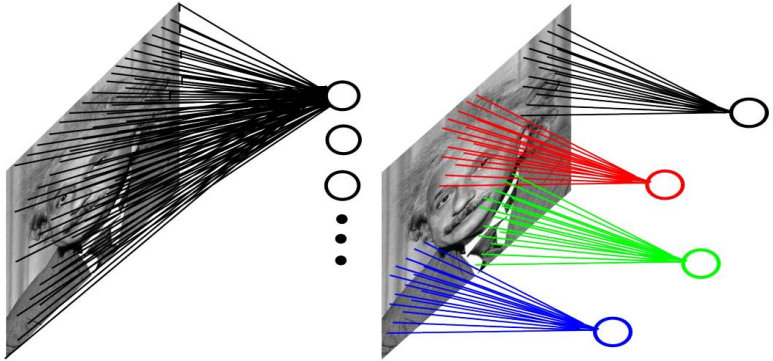




全连接神经网络：维度灾难

- 以200x200图像为例
  - **全连接网络**: 400000神经元,160亿参数
  - **局部连接网络**：利用图像区域的局部相关特性
    - 400000神经元
    - 每个神经元**10x10局部区域**：101个参数
    - 共~4,000,000个参数
    - 只捕捉到特定位置的局部信息

模型尺寸压缩了**4000**倍



图像处理---线性滤波器

- 图像平滑：去噪

平均

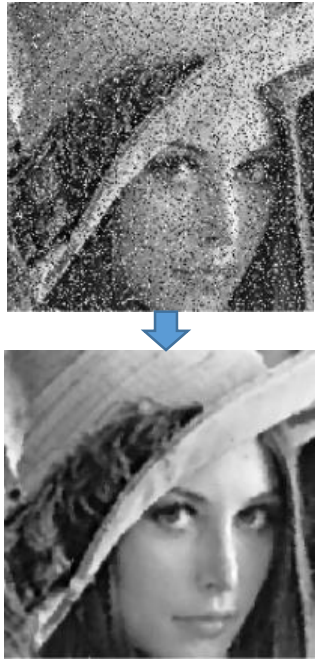
$$\frac{1}{K^2}\begin{bmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & \cdots & 1 \\ \vdots & \vdots & 1 & \vdots \\ 1 & 1 & \cdots & 1 \end{bmatrix}$$

双线性

$$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

高斯

$$\frac{1}{256}\begin{bmatrix} 1 & 4 & 6 & 4 & 1 \\ 4 & 16 & 24 & 16 & 4 \\ 6 & 24 & 36 & 24 & 6 \\ 4 & 16 & 24 & 16 & 4 \\ 1 & 4 & 6 & 4 & 1 \end{bmatrix}$$



图像处理---线性滤波器

- 图像平滑：去噪
  - 平均, 双线性, 高斯
- 边缘提取
  - Sobel

$$\frac{1}{8}\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

输入图像

x-方向

y-方向

边缘图像

图像处理---线性滤波器

- 图像平滑：去噪
  - 平均, 双线性, 高斯

- 边缘提取
  - Sobel

$$\frac{1}{8}\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$

- 角点提取

$$\frac{1}{4}\begin{bmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{bmatrix}$$

卷积神经网络

1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 图像

1	-1	-1
-1	1	-1
-1	-1	1

Filter 1

Matrix

-1	1	-1
-1	1	-1
-1	1	-1

Filter 2

Matrix

卷积神经网络需要学习
每个卷积核（滤波器）

每个滤波器检测一个小区域特征(3 x 3).

卷积神经网络

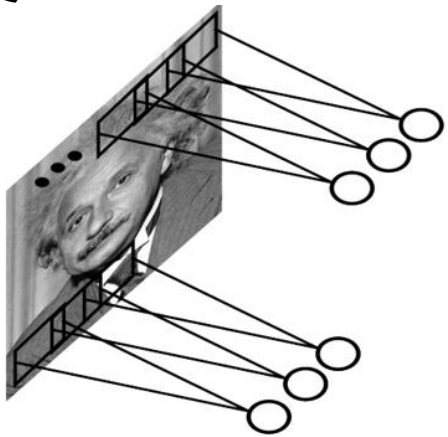
- 局部连接、权值共享
  - 在一个区域内提取的特征表达也可以应用于图像其他区域
  - 所有神经元共享权值
  - 平移等变
    - 当输入区域移动时，输出也会移动，但响应值不变
  - 卷积操作
    - 线性滤波器（学习权值）
    - 非线性：ReLu

$$A_{ij} = \sum_{kl} W_{kl} X_{i+k,j+l}$$

$$Z_{ij} = \max(0, A_{ij})$$

Sobel滤波

$$\frac{1}{8}\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}$$



- 滤波后的“图像”Z被称为“特征图”

卷积神经网络

- 局部连接、权值共享
  - 在一个区域内提取的特征表达也可以应用于图像其他区域
  - 所有神经元共享权值
  - 平移等变
    - 当输入区域移动时，输出也会移动，但响应值不变
  - 卷积操作
    - 线性滤波器（学习权值）
    - 非线性：ReLu

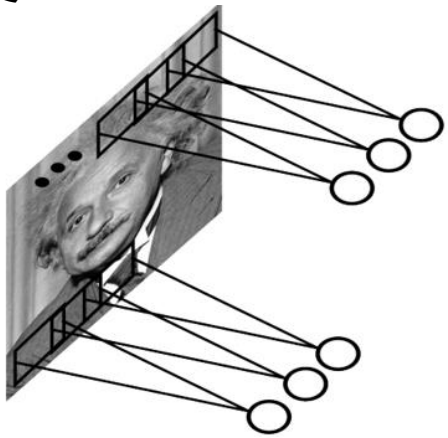
$$A_{ij} = \sum_{kl} W_{kl} X_{i+k,j+l}$$

$$Z_{ij} = \max(0, A_{ij})$$

- 滤波后的“图像”Z被称为“特征图”

以200x200图像为例：

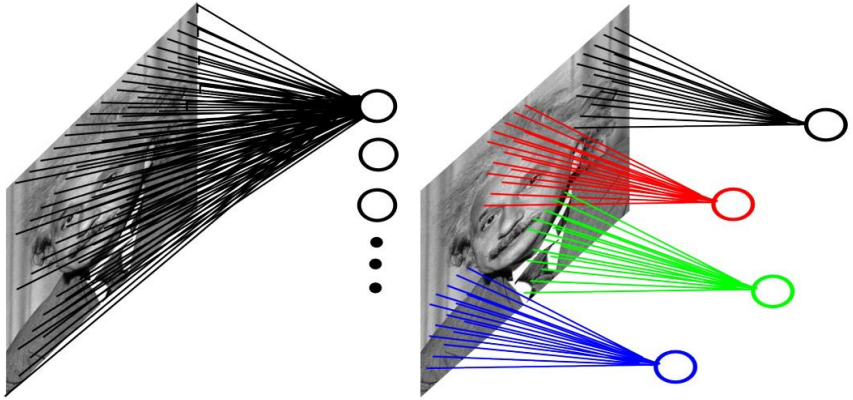
- 10个200x200的特征图，每个滤波器大小10x10，共1000个参数
- 相当于400000局部连接的神经元（10x10局部区域）



卷积神经网络

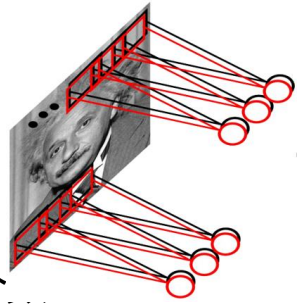
- 以200x200图像为例
  - **全连接网络**:400000神经元,160亿参数
  - **局部连接网络**：利用图像区域的局部相关特性
    - 400000神经元
    - 每个神经元10x10局部区域：100个参数
    - 共4,000,000个参数
    - 只捕捉到特定位置的局部信息
  - **卷积神经网络:**
    - **权值共享，特征图**
    - 10个200x200的特征图
    - 每个滤波器大小10x10
    - 共1000个参数

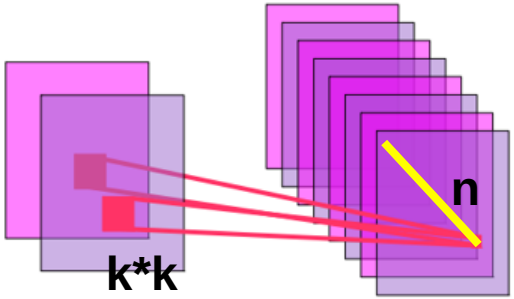
提取图像局部特征
极大幅度压缩网络参数
提高网络的稳定性



卷积神经网络：多卷积核

- 不同卷积核在每个局部区域检测不同特征
- 作用在同一个区域不同神经元的响应组成了这个局部区域的特征向量
- 多个卷积核输出了一个3D矩阵，每层是一个2D特征图





多卷积核

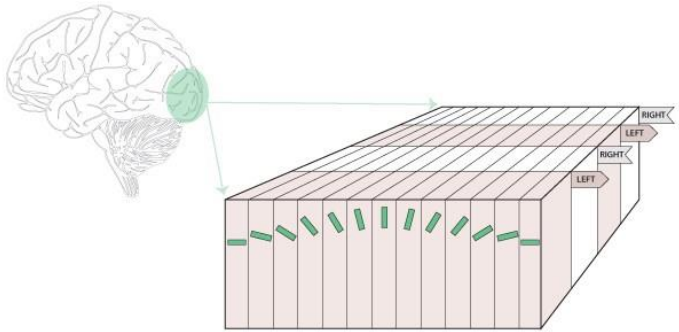
- 一个卷积层参数个
  - c：输入图像的通道数（3 for RGB）
  - k*k: 卷积核大小，神经元的感知野大小
  - n：神经元个数，也是特征图个数

局部连接如何感知复杂的全局信息？

早期的多层级特征模型

- Hubel & Wiesel: 视觉系统
- 视觉皮层中的信息处理
  - 视皮质细胞只对视网膜上的图像的某些特定细节有反应，另一个令人着迷的特点则是这些细胞似乎会自然映射到不同角度。

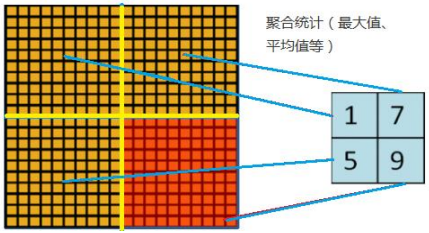
1981年获诺贝尔生理学/医学奖

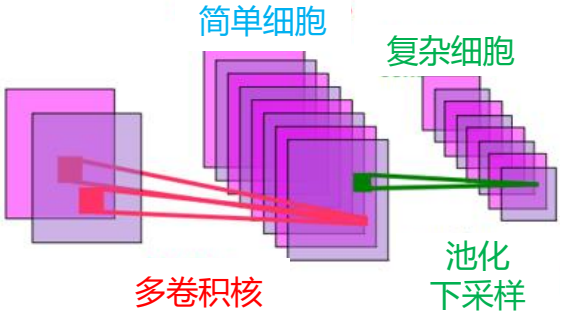


早期的多层级特征模型

- Hubel & Wiesel: 视觉系统
- 视觉皮层中的信息处理
  - 视皮质细胞只对视网膜上的图像的某些特定细节有反应，另一个令人着迷的特点则是这些细胞似乎会自然映射到不同角度。
- **简单细胞**
  - 检测局部特征（方向、亮度等）---多卷积核
- **复杂细胞**
  - 将简单细胞的响应在一个视觉邻域内汇聚（pool），以形成越来越复杂的感受野

1981年获诺贝尔生理学/医学奖





卷积网络模型
(多级Hubel-Wiesel系统)

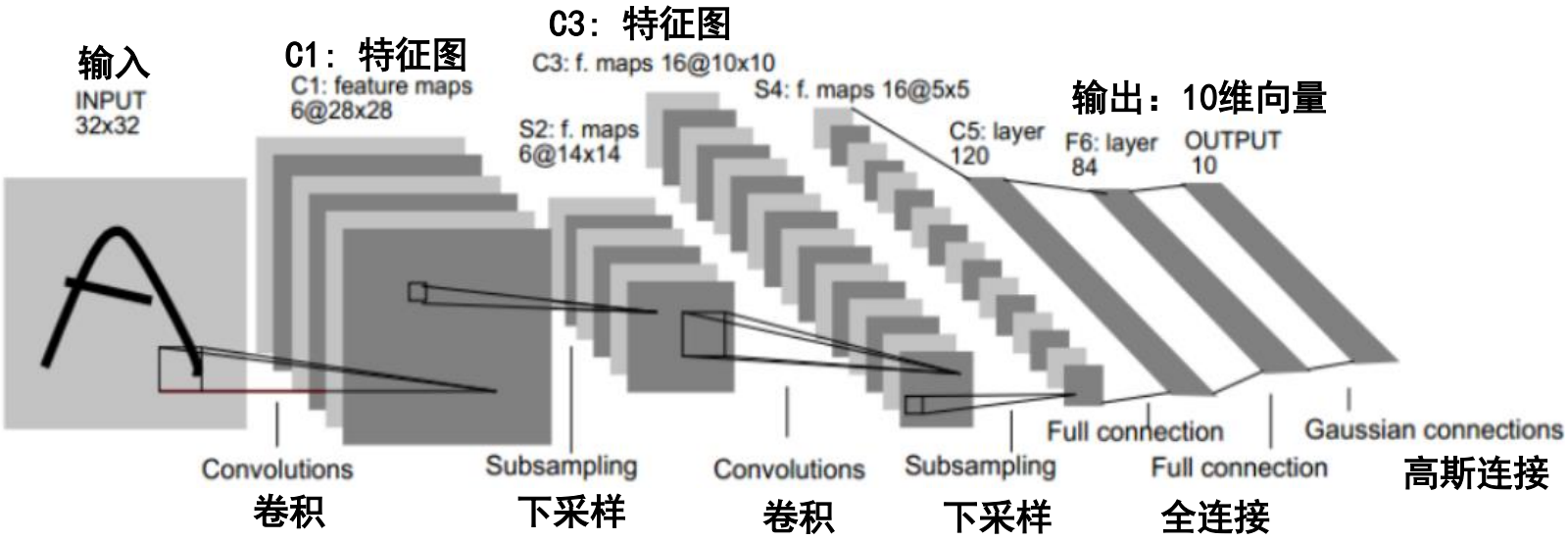
- LeNet-5[LeCun,1994]
  - 手写数字识别

Yann LeCun

Director of AI Research, Facebook
New York University

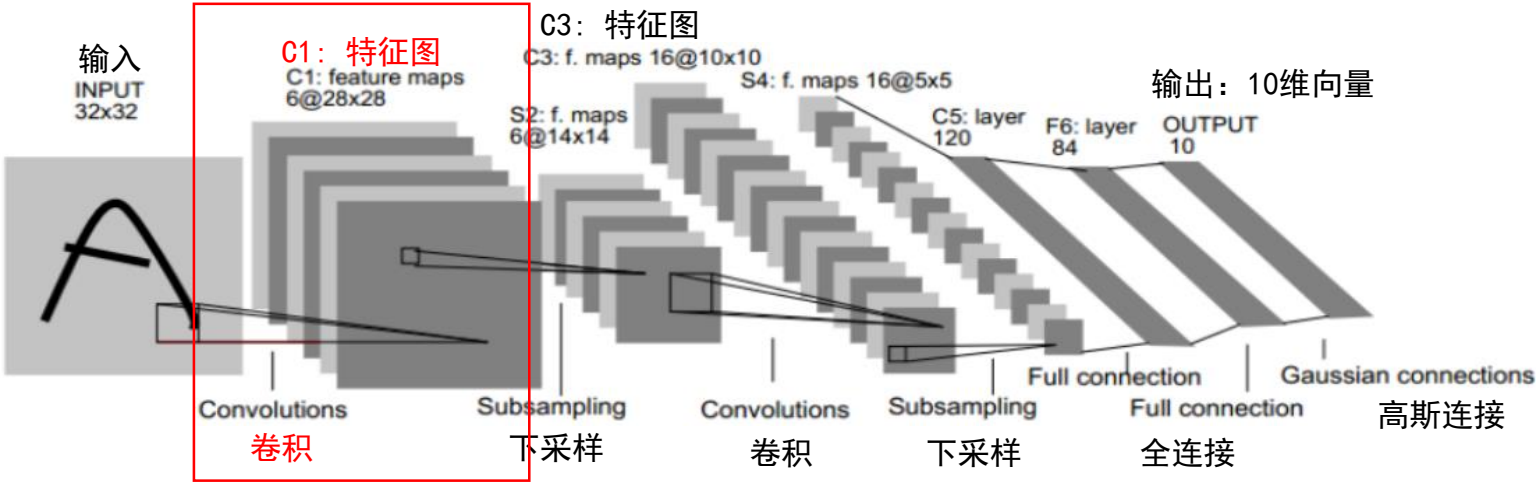
卷积网络模型
(多级Hubel-Wiesel系统)

- LeNet-5[LeCun,1994]
  - 手写数字识别
  - 3个卷积层、2个下采样层、2个全连接层。
  - 32x32 -->10 class



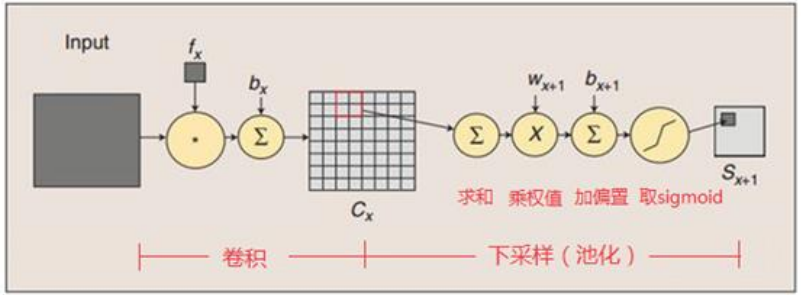
卷积网络模型
LeNet-5[LeCun,1994]

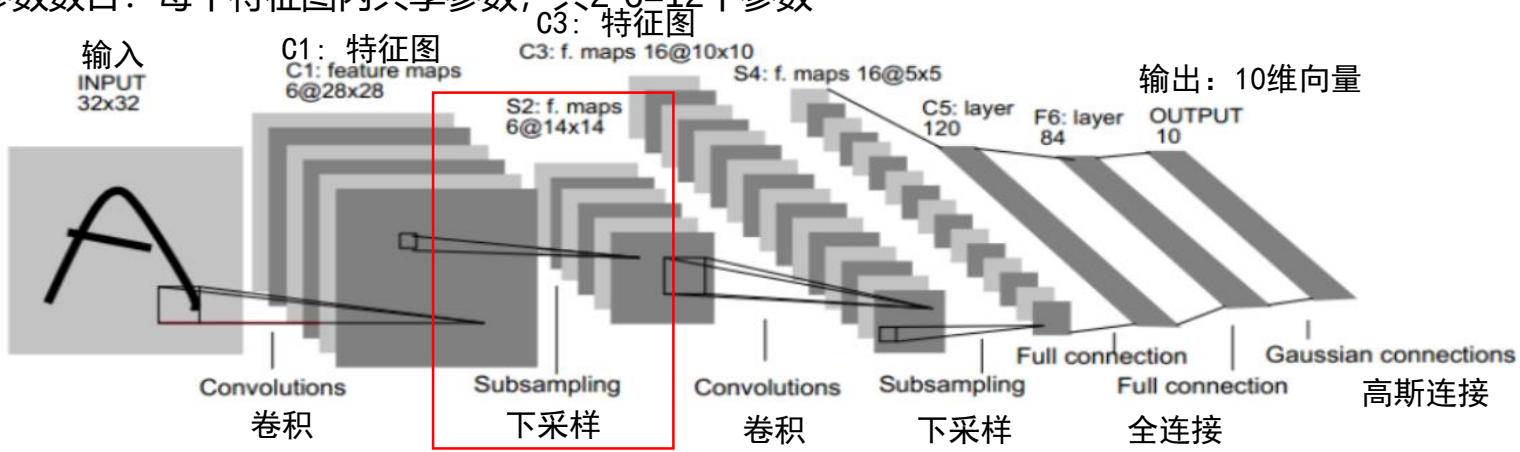
- **C1层：卷积层**
  - 6个特征图，每个卷积核大小5x5
    - 每个特征图中每个神经元与5x5的邻域相连，得到28x28的特征图
  - 每个卷积神经元的参数数目：25个unit参数+1个bias参数 **wx**+b
  - 连接数目： (5*5+1)*6*(28*28)=122,304个连接
  - 参数数目：每个特征图内共享参数，共（5*5+1）*6=156个参数



卷积网络模型
LeNet-5[LeCun,1994]

- **S2层：下采样层/池化层**
  - 池化单元大小2x2，得到6个14x14的特征图。
    - 每个特征图中的每个单元与C1特征图中的一个2x2区域相连不重叠。
  - S2层每个单元的4个输入相加，用sigmoid（wx+b）计算响应
  - 连接数目: (2*2+1)*6*(14*14)=5880个
  - 参数数目：每个特征图内共享参数，共2*6=12个参数





卷积神经网络 ---- 池化

- 最大池化 Max-pooling
- 平均池化 Average-pooling

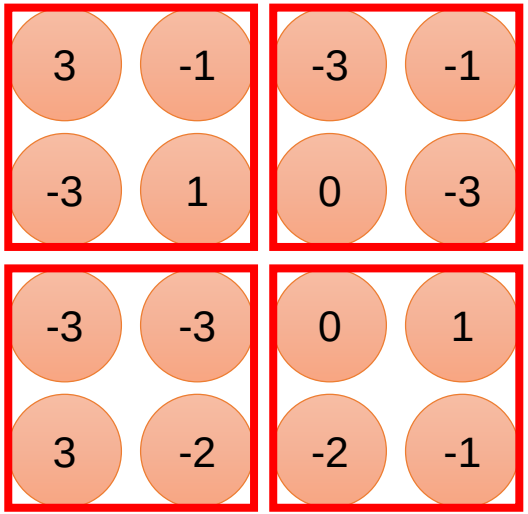
CNN – 最大池化

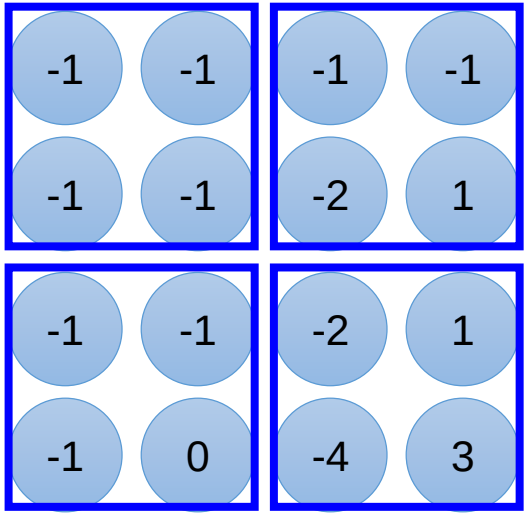
1	-1	-1
-1	1	-1
-1	-1	1

卷积核 1

-1	1	-1
-1	1	-1
-1	1	-1

卷积核 2





CNN –最大池化

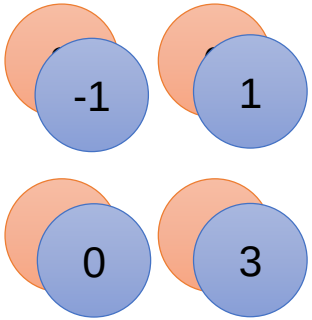
1	0	0	0	0	1
0	1	0	0	1	0
0	0	1	1	0	0
1	0	0	0	1	0
0	1	0	0	1	0
0	0	1	0	1	0

6 x 6 image

卷积
Conv

最大池化
Max
Pooling

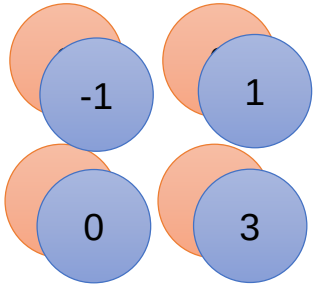
新的图像
尺寸减小



2 x 2 image

每个卷积核
对应一个通道

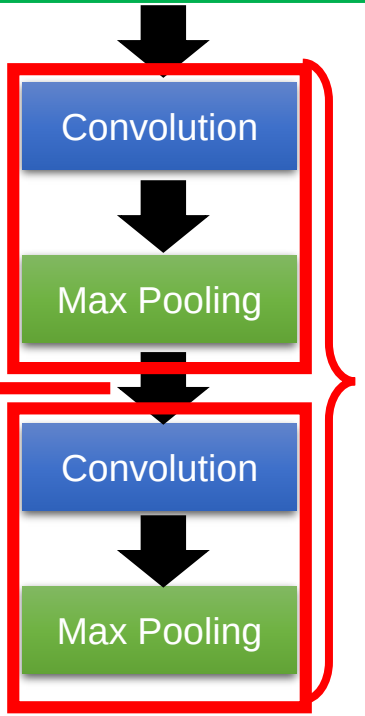
整体卷积神经网络



新的图像/
特征图像

尺寸变小

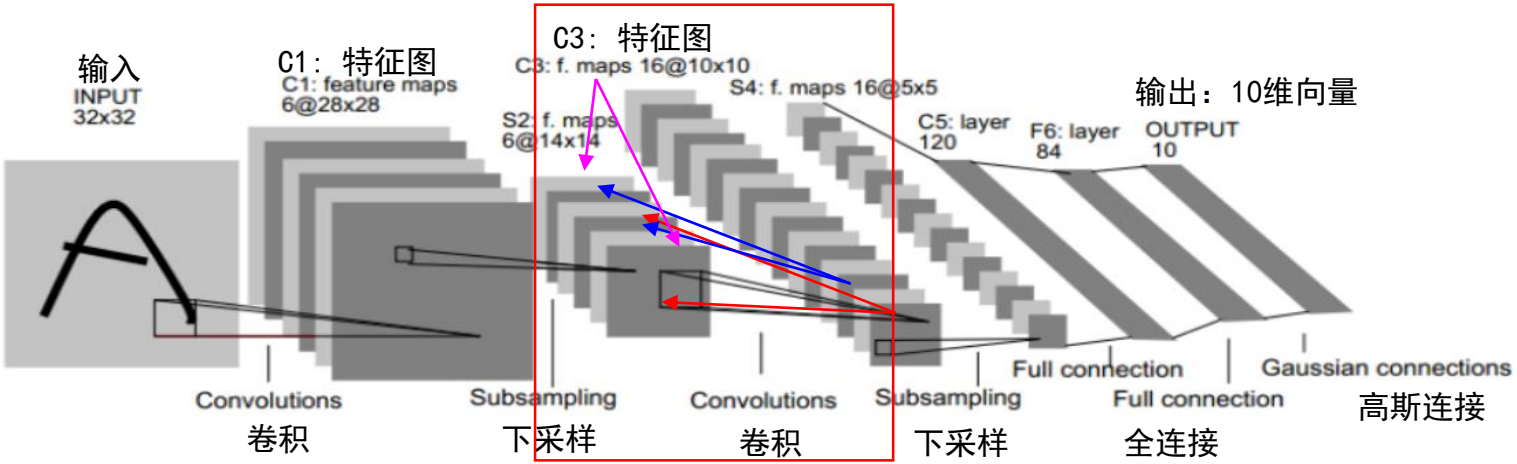
通道数就是滤波器数目



可以多次反复

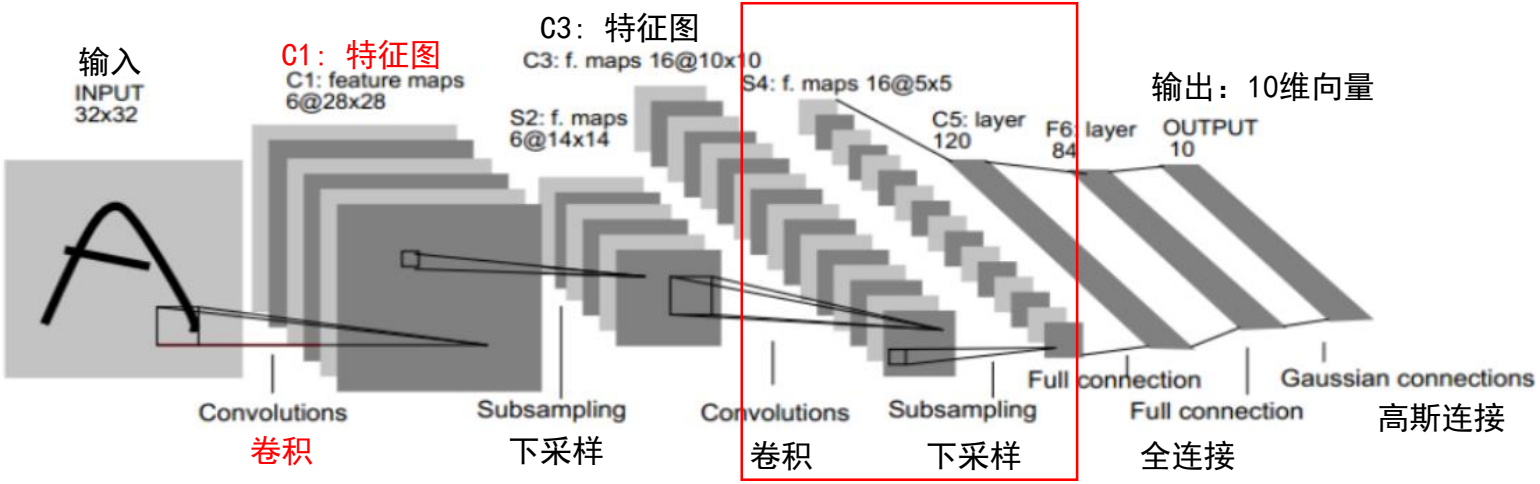
卷积网络模型
LeNet-5[LeCun,1994]

- **C3层：卷积层**
  - 16个卷积核,卷积核大小5×5，
  - 每个特征图中每个神经元与S2层多个特征图5x5的邻域相连，得到10×10的特征图
  - 参数数目：（5×5×3+1）×6 +（5×5×4+1）×9 +5×5×6+1 = 1516
  - 连接数目：10×10×1516=151600

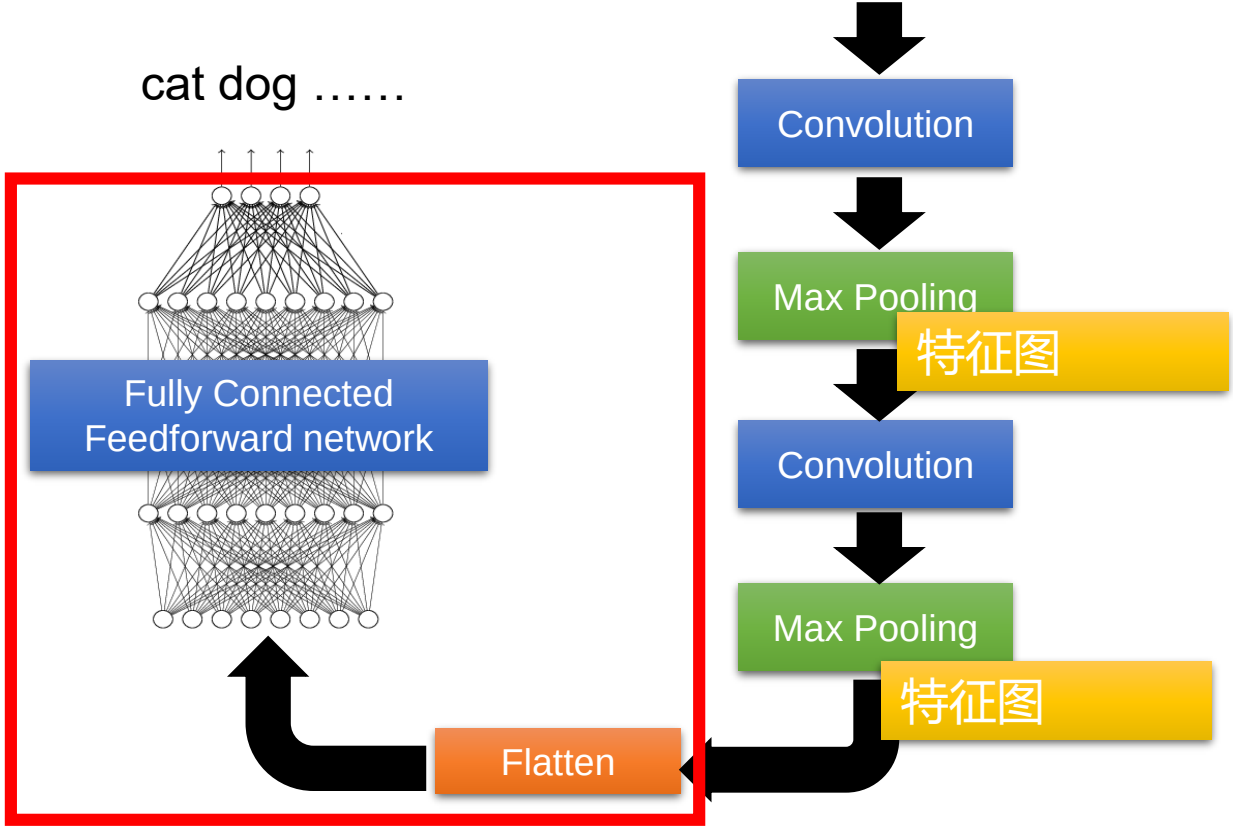


卷积网络模型
LeNet-5[LeCun,1994]

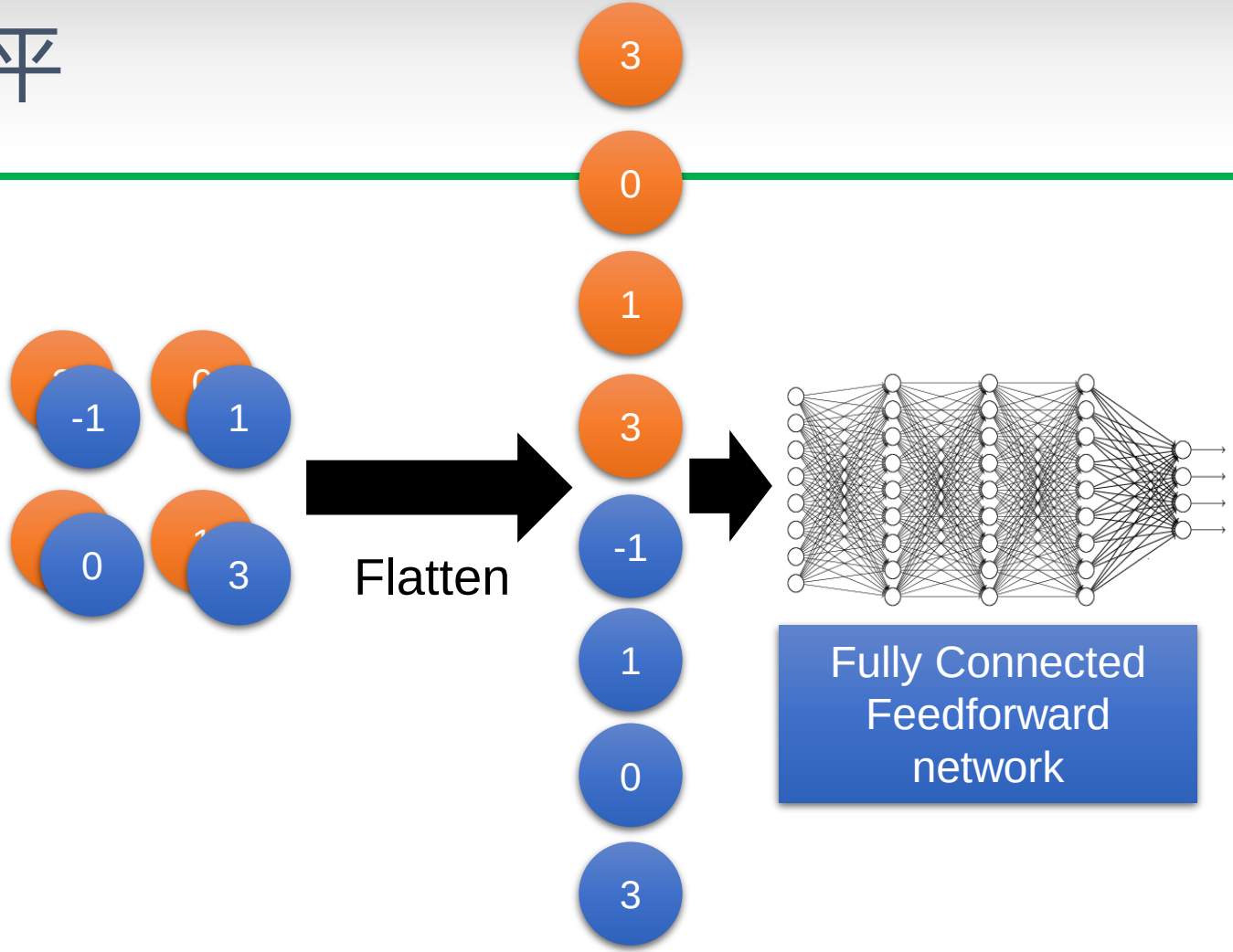
- **S4层：下采样层/池化层**
  - 16个特征图，池化单元大小2x2，特征图大小5x5
  - 连接数目: (2*2+1)*16*(5*5)=2000个
  - 参数数目：每个特征图内共享参数(w,b)，共2*16=32个参数



整体卷积神经网络

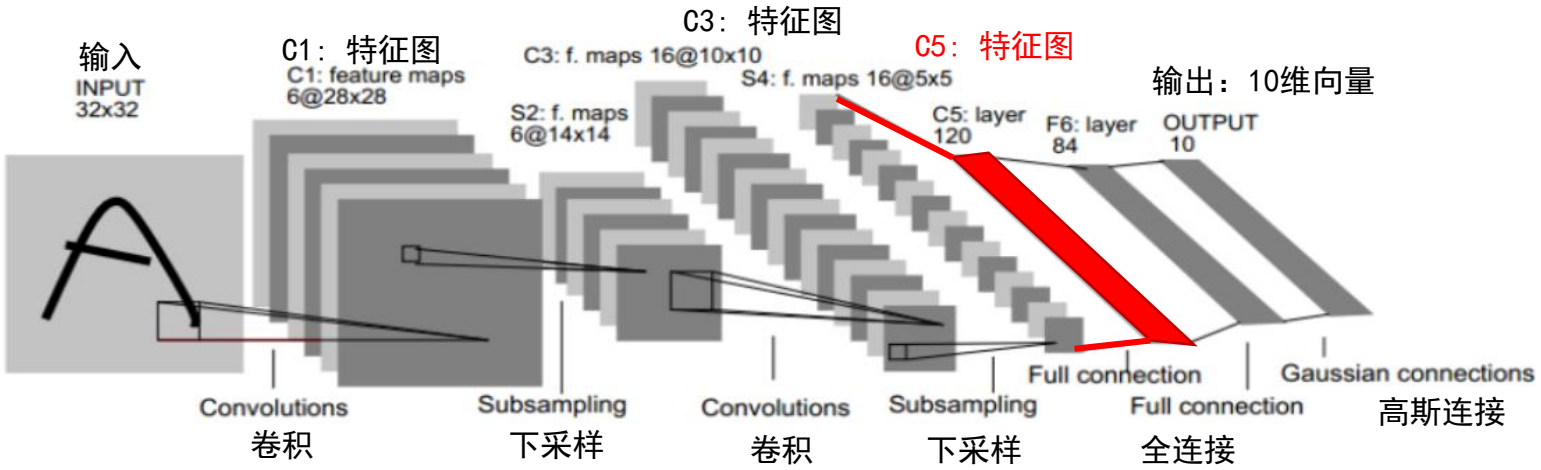

cat dog ……
Convolution
Max Pooling
特征图
Convolution
Max Pooling
特征图
Flatten
Fully Connected
Feedforward network

展平


-1
1
0
3
Flatten
3
0
1
3
-1
1
0
3
Fully Connected Feedforward network

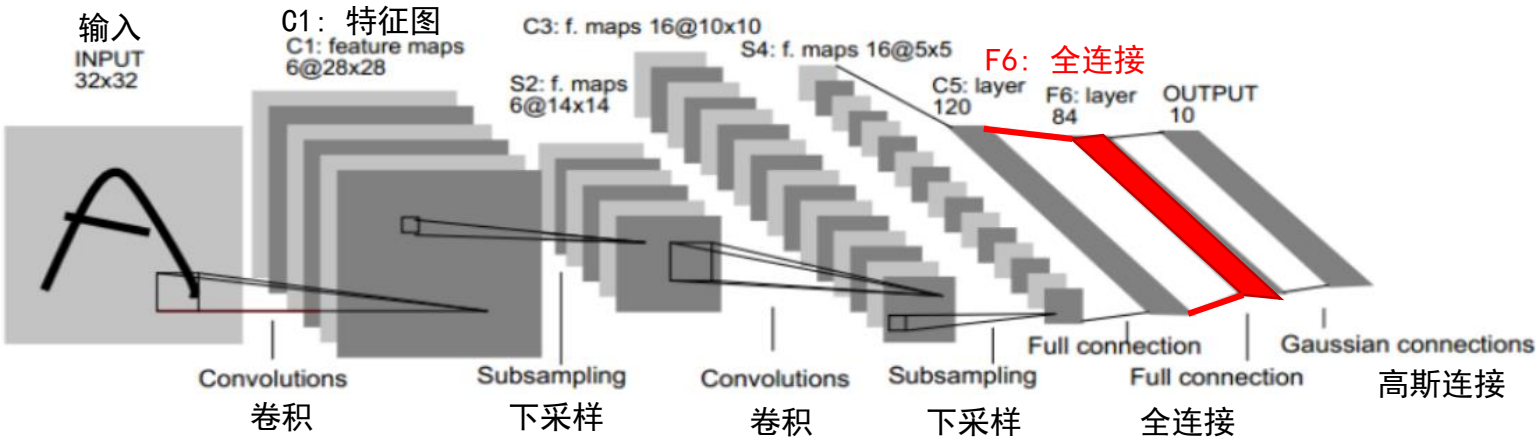
卷积网络模型
LeNet-5[LeCun,1994]

- **C5层：卷积层**
  - 120个神经元，看作120个特征图，每个特征图大小1×1
  - 每个单元与S4层的全部16个单元的5x5的邻域相连（S4和C5之间全连接）
  - 参数数目：（5×5×16+1）×120 = 48120
  - 连接数目=参数数目



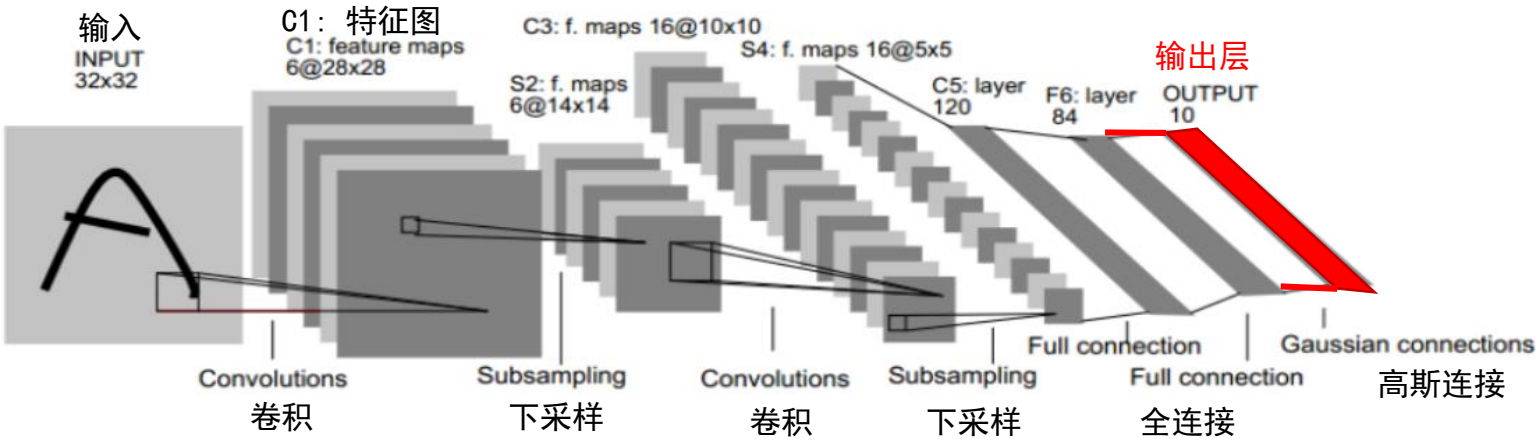
卷积网络模型
LeNet-5[LeCun,1994]

- **F6层：全连接层**
  - 84个神经元，与C5层全相连
  - F6层计算输入向量和权重向量间的点积再加偏置（**wx**+b）
  - 参数数目：（120+1）×84 =10164
  - 84维向量对应7x12的比特图



卷积网络模型
LeNet-5[LeCun,1994]

- **输出层**
  - 10个节点，分别代表数字0-9，第i的节点的数值越接近0，代表与数字i越相似。
  - 连接方式：径向基函数（RBF）$y_i = \sum_{j=1,..84}(x_j - w_{ij})^2$
  - RBF输出可以被理解为F6层配置空间的高斯分布
  - 10个神经元，与F6层全相连
  - 参数数目=连接数目: w_{ij}: 84×10 =840



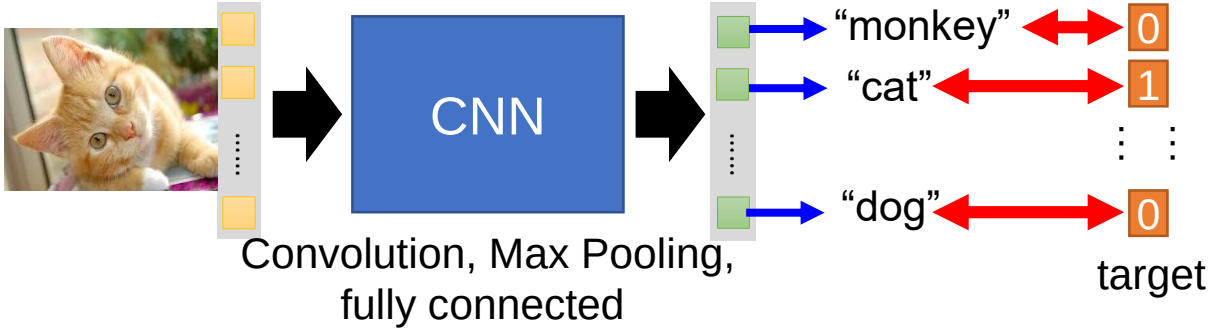
卷积神经网络

第一步: 定义

Convolutional Neural Network

第二步: 衡量函数的好坏

第三步: 选择最优的函数



卷积网络模型
LeNet-5[LeCun,1994]

- **损失函数**

Maximum Likelihood

$$E(W)=\frac{1}{P}\sum_{p=1}^{P} y_{D^p}(Z^p,W)$$

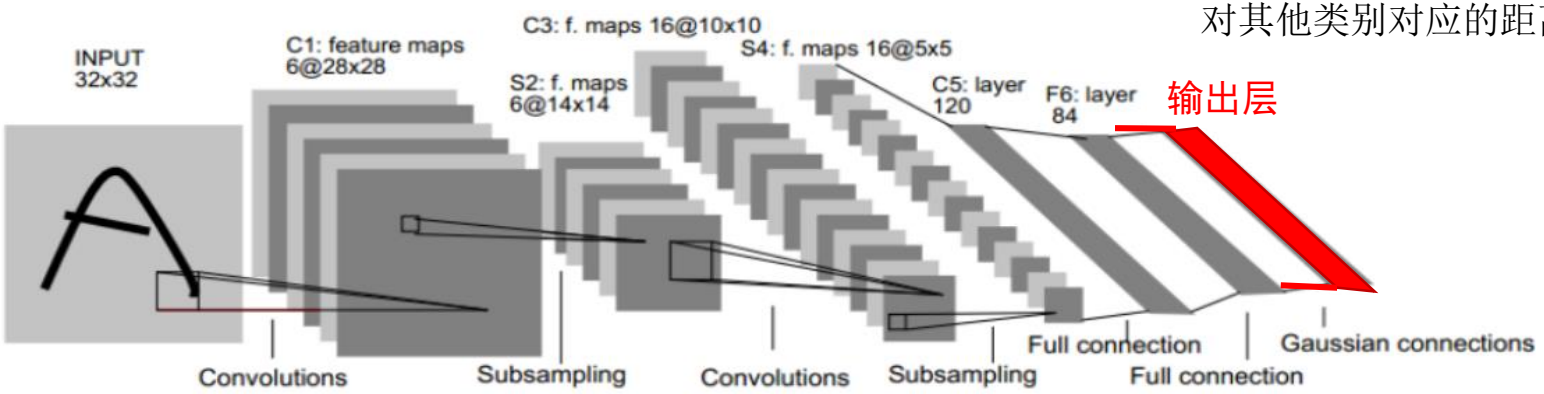
只考虑真实类别对应的向量

Maximum a Posterior

$$E(W)=\frac{1}{P}\sum_{p=1}^{P}\left(y_{D^p}(Z^p,W)\right.$$ 真实类别对应的距离尽可能小

$$\left.+\log\left(e^{-j}+\sum_i e^{-y_i(Z^p,W)}\right)\right)$$

对其他类别对应的距离尽可能大



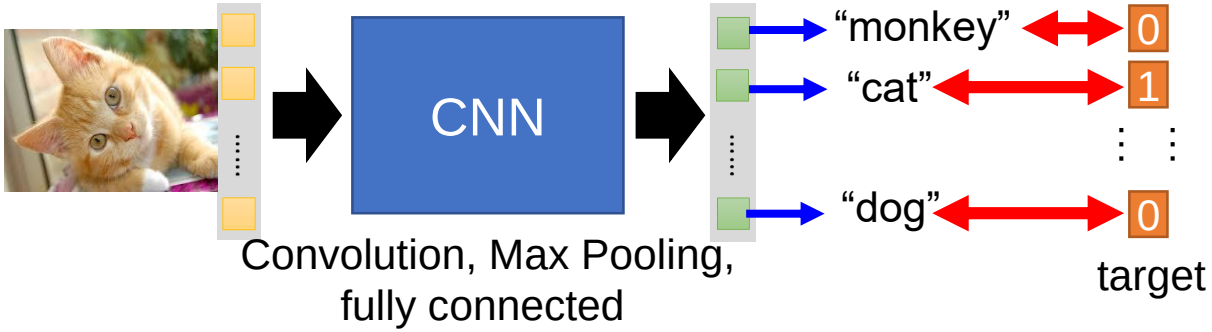
卷积神经网络

第一步: 定义

Convolutional Neural Network

第二步: 衡量函数的好坏

第三步: 选择最优的函数

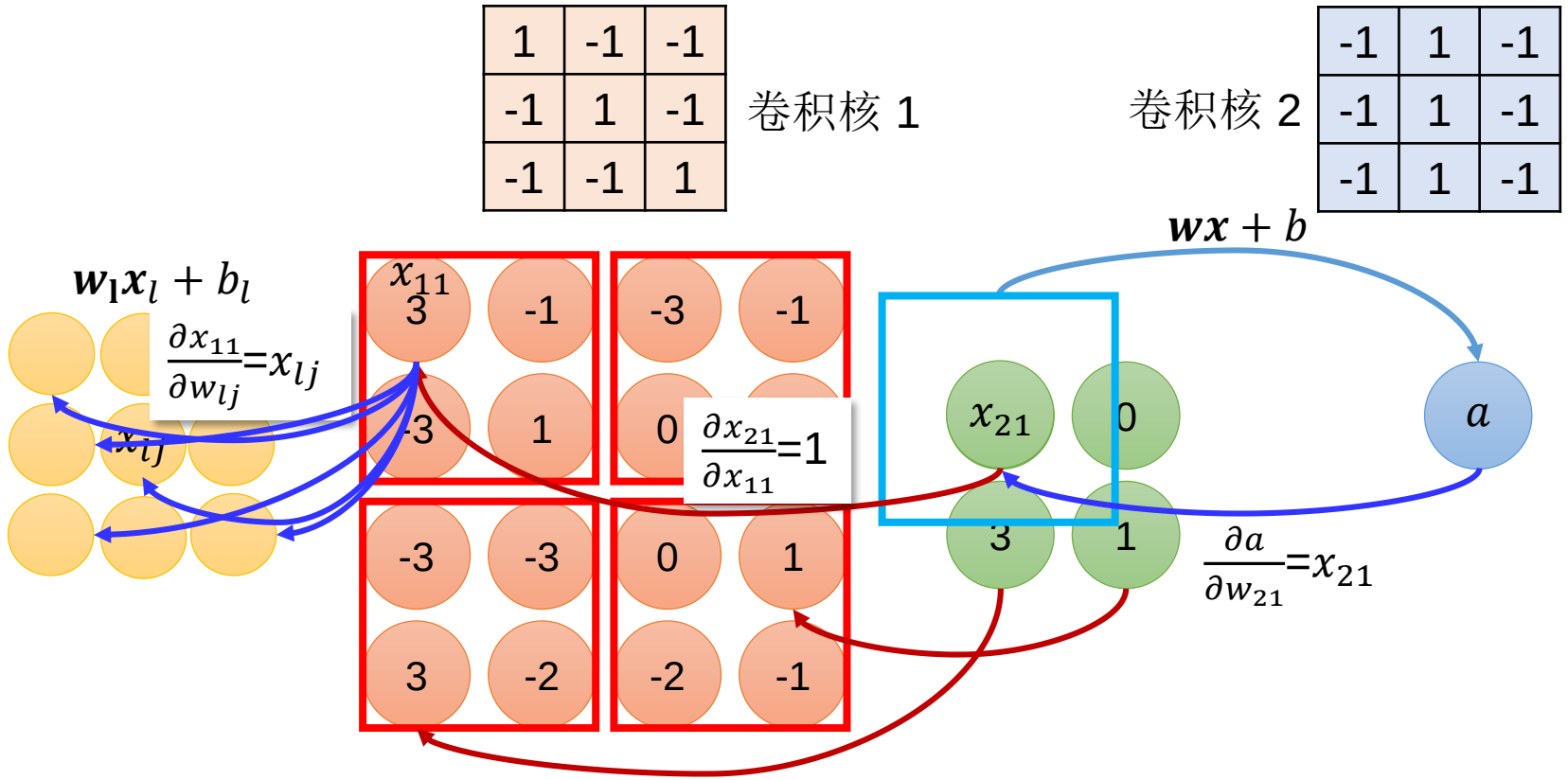


学习过程：梯度下降 ……

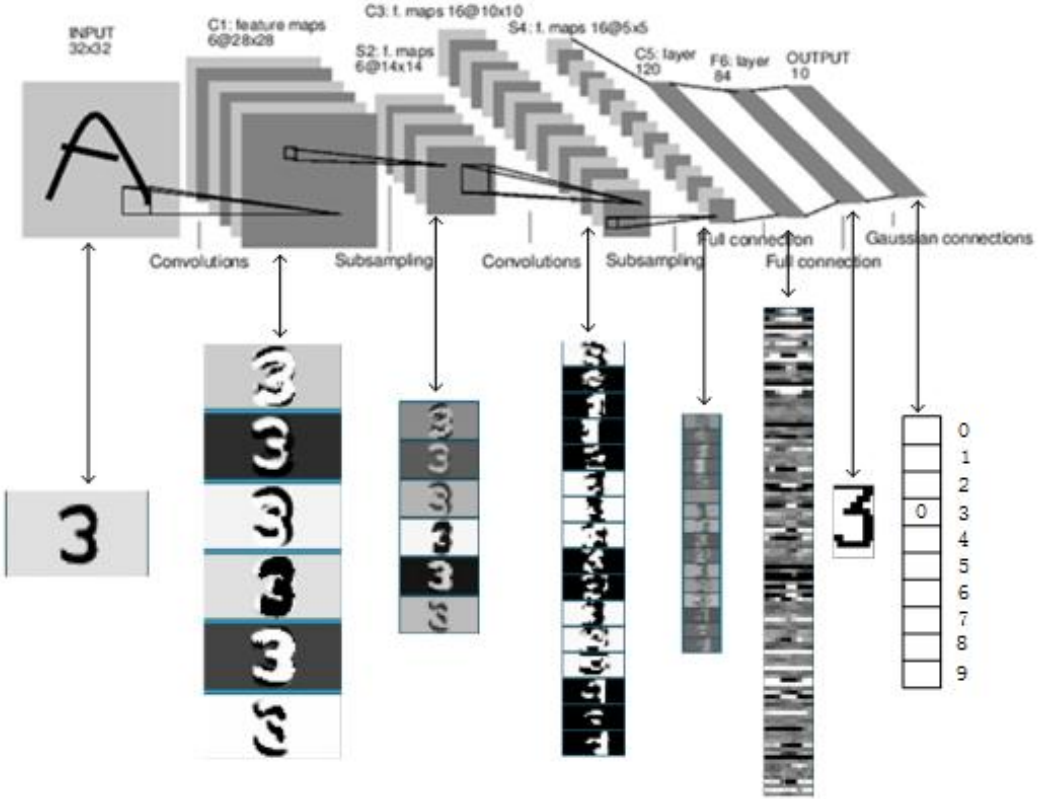
卷积神经网络反向传播

- 最大池化层的梯度反向传播

CNN – 最大池化

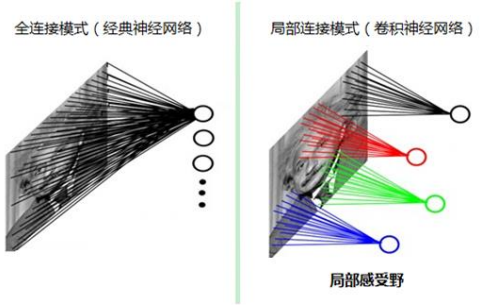

1 -1 -1
-1 1 -1
-1 -1 1
卷积核 1
-1 1 -1
-1 1 -1
-1 1 -1
卷积核 2
$wx + b$
$w_l x_l + b_l$
$\frac{\partial x_{11}}{\partial w_{lj}} = x_{lj}$
x_{lj}
x_{11}
3 -1 -3 -1
-3 1 0 1
-3 -3 0 1
3 -2 -2 -1
$\frac{\partial x_{21}}{\partial x_{11}} = 1$
x_{21} 0
3 1
a
$\frac{\partial a}{\partial w_{21}} = x_{21}$

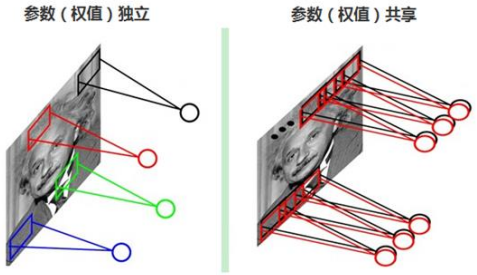
卷积网络模型
LeNet-5[LeCun,1994]

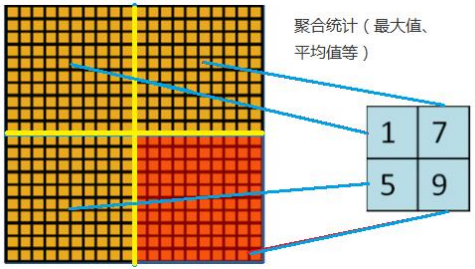


卷积神经网络模型: 总结

- 三大特点：局部感知、参数共享、池化操作
- 大大节省了计算成本（CPU允许）
- 多级Hubel-Wiesel系统



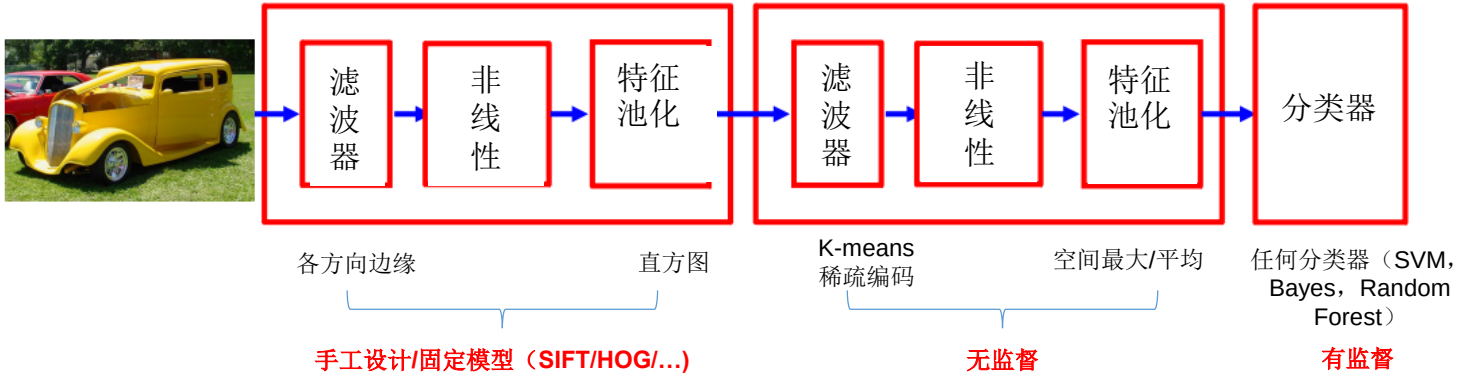




局部感知 **参数共享** **池化**

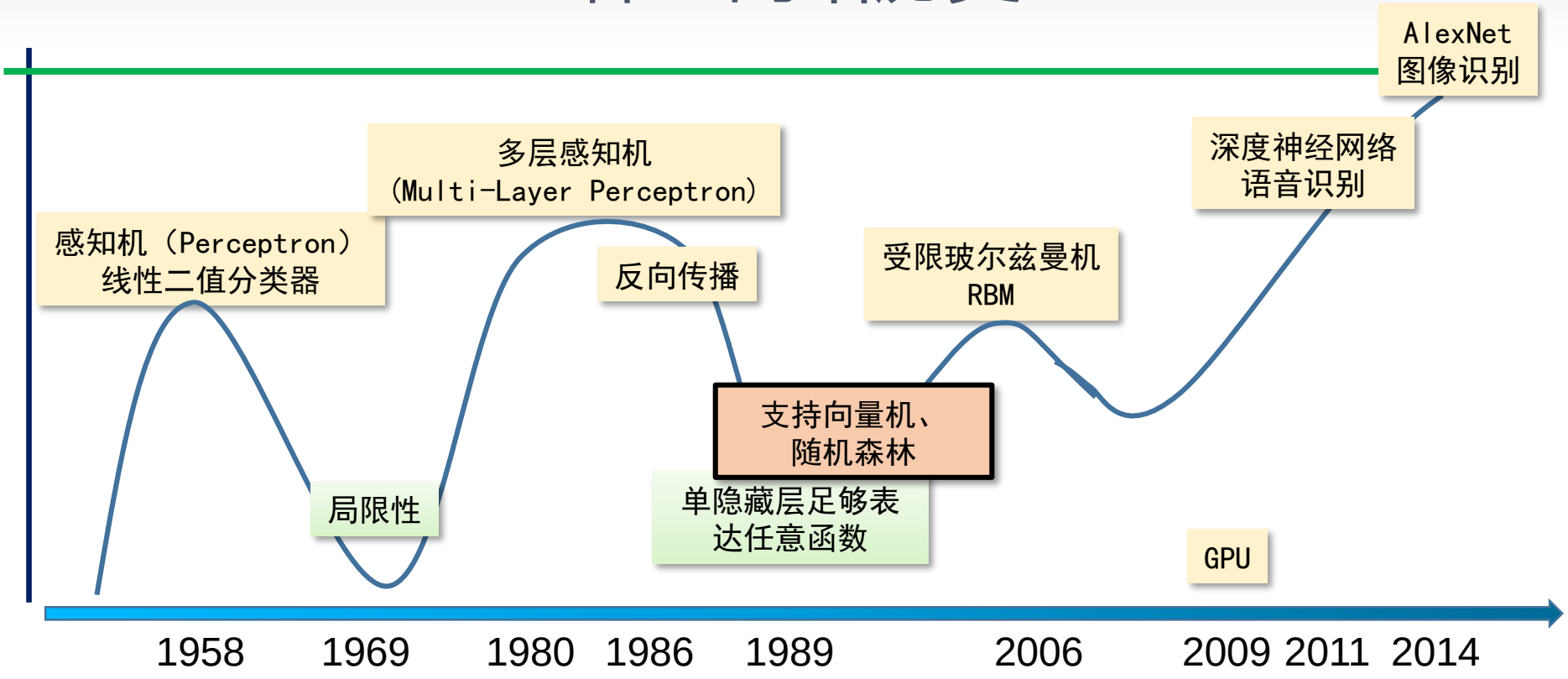
- 后续变种：AlexNet，GoogleNet, VGG,DRL等等

卷积神经网络与传统物体识别框架

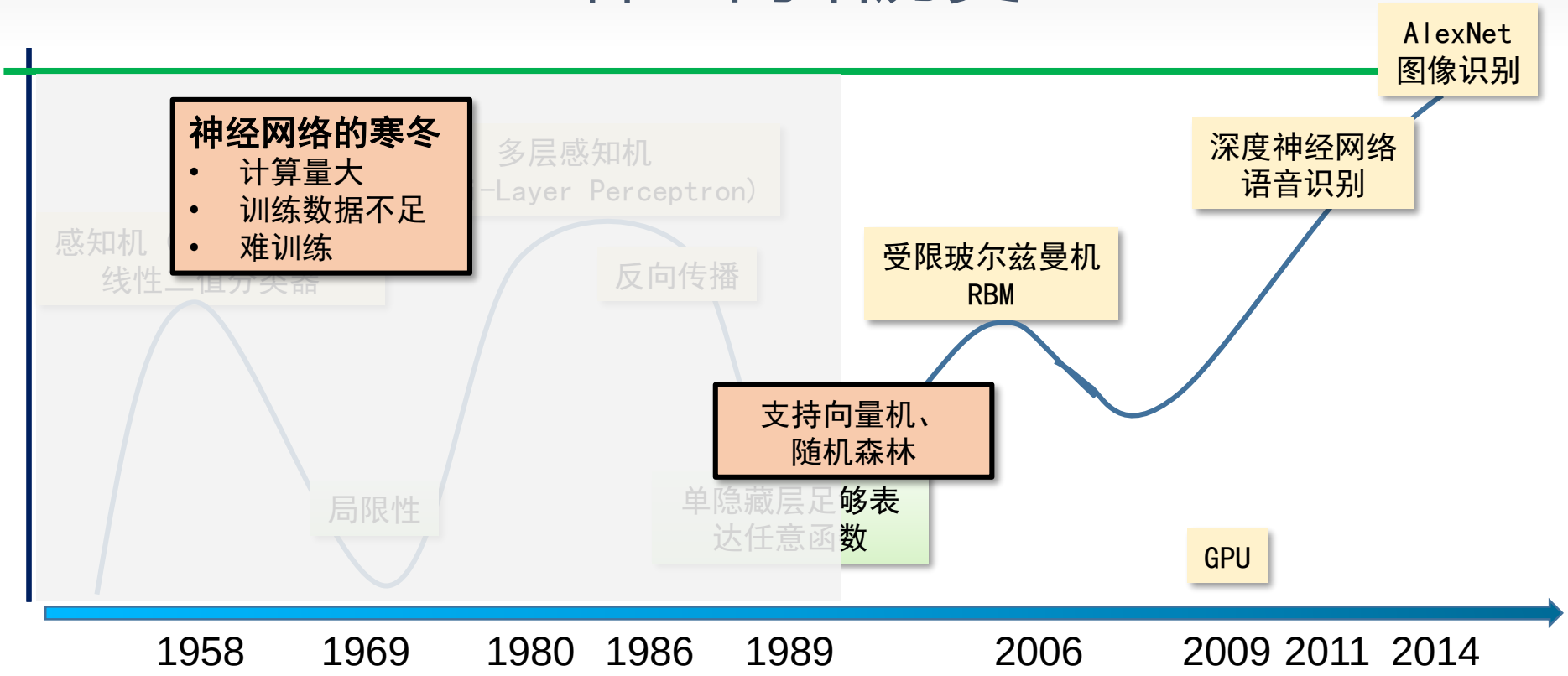


- 传统方法：手工设计局部特征+无监督的中层特征+简单分类器
  - SIFT + 矢量量化 + 金字塔池化 + SVM [Lazebnik et al CVPR 2006]
  - SIFT + 局部稀疏编码 + 金字塔池化 + SVM [Boureau et al. ICCV 2011]
  - SIFT + Fisher Vectors + 可变形部件模型 + SVM [Perronin et al 2012]

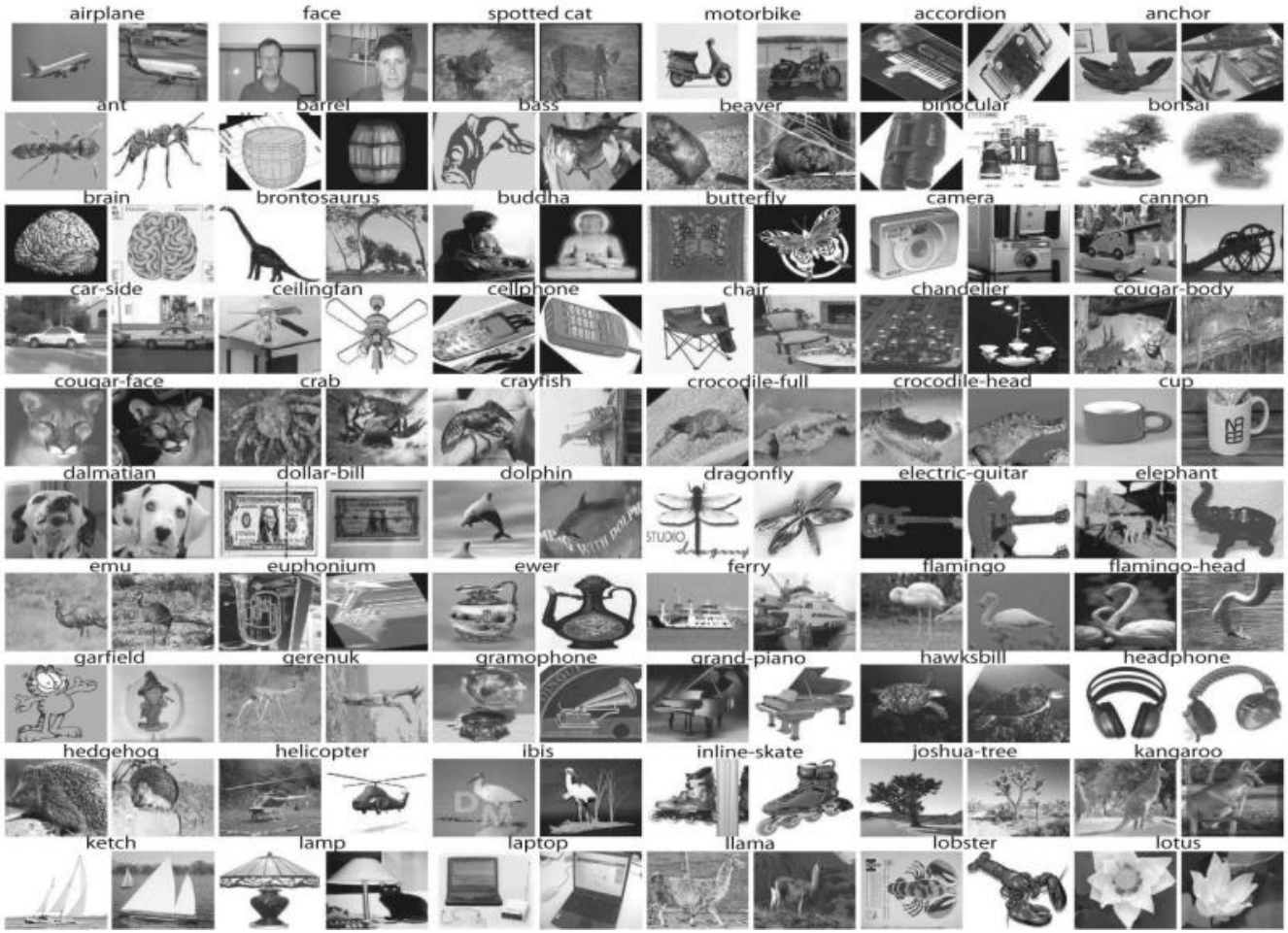
神经网络历史


AlexNet
图像识别
多层感知机
(Multi-Layer Perceptron)
深度神经网络
语音识别
感知机（Perceptron）
线性二值分类器
反向传播
受限玻尔兹曼机
RBM
支持向量机、
随机森林
局限性
单隐藏层足够表
达任意函数
GPU
1958
1969
1980
1986
1989
2006
2009
2011
2014

神经网络历史


AlexNet
图像识别
神经网络的寒冬
• 计算量大
• 训练数据不足
• 难训练
多层感知机
-Layer Perceptron)
感知机
线性二值分类器
反向传播
深度神经网络
语音识别
受限玻尔兹曼机
RBM
支持向量机、
随机森林
够表
数
局限性
单隐藏层足
达任意函
GPU
1958
1969
1980
1986
1989
2006
2009
2011
2014

IMAGENET
15,000,000 images in
22,000 categories
Baidu百科



2004年图像分类任务

101 物体类别

每类40 ~ 800 张图

图像尺寸 ~ 300 x 200

L. Fei-Fei, R. Fergus, and P. Perona, CVPR 2004

IMAGENET

www.image-net.org

22K categories and 14M images

- Animals
  - Bird
  - Fish
  - Mammal
  - Invertebrate
- Plants
  - Tree
  - Flower
- Food
- Materials
- Structures
- Artifact
  - Tools
  - Appliances
  - Structures
- Person
- Scenes
  - Indoor
  - Geological Formations
- Sport Activities

Deng, Dong, Socher, Li, Li, & Fei-Fei, 2009

IMAGENET Large Scale Visual Recognition Challenge

Steel drum

The Image Classification Challenge:

1,000 object classes

1,431,167 images

Output:
Scale
T-shirt
<u>Steel drum</u>
Drumstick
Mud turtle

Output:
Scale
T-shirt
Giant panda
Drumstick
Mud turtle

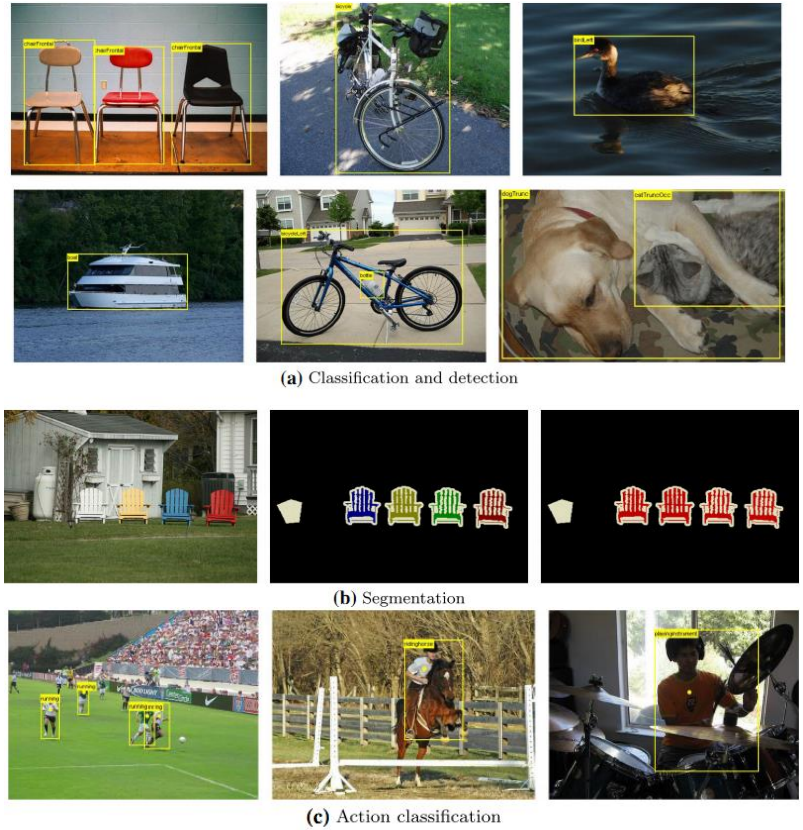
Russakovsky et al. arXiv, 2014

机器视觉任务

Visual Object Classes

PASCAL2
Pattern Analysis, Statistical Modelling and Computational Learning

- 20 classes
- 11,530 images
- 27,450 ROI annotated objects
- 6,929 segmentations



(a) Classification and detection

(b) Segmentation

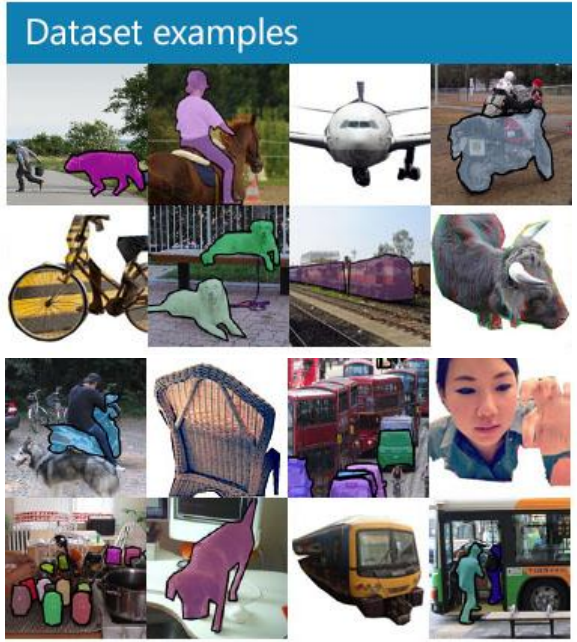
(c) Action classification

Microsoft CoCo Database

COCO is a large-scale object detection, segmentation, and captioning dataset.

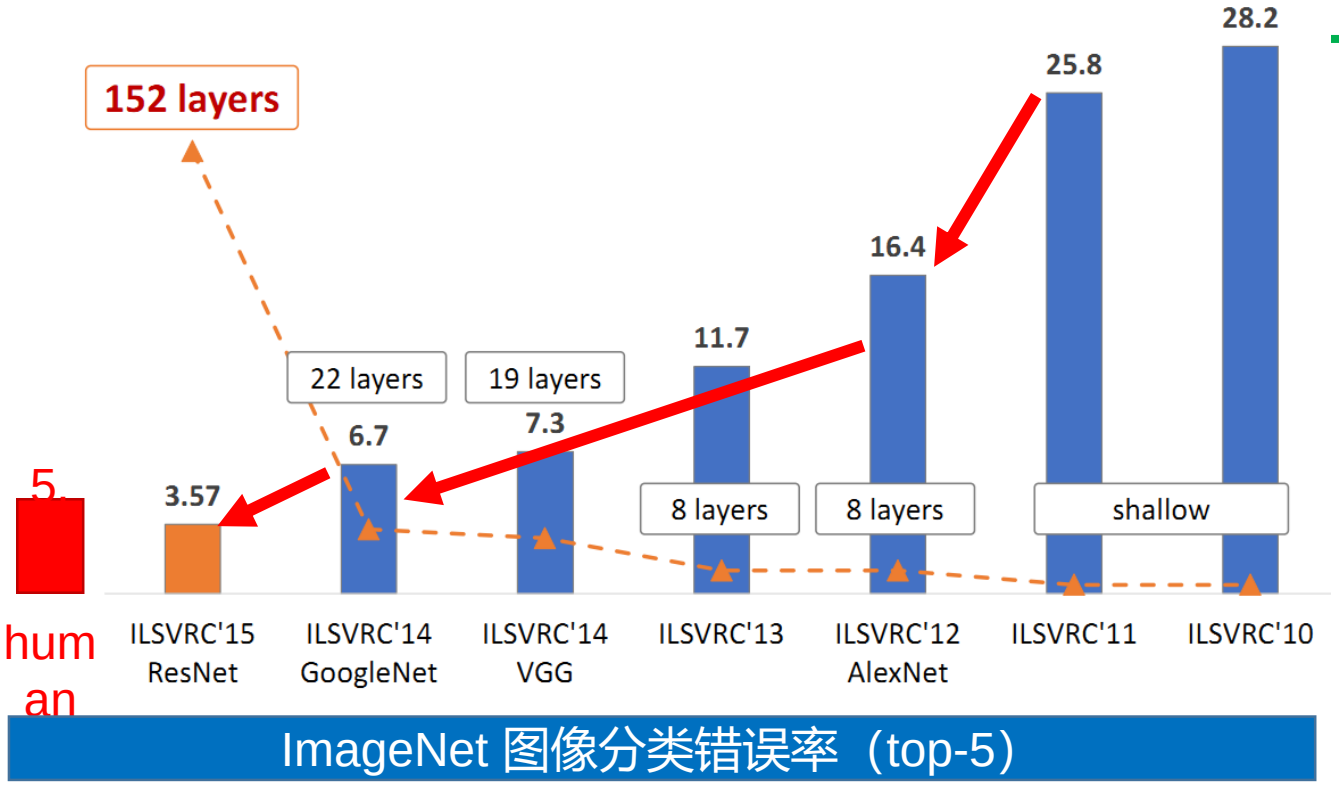
COCO has several features:

- **Object segmentation**
- **Recognition in context**
- **Superpixel stuff segmentation**
- **330K images (>200K labeled)**
- **1.5 million object instances**
- **80 object categories**
- **91 stuff categories**
- **5 captions per image**
- **250,000 people with keypoints**



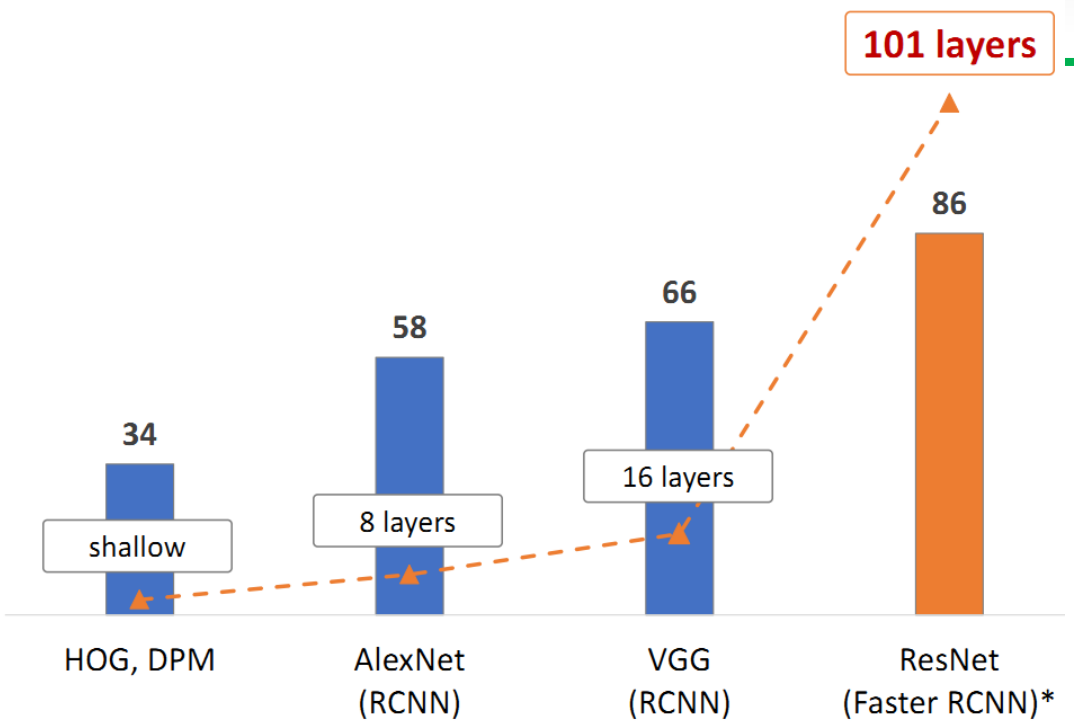
Microsoft COCO: Common Objects in Context, Lin et al, ECCV 2014

图像识别技术的发展



Kaiming He et al, Residual Learning for Image Recognition, CVPR 2016

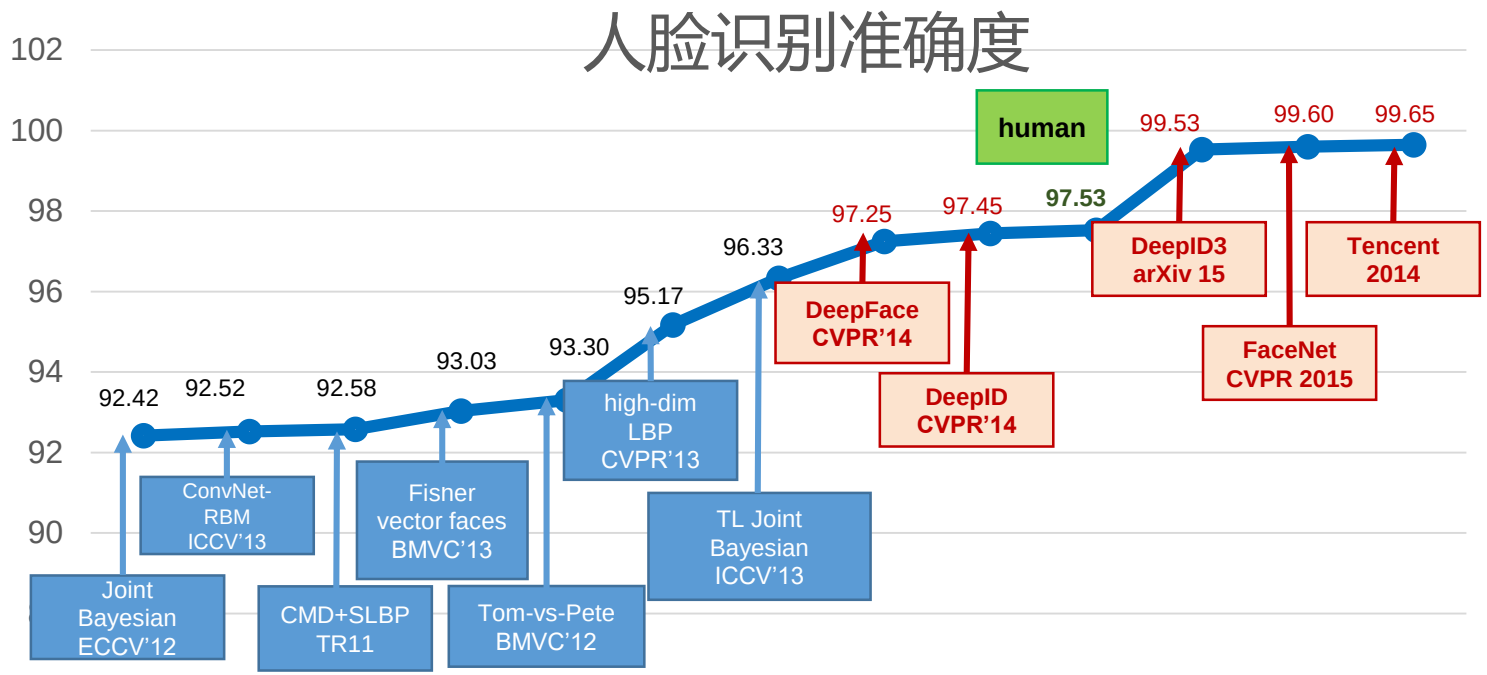
物体检测技术的发展



PASCAL VOC 2007 物体检测准确度

Kaiming He et al, Residual Learning for Image Recognition, CVPR 2016

人脸识别的进展


人脸识别准确度
102
100
98
96
94
92
90
92.42
92.52
92.58
93.03
93.30
95.17
96.33
97.25
97.45
97.53
99.53
99.60
99.65
human
Joint Bayesian ECCV'12
ConvNet-RBM ICCV'13
CMD+SLBP TR11
Fisner vector faces BMVC'13
Tom-vs-Pete BMVC'12
high-dim LBP CVPR'13
TL Joint Bayesian ICCV'13
DeepFace CVPR'14
DeepID CVPR'14
DeepID3 arXiv 15
FaceNet CVPR 2015
Tencent 2014

人工智能产业发展

商汤 sensetime

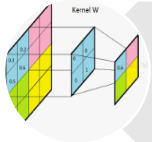
Face++ 旷视

依图 | YITU

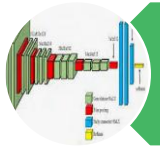
云从科技 CLOUDWALK

Outline

Introduction



CNN Architectures



Classical CNN Models

Discussion

CNN经典模型---AlexNet

ImageNet Classification with Deep Convolutional Neural Networks

NIPS2012

Alex Krizhevsky
University of Toronto
kriz@cs.utoronto.ca

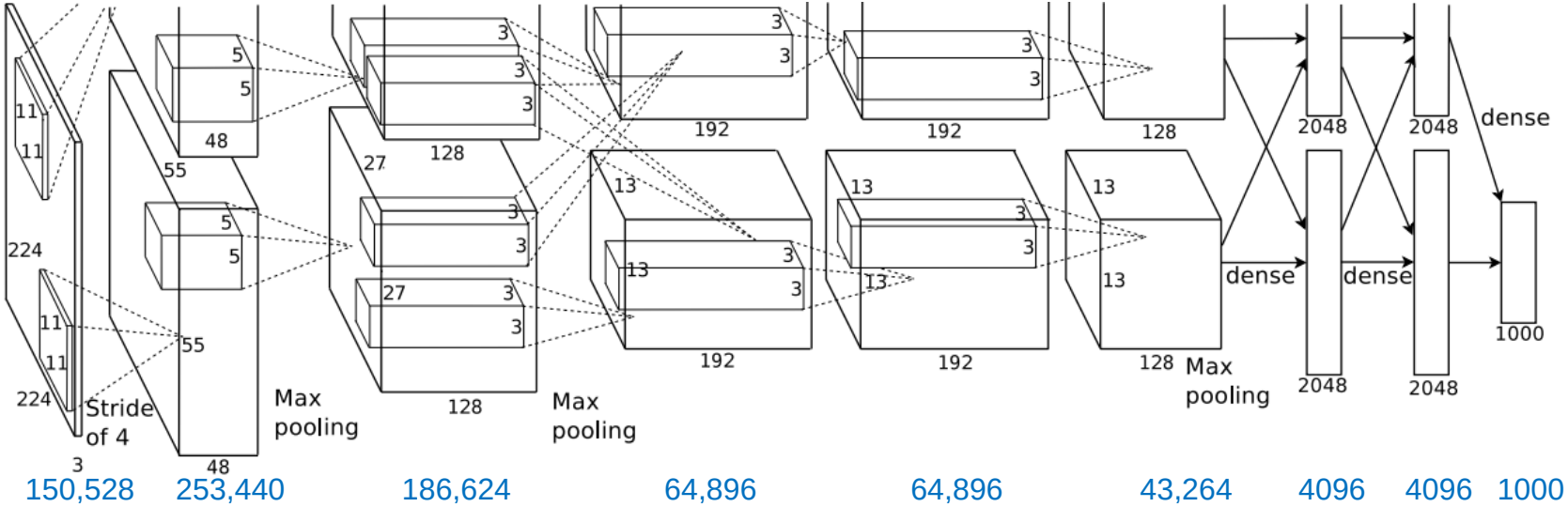
Ilya Sutskever
University of Toronto
ilya@cs.utoronto.ca

Geoffrey E. Hinton
University of Toronto
hinton@cs.utoronto.ca

- 2012年ImageNet LSVRC（大规模视觉识别挑战赛）**冠军**
  - Top-5错误率 15.3% (无数据扩充情况下16.4%), << 第二名 26.2%
  - Top-1错误率37.5%
  - NIPS2012，截止2020年7月29日被引用次数：66572
  - NIPS2012，截止2021年11月8日被引用次数: **89733**
- **主要贡献**
  - 非线性激活函数ReLu
  - 防止过拟合的方法：DropOut+数据扩充（Data Augmentation）
  - 多GPU实现，LRN归一化层

CNN经典模型---AlexNet

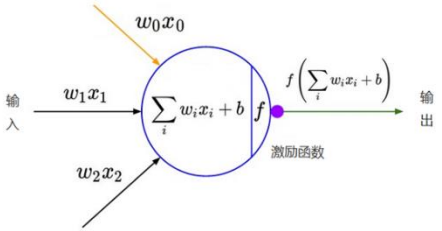
- 网络结构：**2个GPU并行**



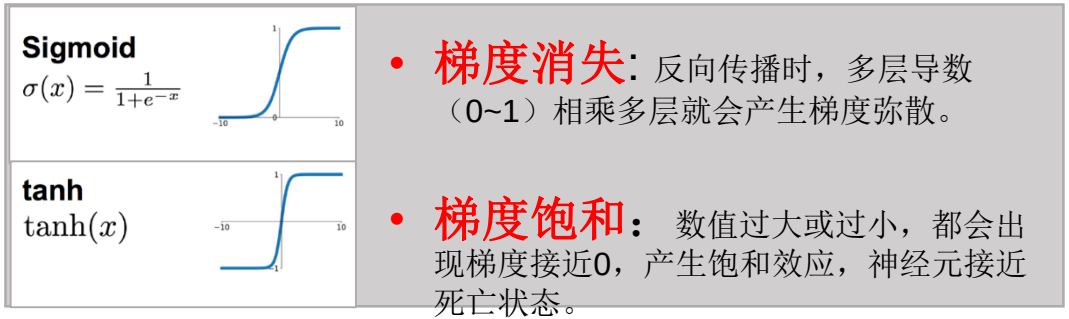
- GTX580，3G内存
- 两个GPU分别负责上下各一半的神经元，在特定层进行通信，加快了训练速度

CNN经典模型---AlexNet

- **激活函数**

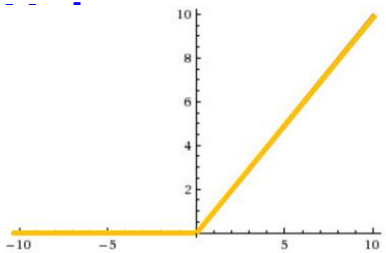


非线性激励函数增强了神经网络的表达能力

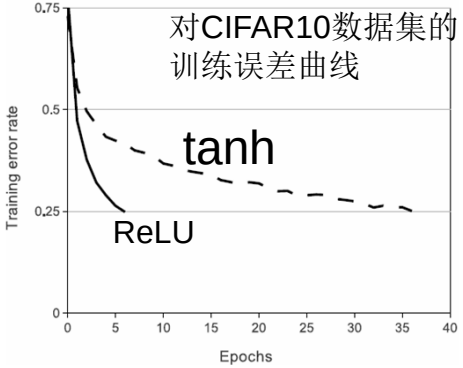


- **梯度消失**：反向传播时，多层导数（0~1）相乘多层就会产生梯度弥散。
- **梯度饱和**：数值过大或过小，都会出现梯度接近0，产生饱和效应，神经元接近死亡状态。

- **ReLU：Rectified Linear**
  - **非线性函数**
  - **导数在x>0时始终为1**
    - **计算量大大减少**
    - **收敛速度快**
- **14M的训练集**

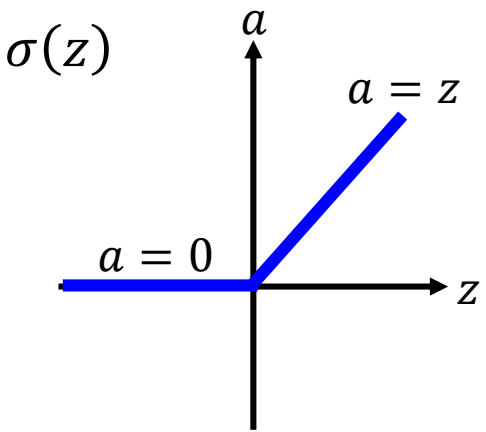


$F(x) = \max(0, x)$



ReLU

- Rectified Linear Unit (ReLU)

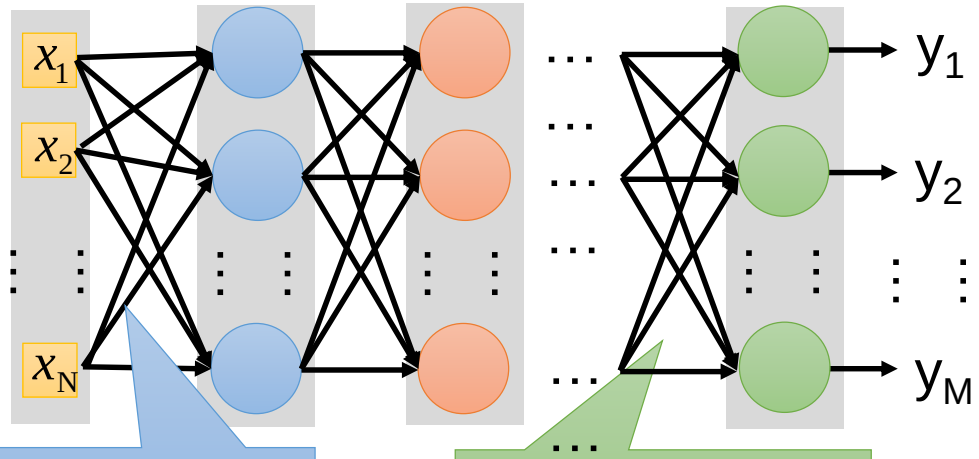


[Xavier Glorot, AISTATS'11]
[Andrew L. Maas, ICML'13]
[Kaiming He, arXiv'15]

原因:

1. 计算较快
2. 生物上：动作电位
3. 相当于无穷多不同bias的sigmoid函数叠加
4. 解决梯度消失问题

梯度消失



梯度较小

更新缓慢

几乎随机

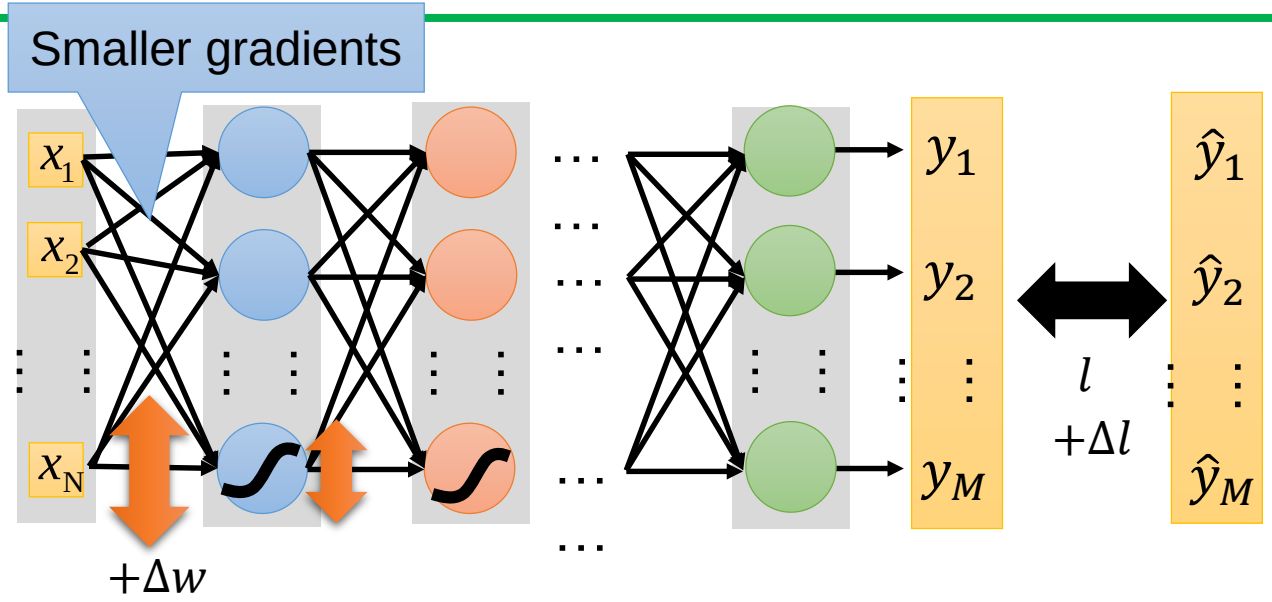
梯度较大

更新较快

接近收敛

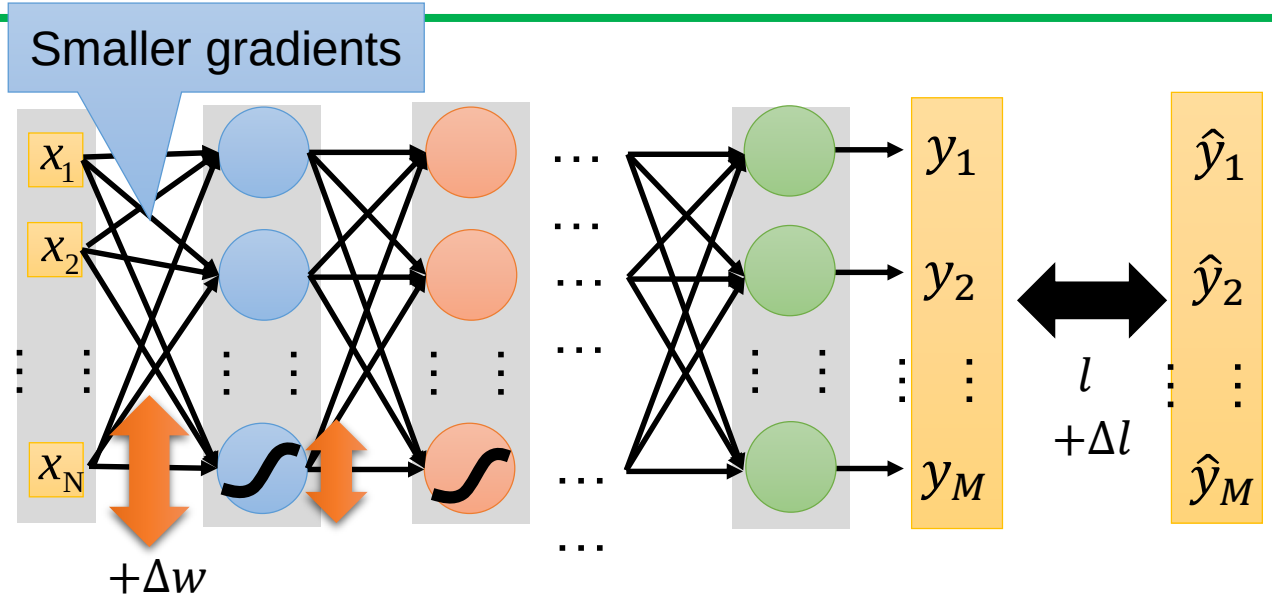
based on random!?

梯度消失



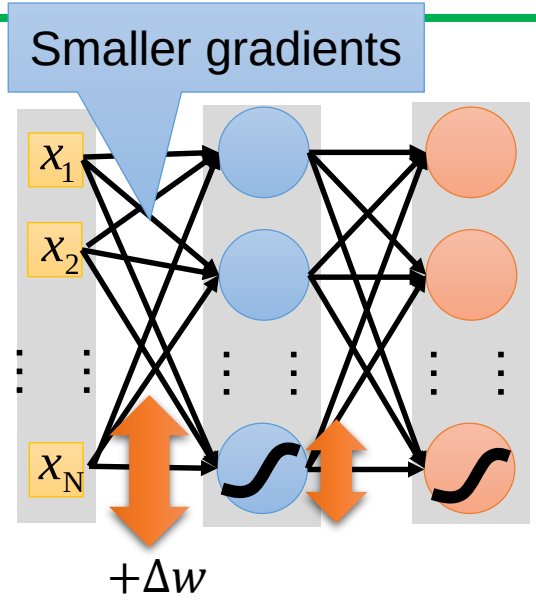
直观解释偏微分： $\frac{\partial l}{\partial w} = ? \frac{\Delta l}{\Delta w}$

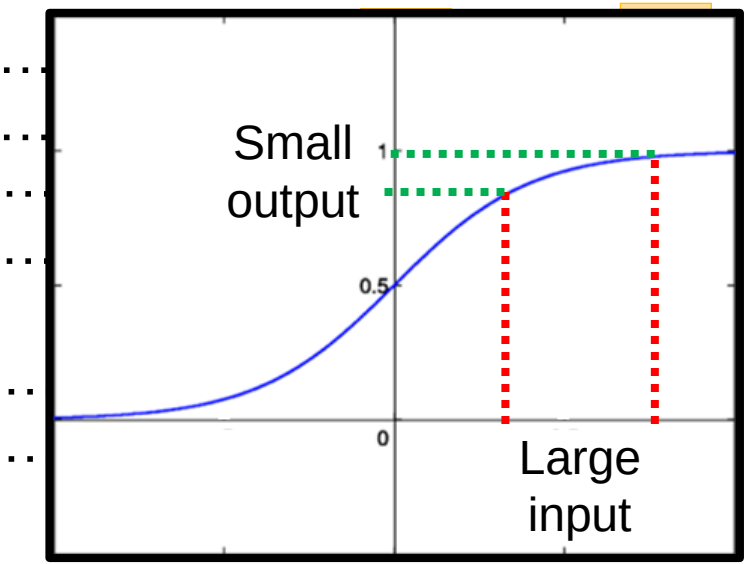
梯度消失



直观解释偏微分： $\frac{\partial l}{\partial w} = ? \frac{\Delta l}{\Delta w}$

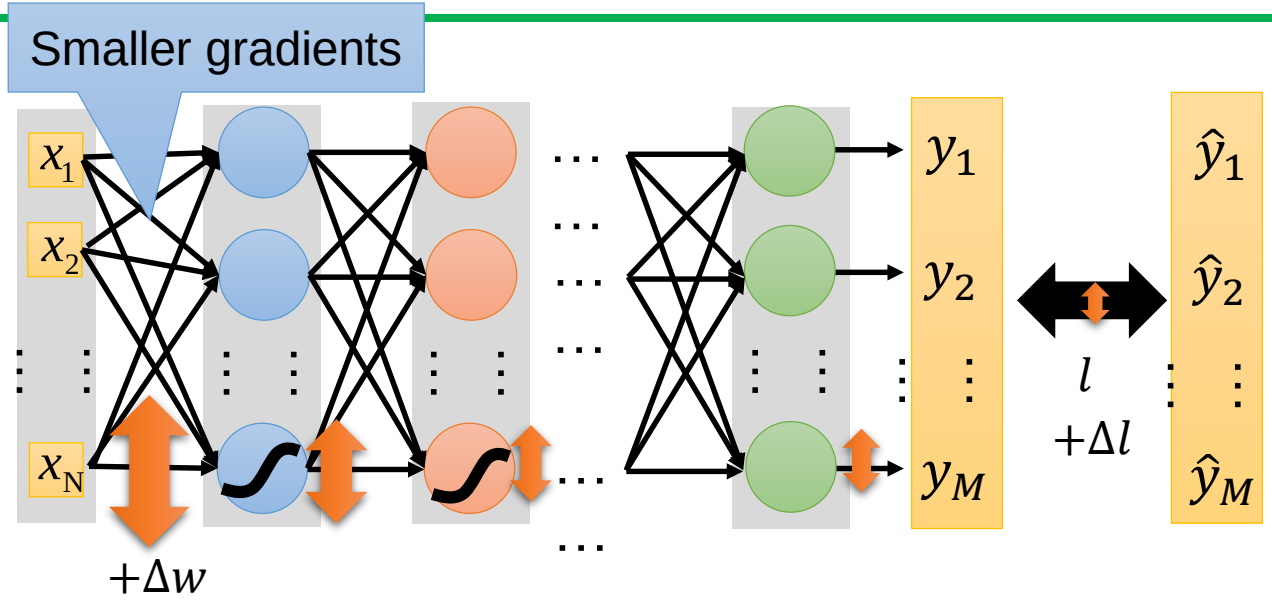
梯度消失





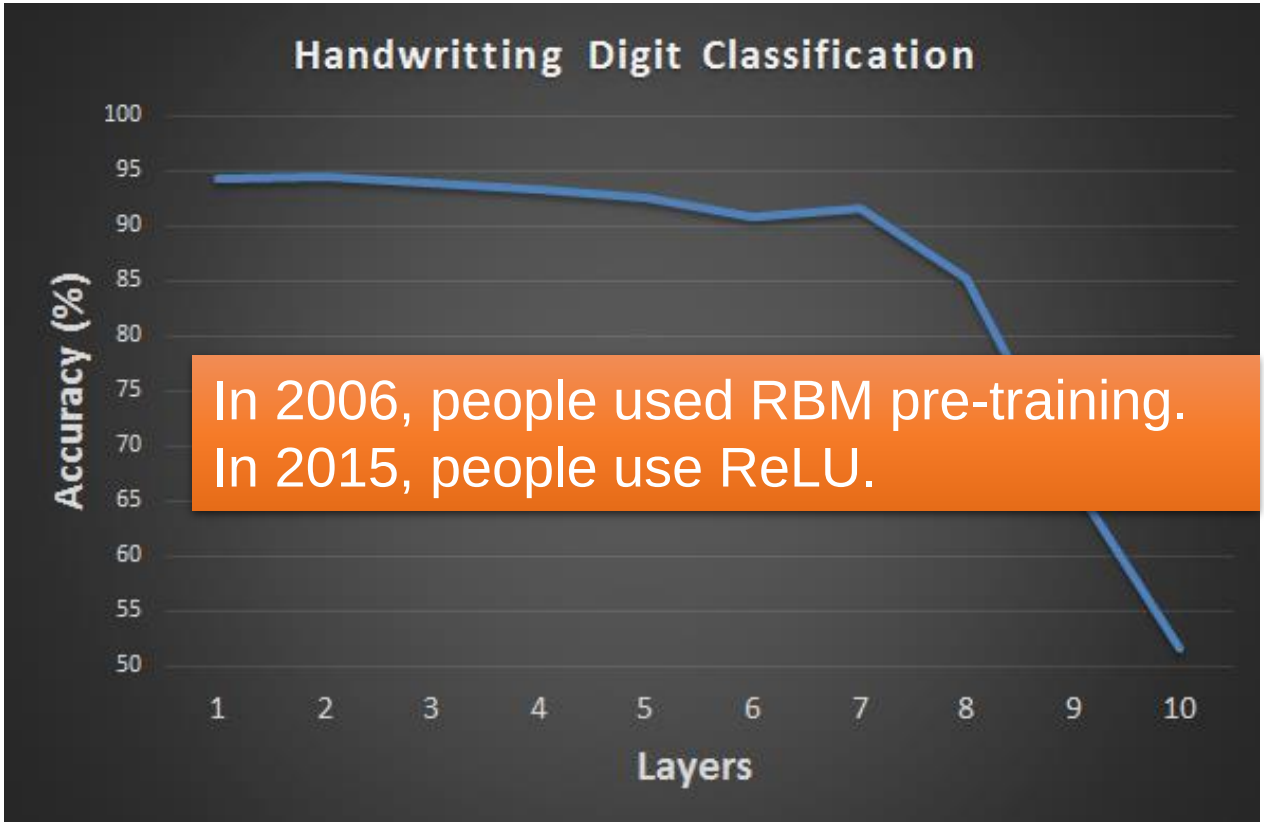
直观解释偏微分： $\frac{\partial l}{\partial w} = ? \frac{\Delta l}{\Delta w}$

梯度消失



直观解释偏微分： $\frac{\partial l}{\partial w} = ? \frac{\Delta l}{\Delta w}$

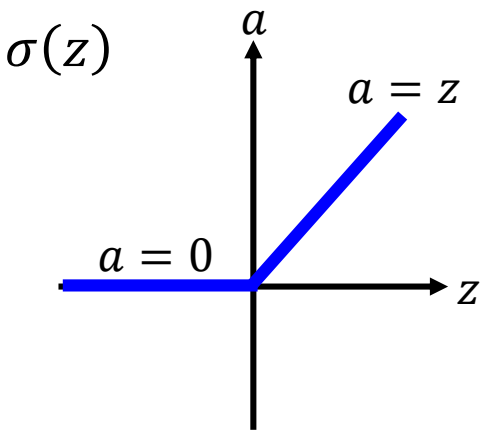
如何充分挖掘深度网络的能力?



In 2006, people used RBM pre-training.
In 2015, people use ReLU.

ReLU

- Rectified Linear Unit (ReLU)

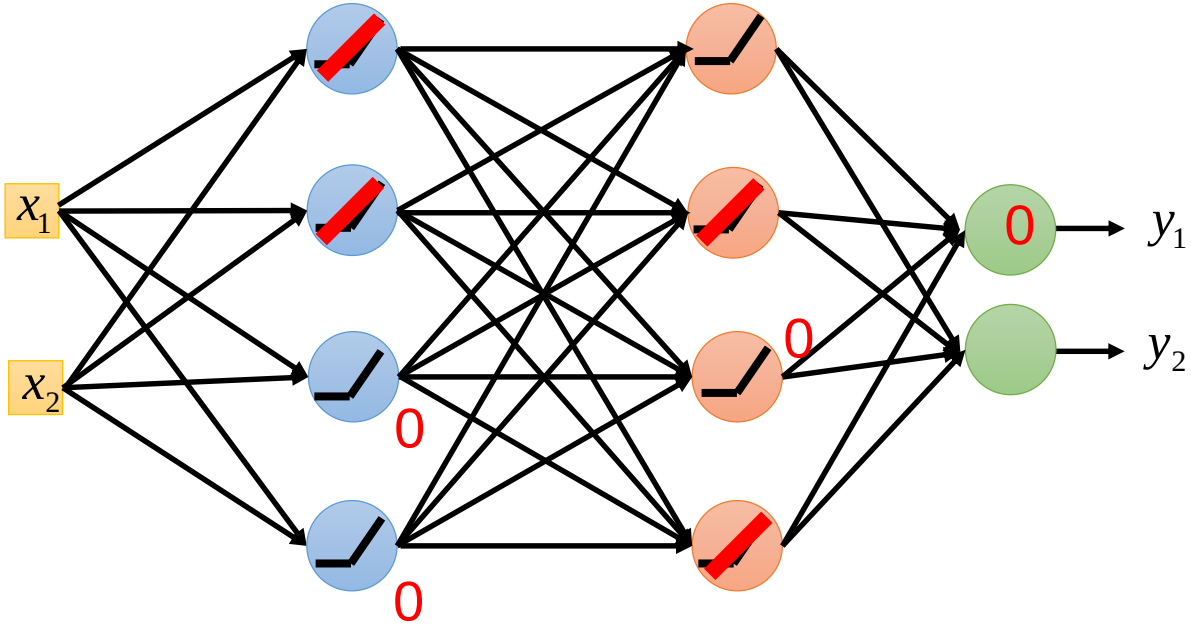


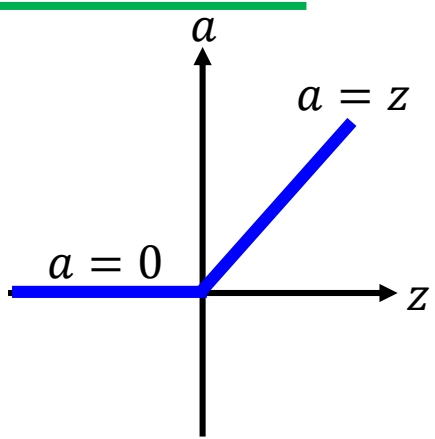
[Xavier Glorot, AISTATS'11]
[Andrew L. Maas, ICML'13]
[Kaiming He, arXiv'15]

原因:

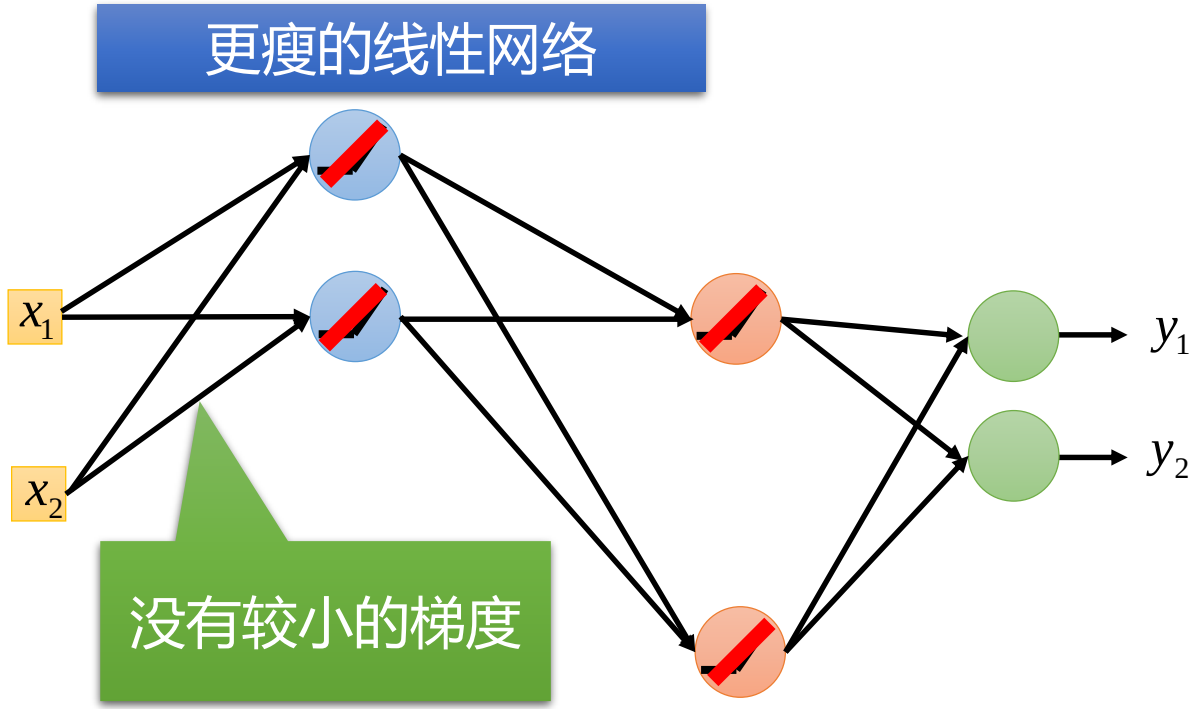
1. 计算较快
2. 生物上：动作电位
3. 相当于无穷多不同bias的sigmoid函数叠加
4. 解决梯度消失问题

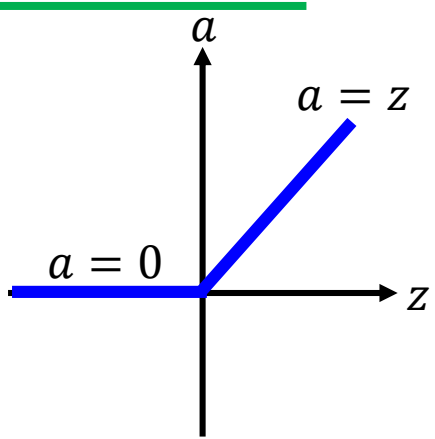
ReLU



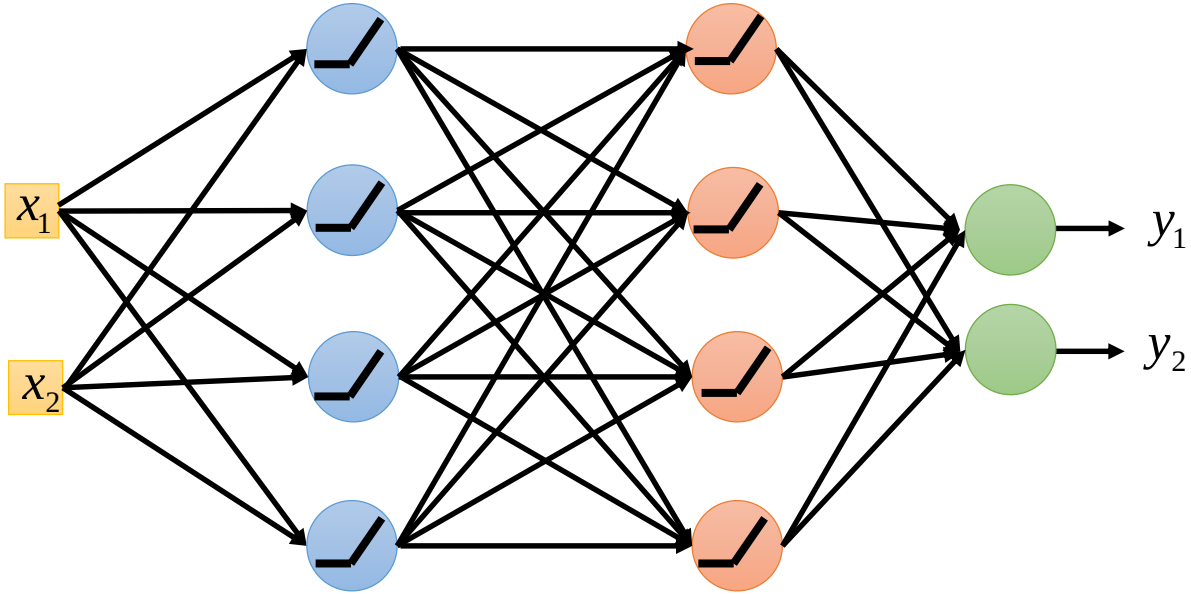


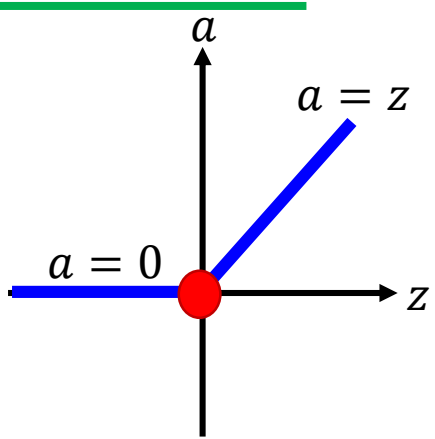
ReLU





ReLU



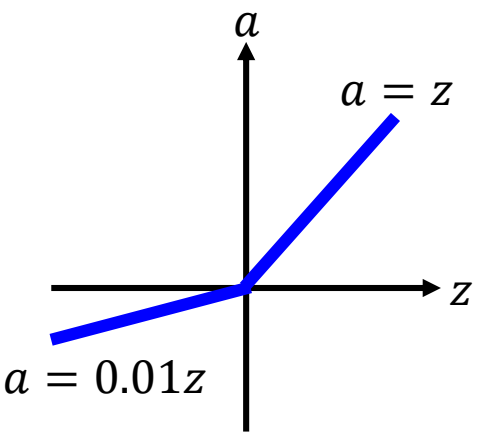


反向传播: 梯度

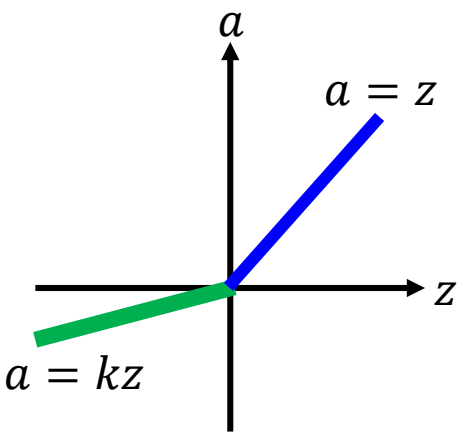
$$\frac{\partial a}{\partial z}\Big|_{z=0} = ?$$

ReLU – 变形

Leaky ReLU



Parametric ReLU



k: 网络参数
通过梯度下降学习

Maxout

- 可学习的激活函数 [Ian J. Goodfellow, ICML’13]

https://www.iangoodfellow.com/

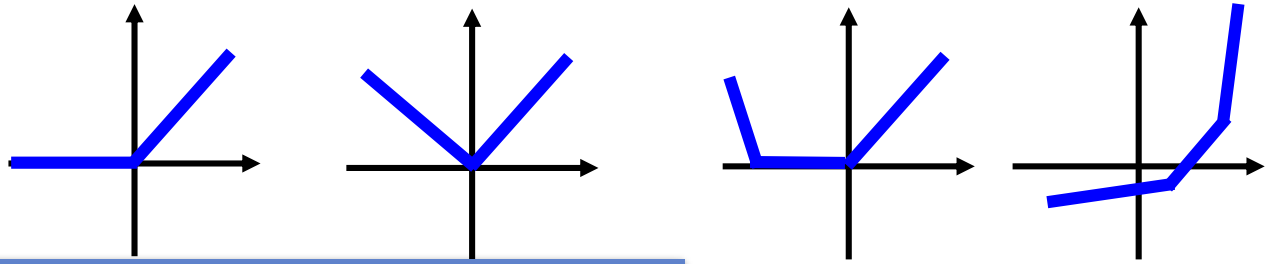
Yoshua Bengio

Maxout

- 可学习的激活函数 [Ian J. Goodfellow, ICML’13]
  - Maxout网络中的激活函数可以是任意分段线性凸函数
  - 分多少段取决于一组里用多少元素

一个组里有2个元素

一个组里有3个元素

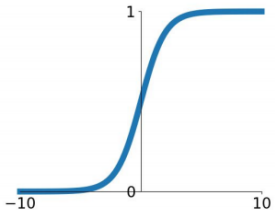


ReLU 是Maxout的一种特殊情况

激活函数

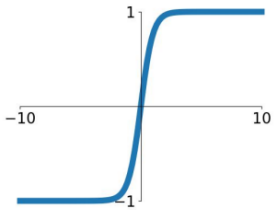
Sigmoid

$$\sigma(x) = \frac{1}{1+e^{-x}}$$



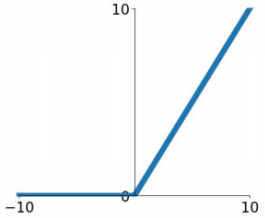
tanh

$$\tanh(x)$$



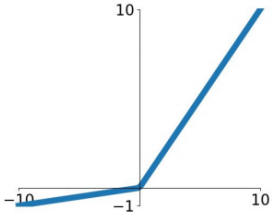
ReLU

$$\max(0, x)$$



Leaky ReLU

$$\max(0.1x, x)$$

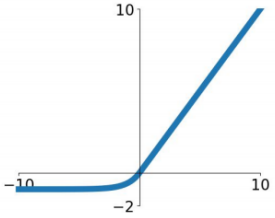


Maxout

$$\max(w_1^T x + b_1, w_2^T x + b_2)$$

ELU

$$\begin{cases} x & x \geq 0 \\ \alpha(e^x - 1) & x < 0 \end{cases}$$



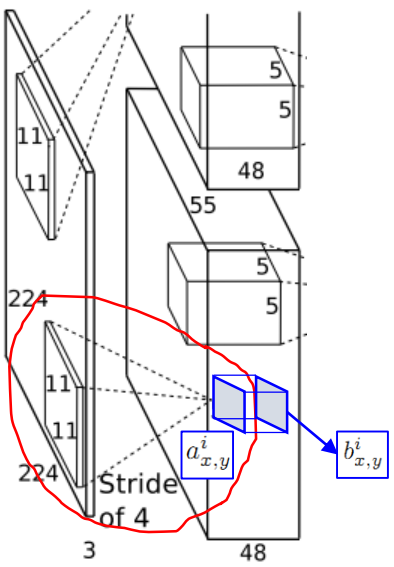
Ref: http://www.cnblogs.com/rgvb178/p/6055213.html

CNN经典模型---AlexNet

- **局部响应归一化**（Local Response Normalization）

$$b^i_{x,y} = a^i_{x,y} / \left(k + \alpha \sum_{j=\max(0,i-n/2)}^{\min(N-1,i+n/2)} (a^j_{x,y})^2 \right)^{\beta}$$

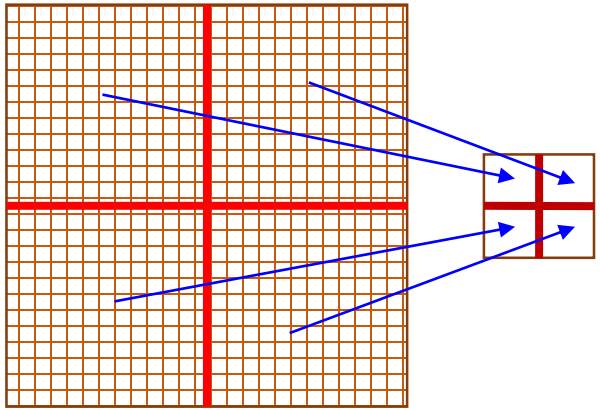
  – ReLU的响应结果是无界的，所以需要归一化
  – 借鉴了神经生物学中“侧抑制”的概念，即被激活的神经元抑制相邻神经元。即在同一个空间位置，通过相邻特征图（神经元）上同一个位置的响应来归一化
  – 特征图的顺序是任意的（训练之前就确定了）

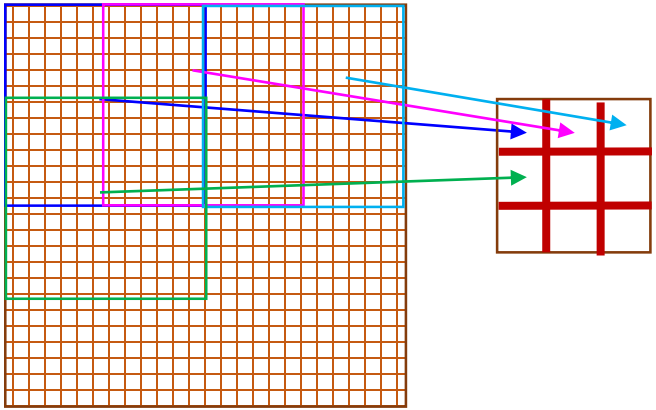


- 提高了**网络的泛化能力**
  – 1.2%（top-5）和1.4%（top-1）的误差减小

CNN经典模型---AlexNet

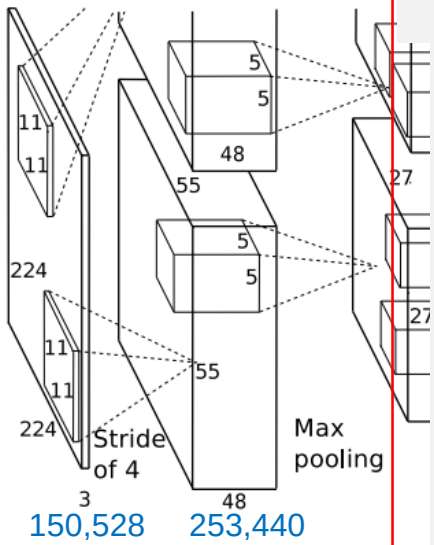
- 重叠池化 (Overlapping Pooling)





- 避免过拟合，提高泛化能力
- 0.3%（top-5）和0.4%（top-1）的误差减小

CNN经典模型---AlexNet

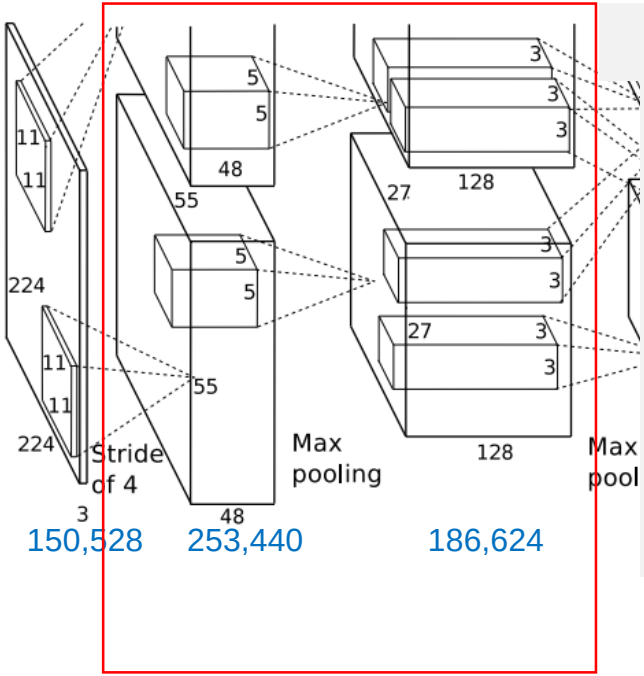


第一层卷积层

- 卷积：224x224x3 → 55x55x96
  - 核大小：11x11，步长 4
  - 填充（Padding ）：防止丢失过多边界信息
  - 224x224 →227x227
  - (227-11)/4+1=55
- ReLU：55x55x96 → 55x55x96
- 重叠池化：55x55x96 → 27x27x96
  - 核大小：3x3，步长 2
  - (55-3)/2+1=27
- 局部响应归一化LRN：27x27x96→ 27x27x96

- 96个特征图分为两组，各自在一个独立的GPU上进行运算

CNN经典模型---AlexNet

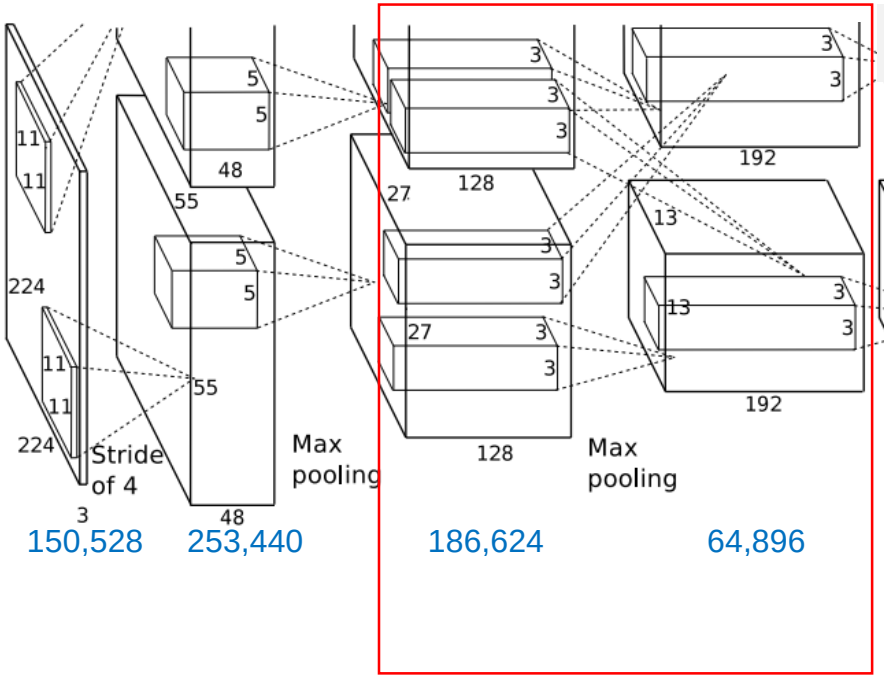
- 

第二层卷积层

- 卷积：27x27x48 → 27x27x128
  - 核大小：5x5x48，步长 1
  - 填充（Padding ）：2+2
  - (27+4-5)+1=27
- ReLU：27x27x128 → 27x27x128
- 重叠池化：27x27x128 → 13x13x128
  - 核大小：3x3，步长 2
  - (27-3)/2+1=13
- 局部响应归一化LRN：13x13x128→ 13x13x128

- 特征图分为两组，各自在一个独立的GPU上进行运算

CNN经典模型---AlexNet

- 

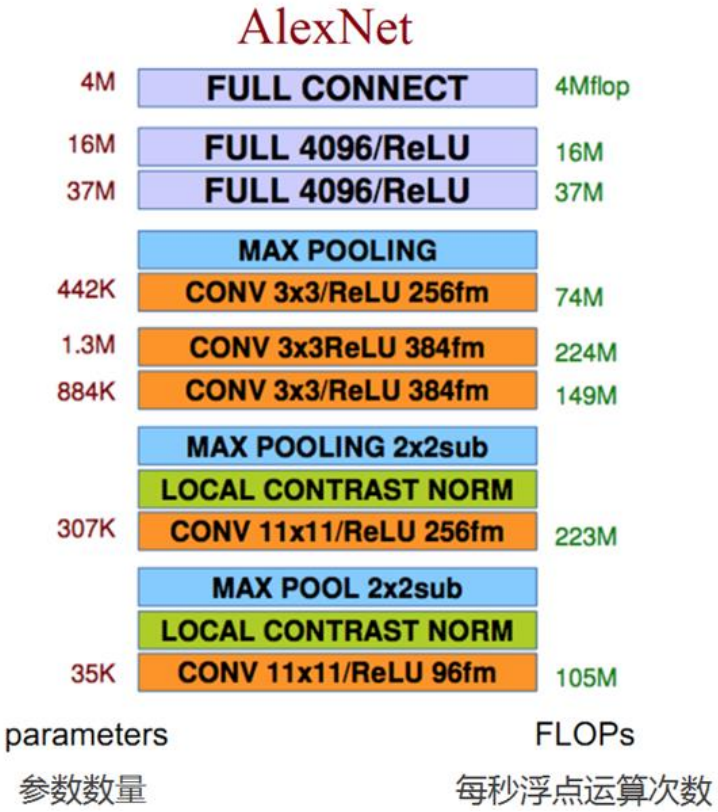
第三层卷积层

- 卷积：13x13x128 → 13x13x192
  - 核大小：3x3x256，步长 1
  - 填充（Padding ）：1+1
  - (27+4-5)+1=27
- ReLU：27x27x128 → 27x27x128

- 两组各192个特征图，每个GPU都要处理来自前一层的所有输入

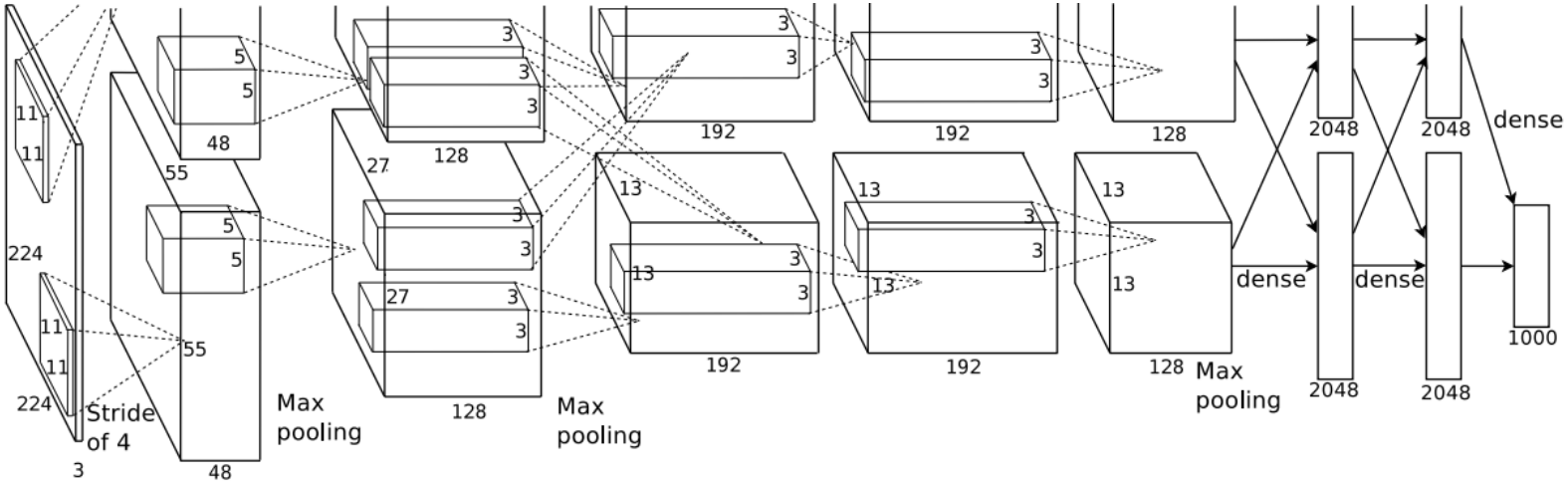
CNN经典模型---AlexNet

- 网络参数及运算次数
  - 约6千万的可训练参数



CNN经典模型---AlexNet

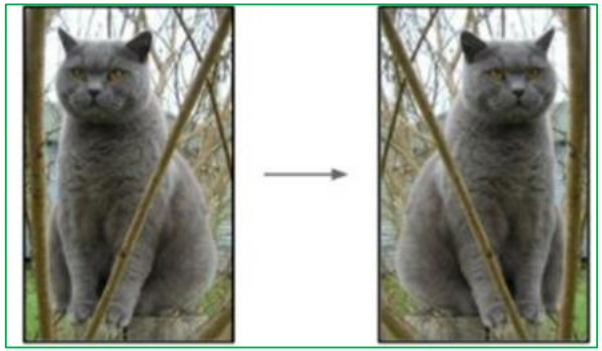
- 网络结构：**6千万参数、1000个类别**

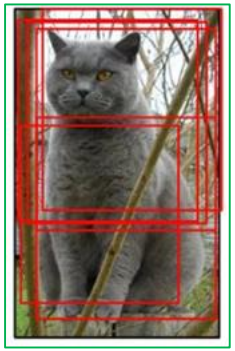


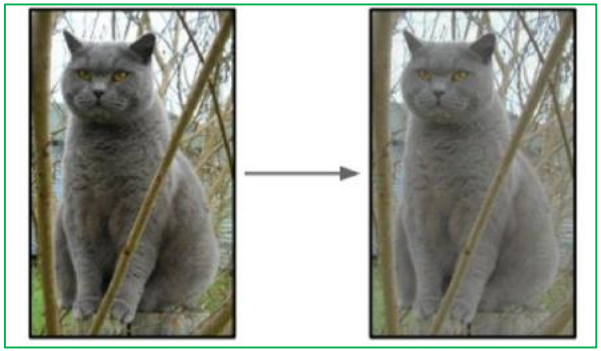
- **过拟合问题（overfitting)**
  - 减少过拟合:在训练时 数据扩充+DropOut机制

CNN经典模型---AlexNet

- **数据扩充**
  - 图像几何变换：平移、镜像
    - 从256x256图像中裁剪224x224大小的图像块
    - (256-224)*(256-224)*2=2048倍的数据扩充
  - 图像灰度变换：
    - 对像素RGB做PCA变换，对主成分做（0,0.1）的高斯扰动

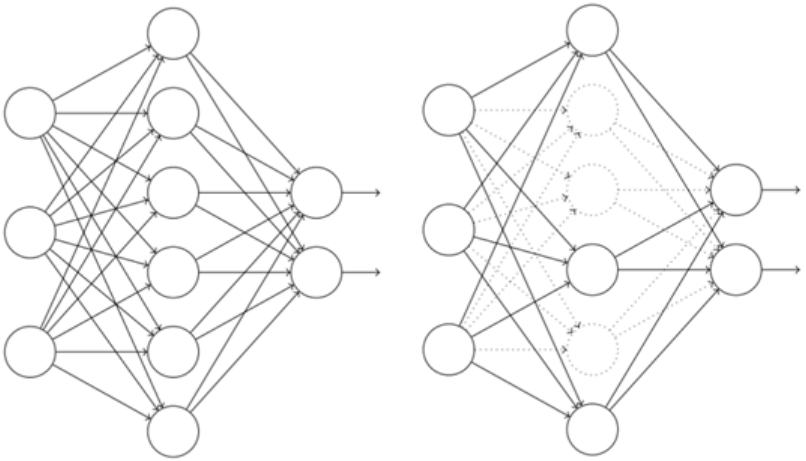






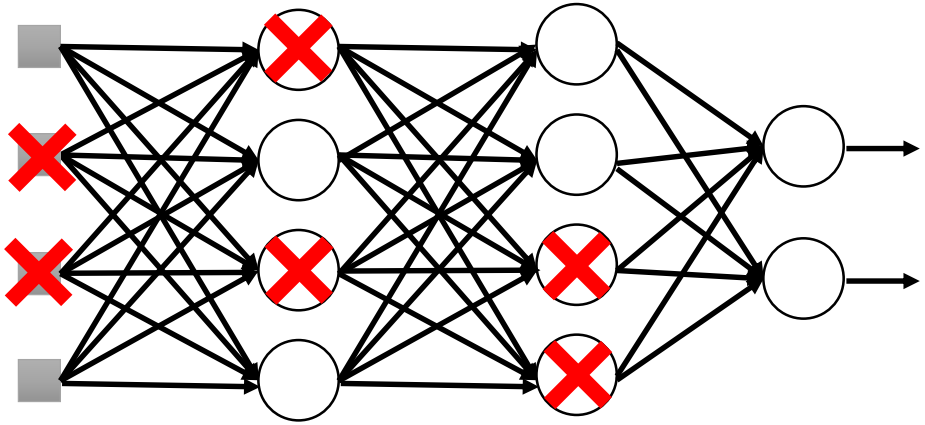
CNN经典模型---AlexNet

- **DropOut**
  - 用于全连接层
  - 每次迭代以某概率（0.5）将神经元输出置零，不参
  - 产生不同的网络结构，进行组合，大大减少了过拟
  - 训练时间增加一倍



Dropout

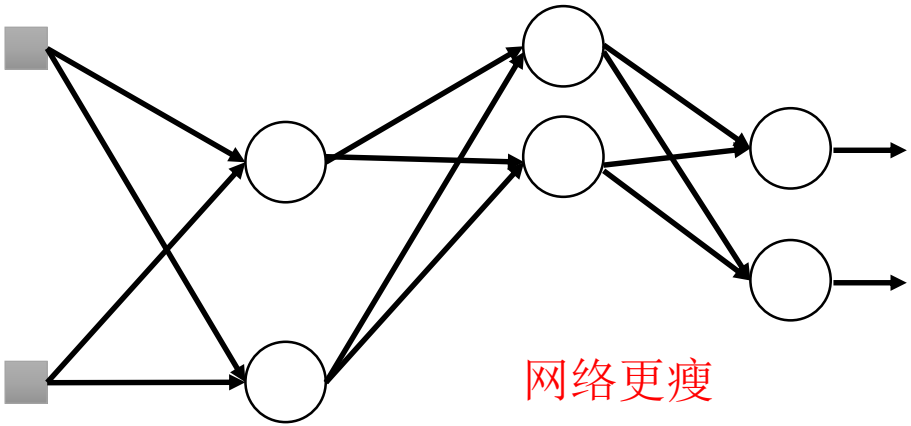
训练:



- **每次更新参数之前**
  - 每个神经元有 p%的概率被剪掉

Dropout

训练:

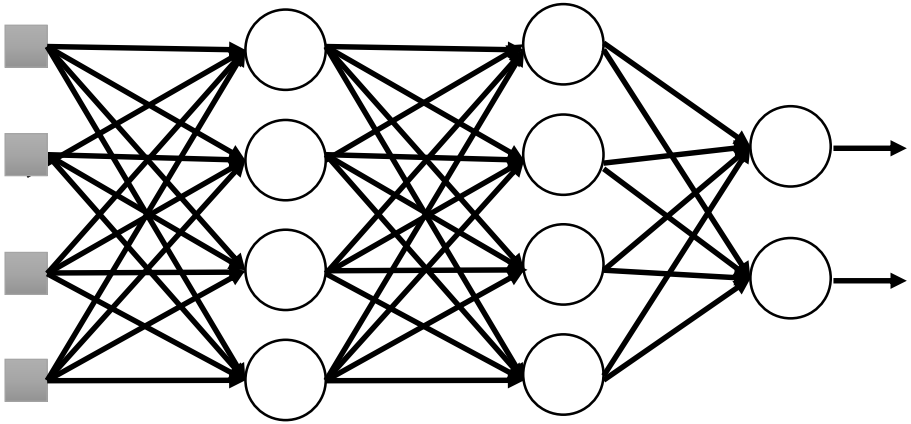


- **每次更新参数之前**
  - 每个神经元有 p%的概率被剪掉 → **网络结构发生变化**
  - 采用变化后网络进行训练

对每个mini-batch, 重新采样神经元

Dropout

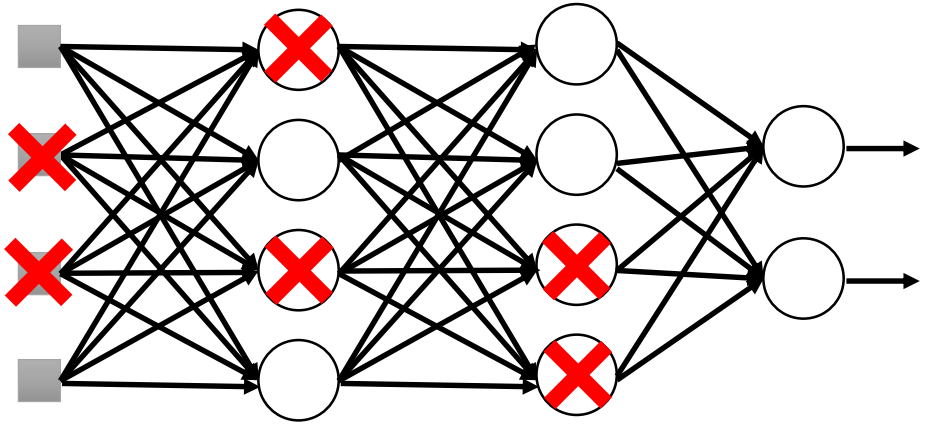
测试:



- **无dropout**
  - 如果训练时dropout 概率是p%, 所有权重w=w*(1-p%)
  - 例：如dropout概率是50%，则一个w=1的权重在测试阶段变成 $w = 0.5$

Dropout

训练:



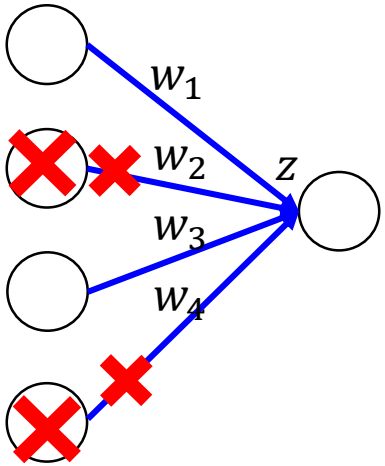
- 无Dropout，全都放在一起，总有神经元滥竽充数
- 有dropout，每组神经元单独拿出来都可以很强

Dropout – 直观解释

- 为何测试时 权重 w*(1-p)% (dropout rate p%)

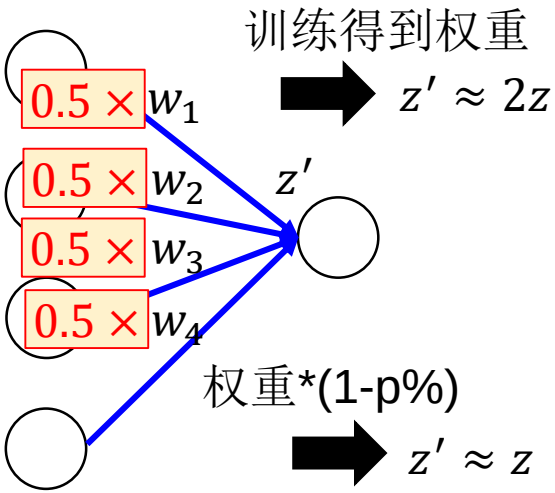
Training of Dropout

假设dropout rate=50%

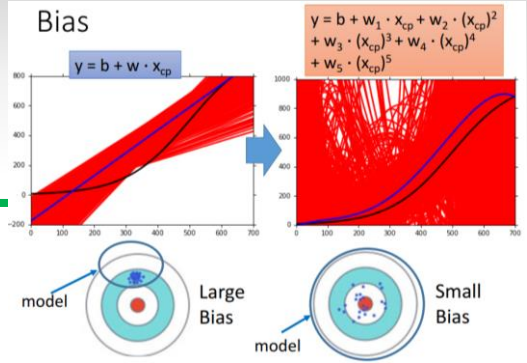


Testing of Dropout

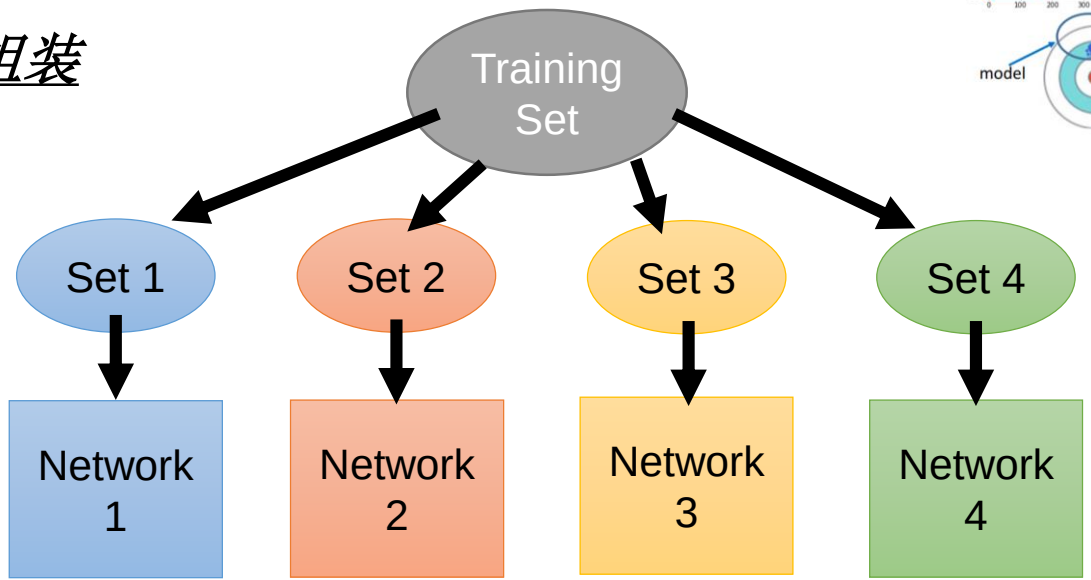
无 dropout



Dropout ~ 组装

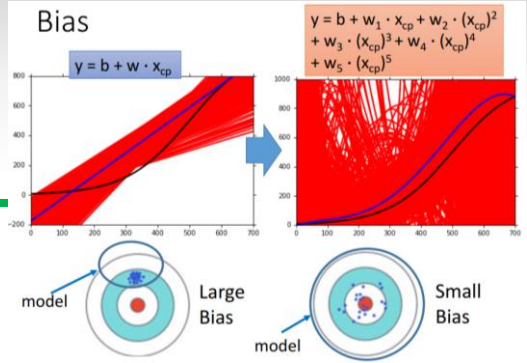


网络组装

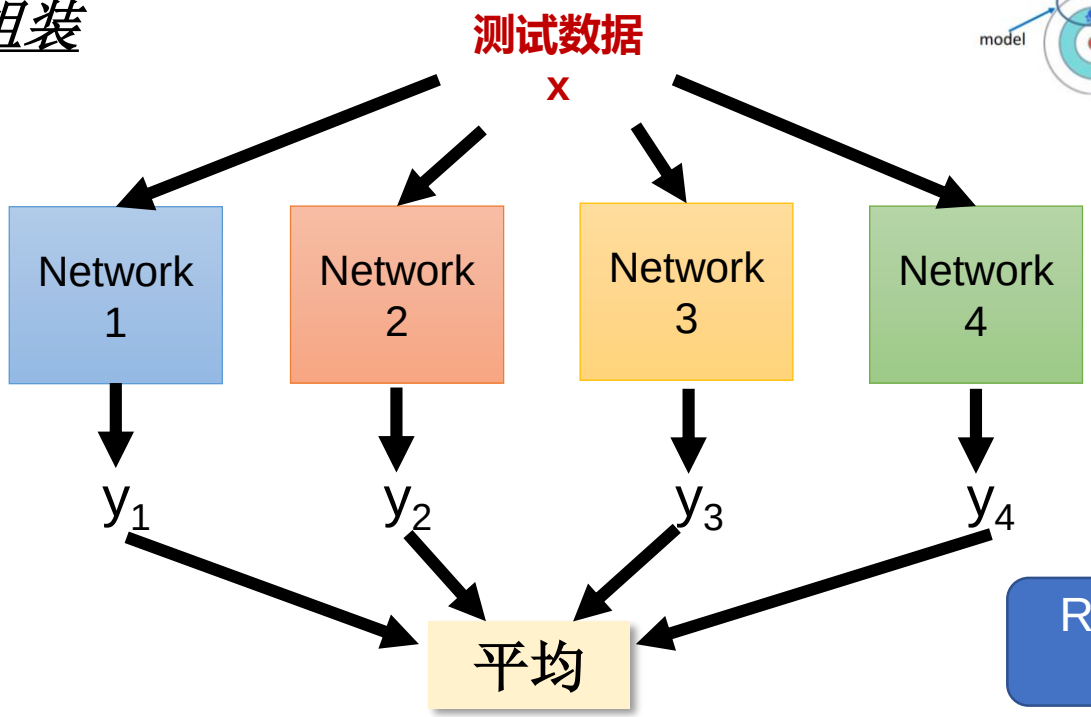


训练了一组不同结构的神经网络

Dropout ~ 组装

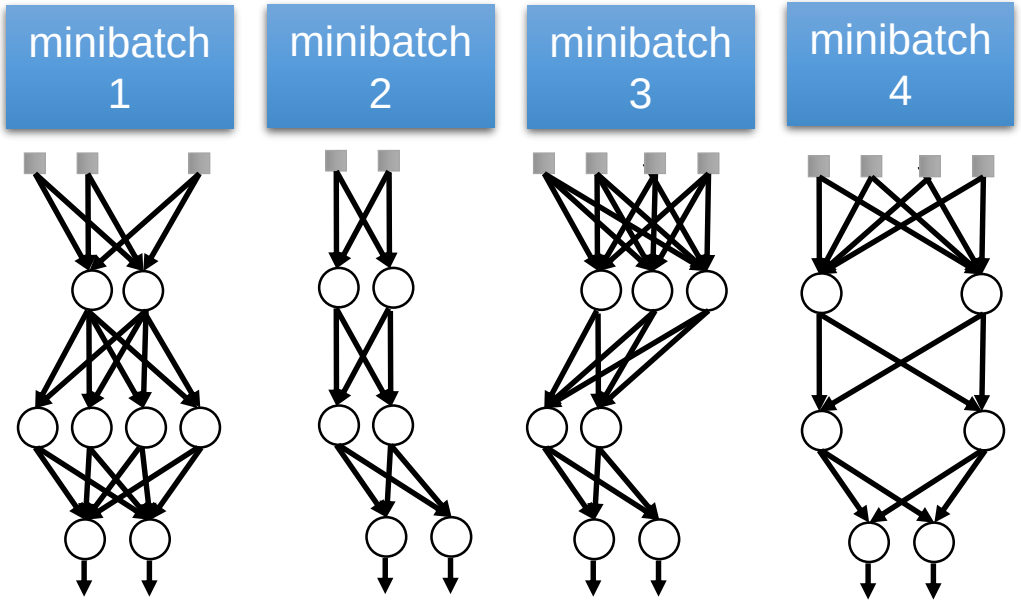


网络组装



Random Forest
随机森林

Dropout ~ 组装



Training of Dropout

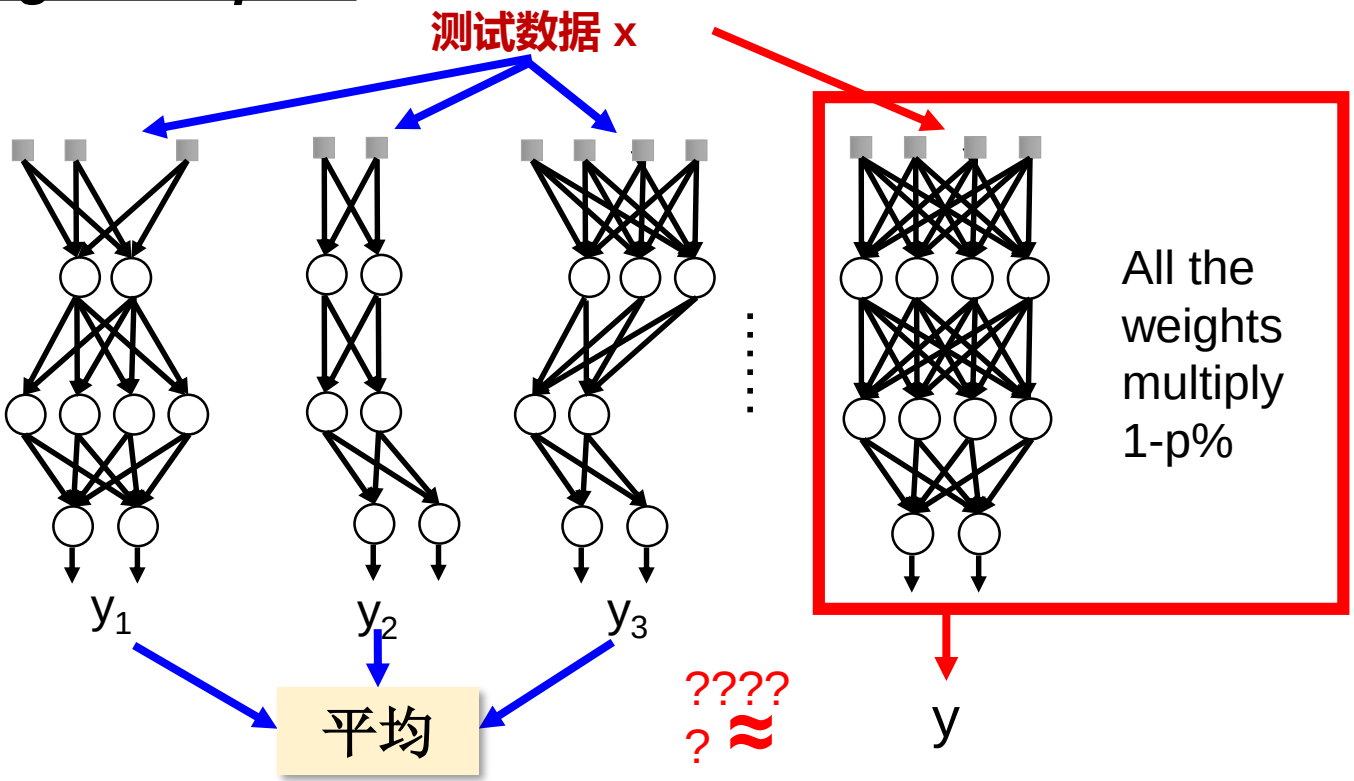
M个神经元

2^M 可能的网络

- 一个mini-batch训练一个网络
- 部分网络共享一些参数

Dropout is a kind of ensemble.

Testing of Dropout



Dropout参考资料

- More reference for dropout [Nitish Srivastava, JMLR'14] [Pierre Baldi, NIPS'13][Geoffrey E. Hinton, arXiv'12]
- Dropout works better with Maxout [Ian J. Goodfellow, ICML'13]
- Dropconnect [Li Wan, *ICML'13*]
  - Dropout delete neurons
  - Dropconnect deletes the connection between neurons
- Annealed dropout [S.J. Rennie, SLT'14]
  - Dropout rate decreases by epochs
- Standout [J. Ba, NISP'13]
  - Each neural has different dropout rate

CNN经典模型---AlexNet

- **训练过程**
  - 随机梯度下降
    - Batch size = 128
    - Momentum = 0.9
    - Weight Decay of 0.0005

$$
\begin{aligned}
v_{i+1} &:= 0.9 \cdot v_i - 0.0005 \cdot \epsilon \cdot w_i - \epsilon \cdot \left\langle \frac{\partial L}{\partial w}\Big|_{w_i} \right\rangle_{D_i} \\
w_{i+1} &:= w_i + v_{i+1}
\end{aligned}
$$

  - 初始化
    - 随机初始化，高斯分布（0,0.01）
    - 神经元偏置设为1：ReLU可以获得正输入
  - 5~6天：两块NVIDIA GTX 580 3GB GPUs

CNN经典模型---AlexNet

- 网络学到了什么?

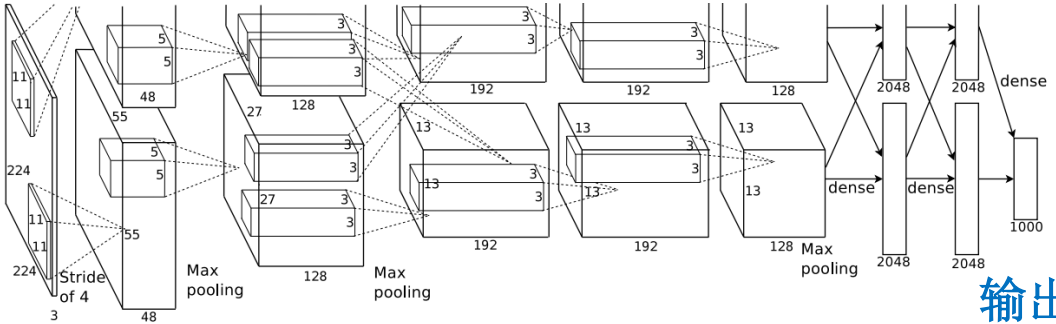
Model	Top-1	Top-5
Sparse coding [2]	*47.1%*	*28.2%*
SIFT + FVs [24]	*45.7%*	*25.7%*
CNN	**37.5%**	**17.0%**

第一层卷积的96个卷积核：
一系列与**频率**、**方向**相关的卷积核，和**斑点**检测核

CNN经典模型---AlexNet

- **测试阶段**

任意图像
尺寸



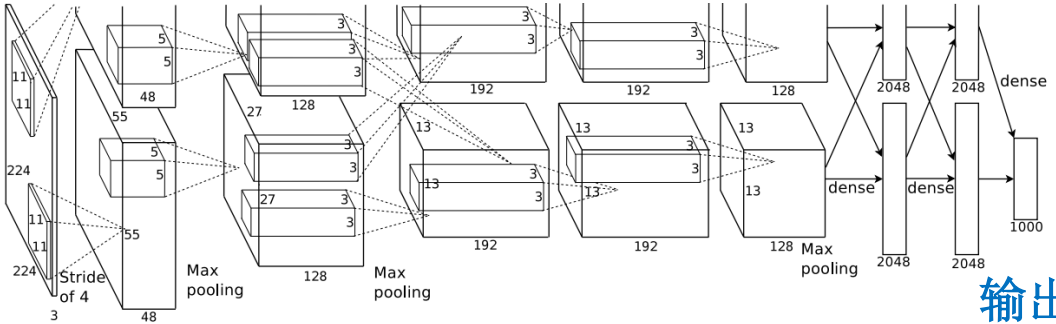
输入
224x224

输出
1000

CNN经典模型---AlexNet

- **测试阶段**

任意图像
尺寸



输入
224x224

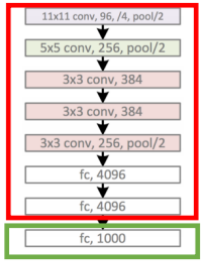
输出
1000

- 对左上、右上、左下、右下、中间分别做5次裁剪，然后翻转，对10个预测结果求平均
- 问题?
  - **卷积重复计算**

经典卷积神经网络

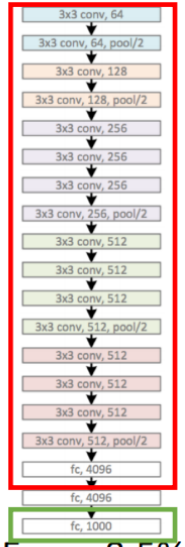
Deep ConvNet for Visual Recognition

2012: AlexNet
5 conv. layers



Error: 15.3%

2014: VGG
16 conv. layers



Error: 8.5%

2015: GoogLeNet
22 conv. layers

Error: 7.8%

2016: ResNet
>100 conv. layers

Error: 4.4%

Slide from Bolei Zhou, M

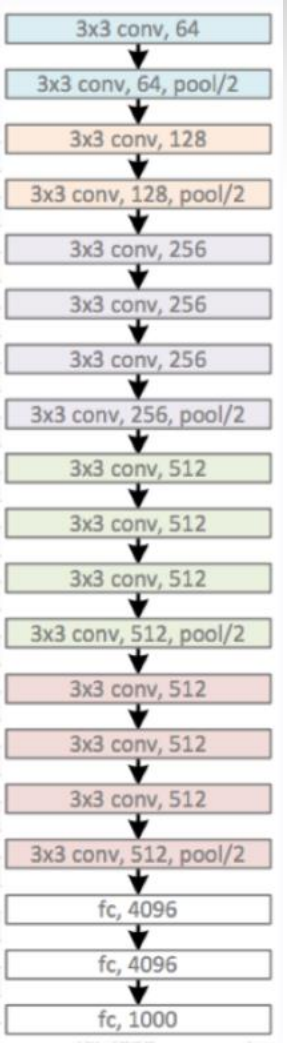
CNN经典模型---VGGNet 2014

VERY DEEP CONVOLUTIONAL NETWORKS FOR LARGE-SCALE IMAGE RECOGNITION

Karen Simonyan* **& Andrew Zisserman+**
Visual Geometry Group, Department of Engineering Science, University of Oxford
{karen,az}@robots.ox.ac.uk

ICRL 2015, 2020年被引用次数：41805
截止2021年11月8日被引用次数: **66231**

- **网络深度在卷积神经网络设计中的作用**
  - 使用固定尺寸的小卷积核 3x3
  - 16-19层的网络结构
  - 在多个数据集上带来较大的性能提升



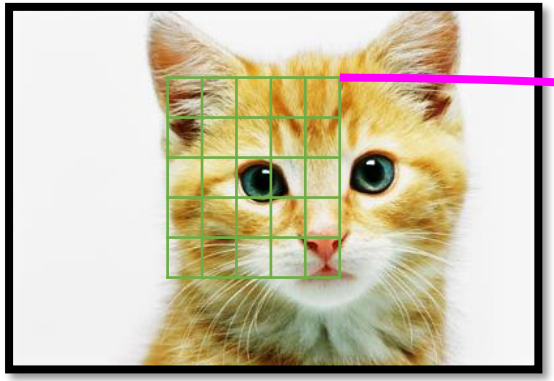
CNN经典模型---VGGNet

- 6个不同配置的网络结构，11-19层
- 堆叠相同的卷积核
  - 3x3，Stride=1
  - 感知野随深度增加
- 池化层：5层maxpool,2x2,stride=2
- 特征图的通道数随深度增加（64->512）
  - 底层特征（边界、方向、斑点）
  - 高层特征

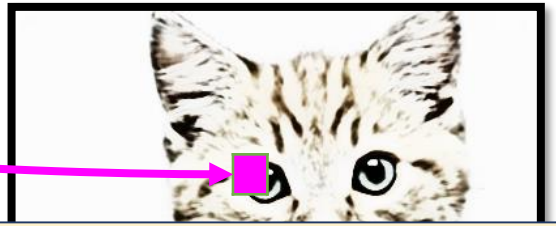
思考: 3层3x3的卷积层与一层7x7的卷积层的区别？

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

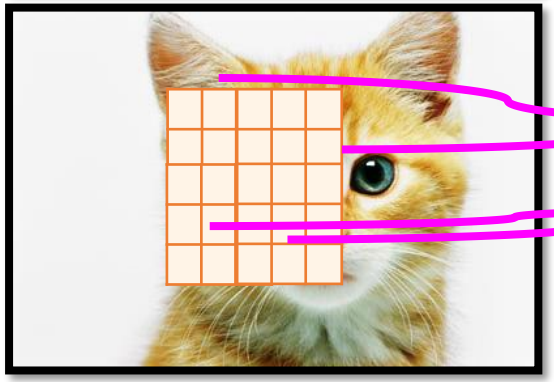
卷积核堆叠

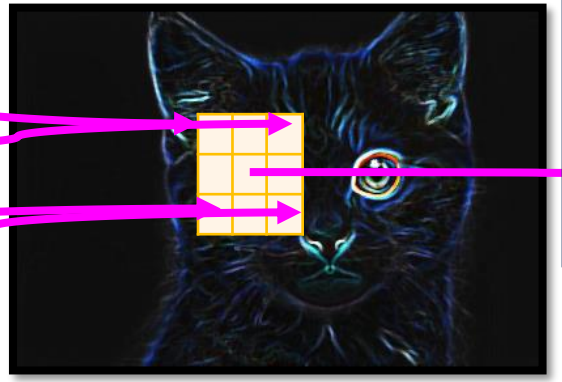


1层5x5卷积



2层3x3卷积





- 2层3x3与1层5x5卷积 感知野相同
- 参数量：9x2 vs 25
- 两层叠加有三次非线性变换（ReLU），组合成更复杂的决策函数，使其判别能力大大增强

CNN经典模型---VGGNet

- 6个不同配置的网络结构，11-19层
- 堆叠相同的卷积核
  - 3x3，Stride=1
  - 感知野随深度增加
- 池化层：5层maxpool,2x2,stride=2
- 特征图的通道数随深度增加（64->512）
  - 底层特征（边界、方向、斑点）
  - 高层特征

思考: 3层3x3的卷积层与一层7x7的卷积层

- ◆3层卷积层叠加有三次非线性变换（ReLU），组合成
- ◆更复杂的决策函数，使其判别能力大大增强
- ◆参数减少：27 vs 49

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

CNN经典模型---VGGNet

- 6个不同配置的网络结构，11-19层
- 堆叠相同的卷积核
  - 3x3，Stride=1
  - 感知野随深度增加
- 池化层：5层maxpool,2x2,stride=2
- 特征图的通道数随深度增加（64->512）
  - 底层特征（边界、方向、斑点）
  - 高层特征

◆局部响应归一化LRN的作用
  ◆在VGGNet中并未有性能提升
  ◆内存和计算资源消耗更多

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

CNN经典模型---VGGNet

- 6个不同配置的网络结构，11-19层
- 堆叠相同的卷积核
  - 3x3，Stride=1
  - 感知野随深度增加
- 池化层：5层maxpool,2x2,stride=2
- 特征图的通道数随深度增加（64->512）
  - 底层特征（边界、方向、斑点）
  - 高层特征

◆1x1卷积核
- ◆ 不影响感知野
- ◆增加决策函数的非线性（ReLU）

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

CNN经典模型---VGGNet

- 6个不同配置的网络结构，11-19层
- 堆叠相同的卷积核
  - 3x3，Stride=1
  - 感知野随深度增加
- 池化层：5层maxpool,2x2,stride=2
- 特征图的通道数随深度增加（64->512）
  - 底层特征（边界、方向、斑点）
  - 高层特征

不同网络结构的参数（百万）

Network	A,A-LRN	B	C	D	E
Number of parameters	133	133	134	138	144

ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 LRN	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	conv3-128	conv3-128 **conv3-128**	conv3-128 conv3-128	conv3-128 conv3-128	conv3-128 conv3-128
maxpool					
conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256	conv3-256 conv3-256 **conv1-256**	conv3-256 conv3-256 **conv3-256**	conv3-256 conv3-256 conv3-256 **conv3-256**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512	conv3-512 conv3-512 **conv1-512**	conv3-512 conv3-512 **conv3-512**	conv3-512 conv3-512 conv3-512 **conv3-512**
maxpool					
FC-4096					
FC-4096					
FC-1000					
soft-max					

CNN经典模型---VGGNet

- 训练
  - 图像集：
    - 每个尺度下随机裁剪224x224的图像，进行随机裁剪、反转、颜色调整等数据扩充方法
    - **图像缩放**：不同物体尺寸、以及裁剪后部分物体
    - 两个固定尺寸：256和384。
      - 384大小的图像集可以使用256图像集训练的基础上进行调优（fine-tuning）
    - **多尺度图像采样**：将原图像随机缩放到[S_min, S_max]内的一个数值，再进行裁剪。
      - 在384图像集训练模型的基础上调优
  - 训练参数
    - **优化方法**(optimizer)：含有动量的随机梯度下降SGD+momentum(0.9)。
    - **批尺寸**(batch size)：256.
    - **正则化**(regularization):采用L2正则化，weight decay是5e-4。
    - **Dropout**在前两个全连接层后，p=0.5。
  - 在装有4个NVIDIA Titan Black GPUs的电脑上，训练一个网络需要2-3周。

CNN经典模型---VGGNet

- **测试阶段**
  - **多尺度求平均?**

CNN经典模型---VGGNet

- **实验分析:单一测试尺度**
  - local response normalization(A-LRN)并不能改善A网络性能。
  - 分类误差随着深度增加而降低。
  - 在训练时采用图像尺度抖动(scale jittering)可以改善图像分类效果。

Table 3: **ConvNet performance at a single test scale.**

ConvNet config. (Table 1)	smallest image side		top-1 val. error (%)	top-5 val. error (%)
	train (S)	test (Q)		
A	256	256	29.6	10.4
A-LRN	256	256	29.7	10.5
B	256	256	28.7	9.9
C	256	256	28.1	9.4
	384	384	28.1	9.3
	[256;512]	384	27.3	8.8
D	256	256	27.0	8.8
	384	384	26.8	8.7
	[256;512]	384	25.6	8.1
E	256	256	27.3	9.0
	384	384	26.9	8.7
	[256;512]	384	**25.5**	**8.0**

CNN经典模型---VGGNet

- **实验分析:多测试尺度**
  - 相对于单一尺度评估，多尺度评估提高了分类精度。
  - 在训练时采用图像尺度抖动(scale jittering)可以改善图像分类效果。

Table 4: **ConvNet performance at multiple test scales.**

ConvNet config. (Table 1)	smallest image side		top-1 val. error (%)	top-5 val. error (%)
	train (S)	test (Q)		
B	256	224,256,288	28.2	9.6
C	256	224,256,288	27.7	9.2
	384	352,384,416	27.8	9.2
	[256; 512]	256,384,512	26.3	8.2
D	256	224,256,288	26.6	8.6
	384	352,384,416	26.5	8.6
	[256; 512]	256,384,512	**24.8**	**7.5**
E	256	224,256,288	26.9	8.7
	384	352,384,416	26.7	8.6
	[256; 512]	256,384,512	**24.8**	**7.5**

CNN经典模型---VGGNet

- **实验分析:多裁剪评估**
  - 多裁剪(multi-crop)评估比起密集(dense)评估，效果更好。而且两者具有互补作用，结合两种方式，效果更好。

Table 5: **ConvNet evaluation techniques comparison.** In all experiments the training scale S was sampled from $[256; 512]$, and three test scales Q were considered: $\{256, 384, 512\}$.

ConvNet config. (Table 1)	Evaluation method	top-1 val. error (%)	top-5 val. error (%)
D	dense	24.8	7.5
	multi-crop	24.6	7.5
	multi-crop & dense	**24.4**	**7.2**
E	dense	24.8	7.5
	multi-crop	24.6	7.4
	multi-crop & dense	**24.4**	**7.1**

CNN经典模型---VGGNet

- **实验分析:多网络融合**
  - 如果结合多个卷积网络的sofamax输出，分类效果会更好。

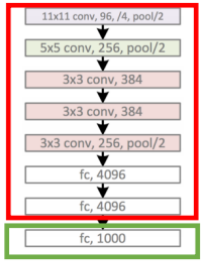
Table 6: **Multiple ConvNet fusion results.**

Combined ConvNet models	Error		
	top-1 val	top-5 val	top-5 test
ILSVRC submission			
(D/256/224,256,288), (D/384/352,384,416), (D/[256;512]/256,384,512) (C/256/224,256,288), (C/384/352,384,416) (E/256/224,256,288), (E/384/352,384,416)	24.7	7.5	7.3
post-submission			
(D/[256;512]/256,384,512), (E/[256;512]/256,384,512), dense eval.	24.0	7.1	7.0
(D/[256;512]/256,384,512), (E/[256;512]/256,384,512), multi-crop	23.9	7.2	-
(D/[256;512]/256,384,512), (E/[256;512]/256,384,512), multi-crop & dense eval.	**23.7**	**6.8**	**6.8**

经典卷积神经网络

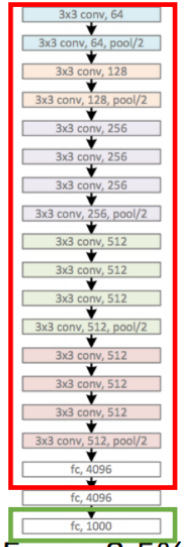
Deep ConvNet for Visual Recognition

2012: AlexNet
5 conv. layers



Error: 15.3%

2014: VGG
16 conv. layers



Error: 8.5%

2015: GoogLeNet
22 conv. layers

Error: 7.8%

2016: ResNet
>100 conv. layers

Error: 4.4%

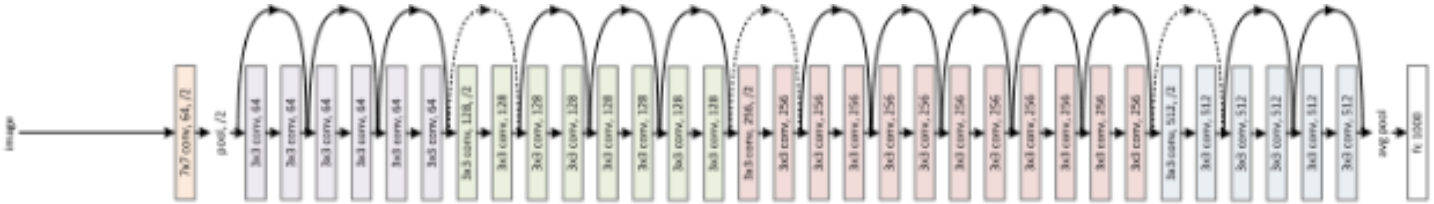
Slide from Bolei Zhou, I

CNN经典模型
深度残差网络Deep ResNet

Deep Residual Learning for Image Recognition

Kaiming He Xiangyu Zhang Shaoqing Ren Jian Sun
Microsoft Research
{kahe, v-xiangz, v-shren, jiansun}@microsoft.com

CVPR 2016 Best Paper, 被引用次数51507 (by 2020/7/28)
被引用次数95133(by 2021/11/08)



CNN经典模型
深度残差网络Deep ResNet

- 网络越深，效果越好?
  - 非线性激励函数,无限逼近定理
  - 梯度消失：随机梯度下降、归一化层等

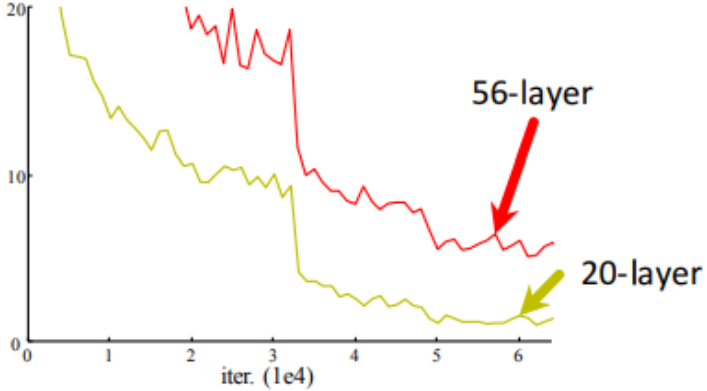
图像识别模型	面世时间	网络层次	备注
LeNet	1998	5	经典的手写数字识别模型
AlexNet	2012	8	ILSVRC 2012比赛第1名
VGG	2014	19	ILSVRC 2014比赛第2名
GoogLeNet	2014	22	ILSVRC 2014比赛第1名

注：ILSVRC是全球最权威的计算机视觉竞赛

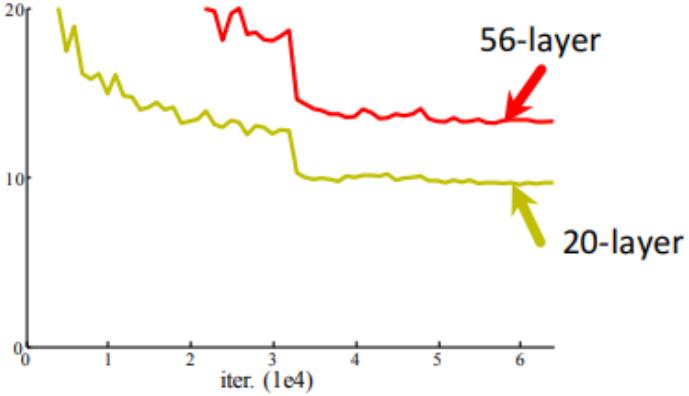
CNN经典模型
深度残差网络Deep ResNet

- 如果单纯的堆叠更深的网络：3x3卷积层

训练误差（%）



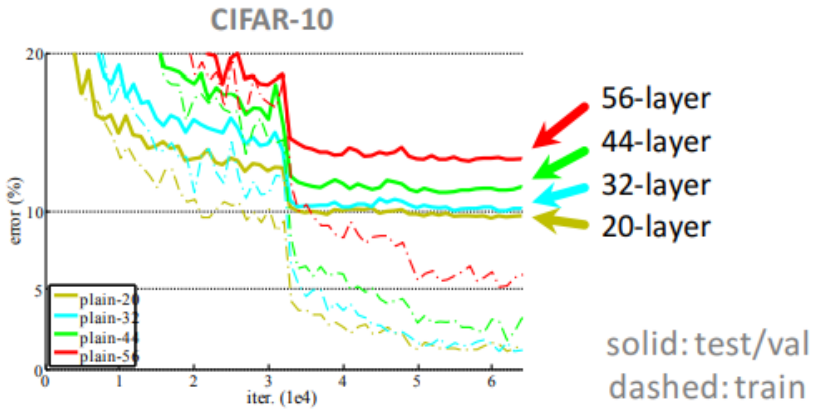
测试误差（%）

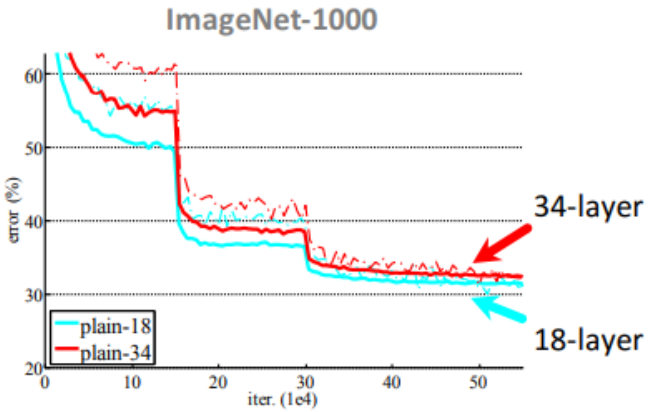


- 56层网络比20层网络有更高的训练和测试误差
- 非过拟合问题，而是网络更难训练

CNN经典模型
深度残差网络Deep ResNet

- 如果单纯的堆叠更深的网络：3x3卷积层

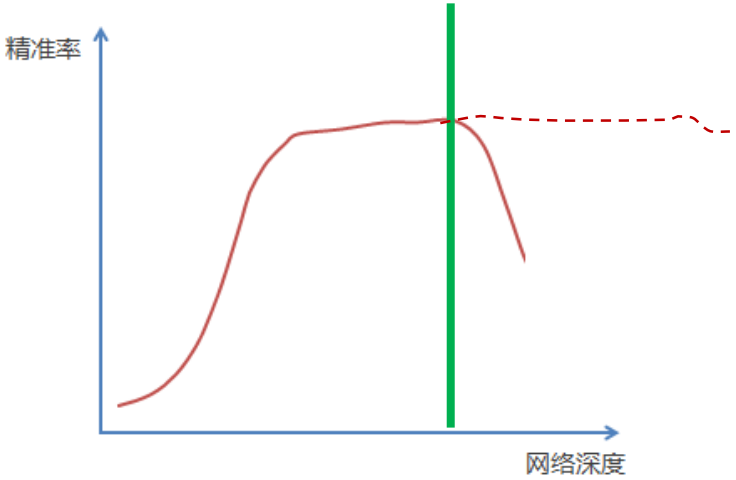




- 56层网络比20层网络有更高的训练和测试误差
- 在很多数据集上表现类似

CNN经典模型
深度残差网络Deep ResNet

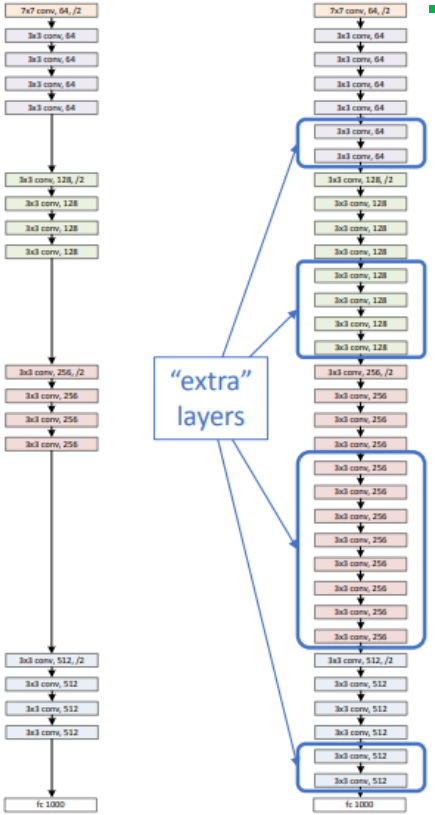
- 网络越深越好?



梯度消失导致无法对前面网络层的权重进行有效的调整

CNN经典模型
深度残差网络Deep ResNet

浅层模型18层



"extra" layers

深层副本34层

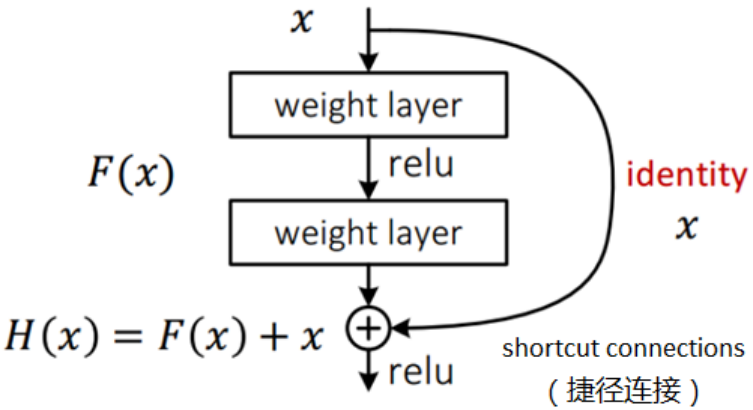
identity

- 深层模型表达能力更强，不应该产生更高的训练误差。
- 简单构建深层网络的方案
  - 从浅层模型中拷贝已有层
  - 将多余的网络层设为“恒等”映射
  - 至少是同样的训练误差
- 优化困难
  - 梯度消失

CNN经典模型
深度残差网络Deep ResNet

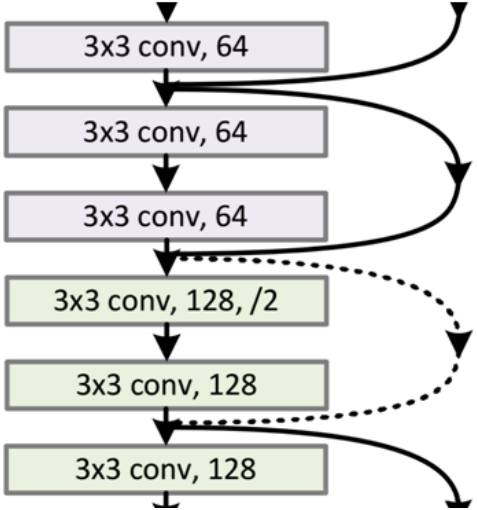
- 深度残差网络：
  - 学习一个复杂映射$x \to H(x)$
  - 学习一个残差映射$x \to F(x) = H(x) - x$
  - 有参考的学习
  - 在最优情况下，学习$F(x) = 0$相较于学习$H(x) = x$更容易

- 捷径连接（Shortcut Connections）
  - 训练目标是$F(x) = 0$
  - 允许跨多层进行连接
  - 保证了网络层数加深不降低准确率



CNN经典模型
深度残差网络Deep ResNet

- 残差单元



通道数目相同时， $H(x) = F(x) + x$

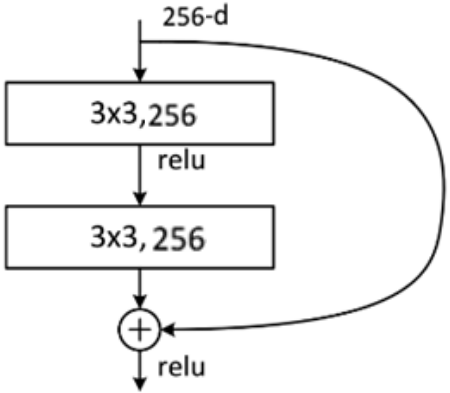
通道数目不同时， $H(x) = F(x) + Wx$

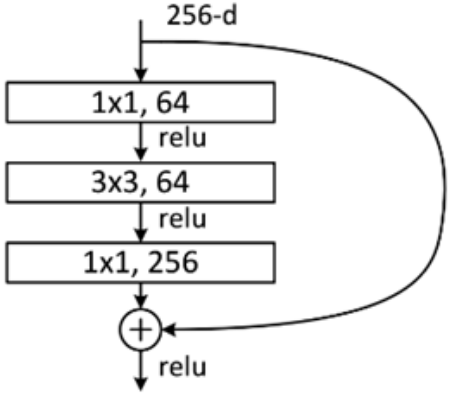
W为卷积操作，
用来调整维度

CNN经典模型
深度残差网络Deep ResNet

- 残差单元

3x3x256x256x2=1179648



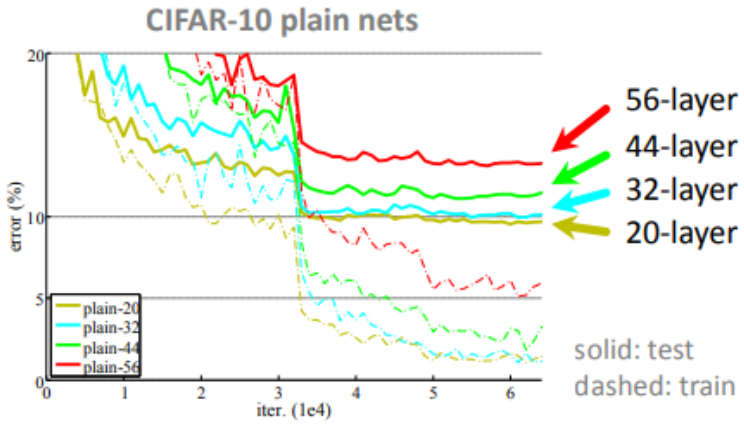


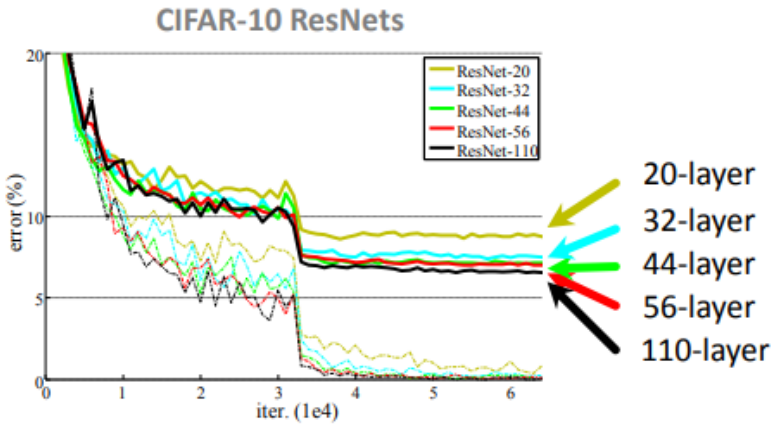
1x1x256x64+3x3x64z64+1x1x64x256=69632

用于深层网络（>100层）
减少计算和参数量

CNN经典模型
深度残差网络Deep ResNet

- CIFAR-10上的实验结果

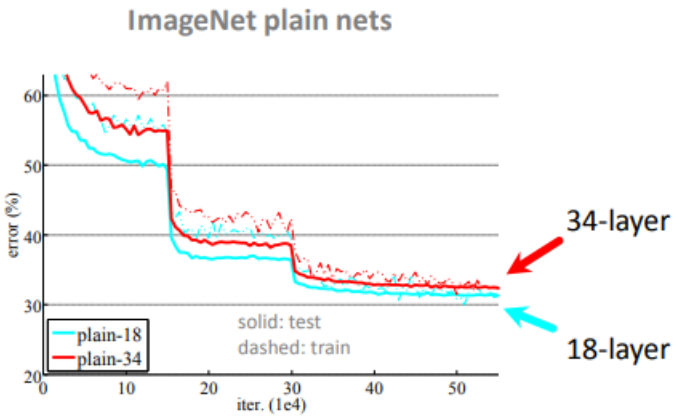


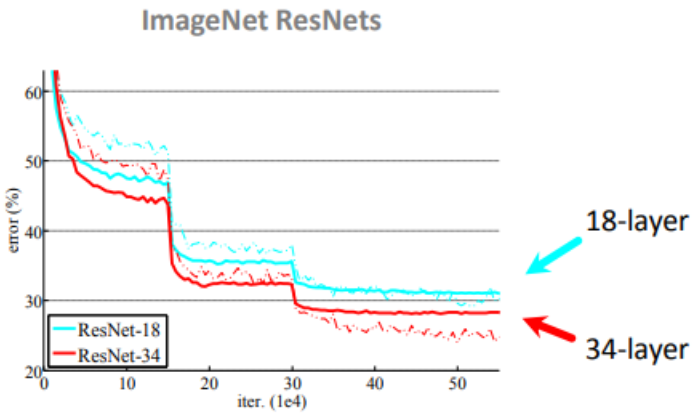


- 深度残差网络更容易训练
- 获得更低的训练误差和测试误差

CNN经典模型
深度残差网络Deep ResNet

- ImageNet上的实验结果





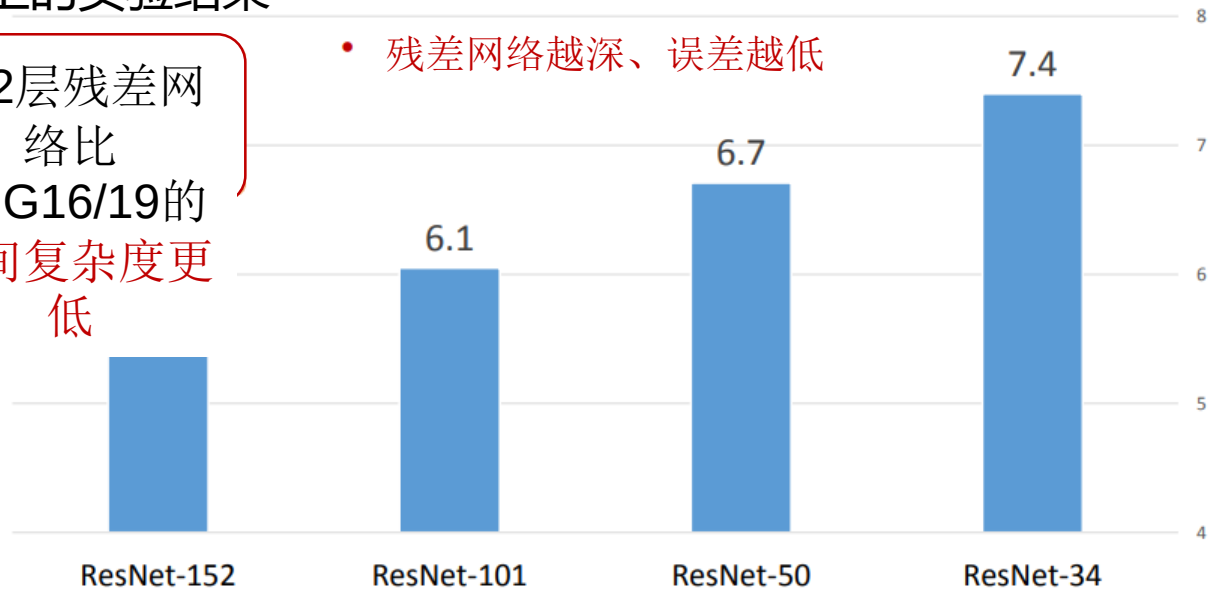
- 深度残差网络更容易训练
- 获得更低的训练误差和测试误差

CNN经典模型
深度残差网络Deep ResNet

- ImageNet上的实验结果

152层残差网络比VGG16/19的时间复杂度更低

- 残差网络越深、误差越低

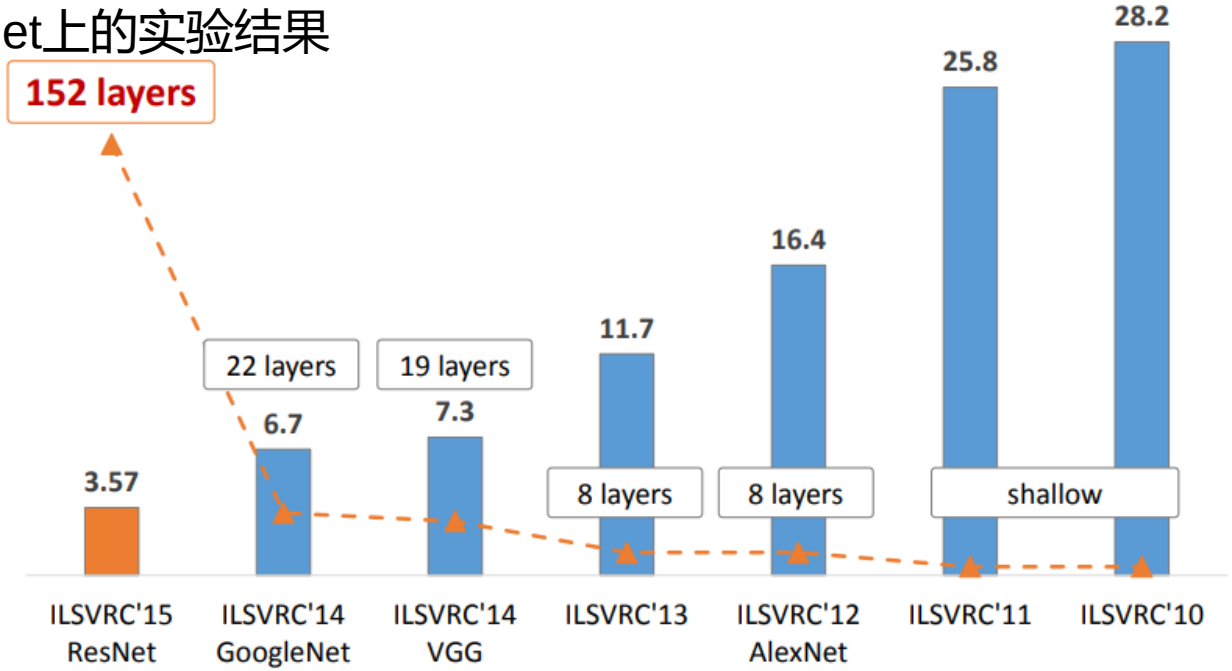


10次裁剪、top-5验证误差

Slides from Kaiming He

CNN经典模型
深度残差网络Deep ResNet

- ImageNet上的实验结果



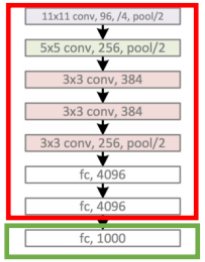
top-5误差

Slides from Kaiming He

经典卷积神经网络

60M

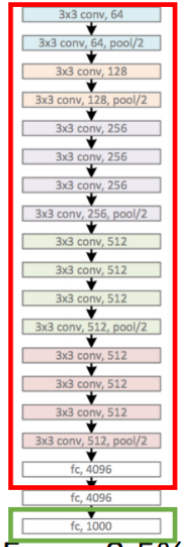
2012: AlexNet
5 conv. layers



Error: 15.3%

140M

2014: VGG
16 conv. layers



Error: 8.5%

5M

2015: GoogLeNet
22 conv. layers

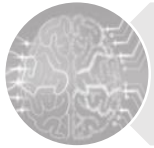
Error: 7.8%

1M

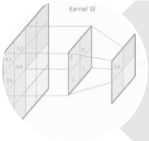
2016: ResNet
>100 conv. layers

Error: 4.4%

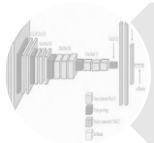
Outline



Introduction



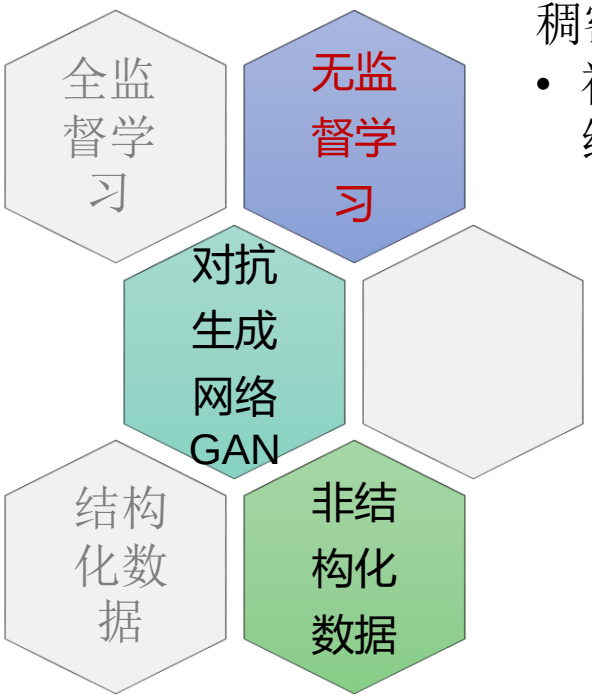
CNN Architectures



Classical CNN Models

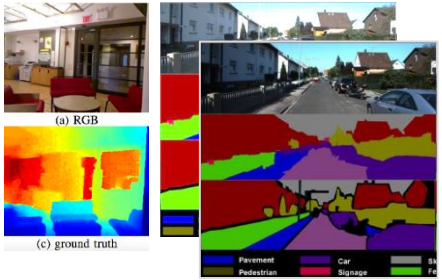
Discussion

总结与讨论



稠密有标数据获取困难：

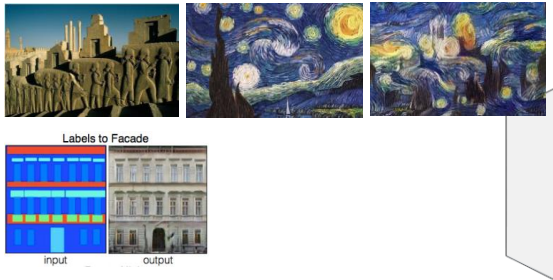
- 视频分割、深度估计、三维分割



总结与讨论

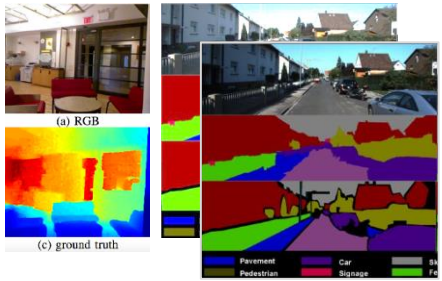
数据生成：

- 纹理外观、几何形状、空间关系…



稠密有标数据获取困难：

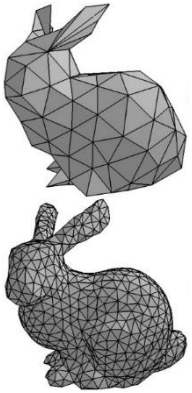
- 视频分割、深度估计、三维分割

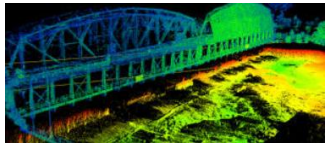


总结与讨论

数据生成：

- 纹理外观、几何形状、空间关系…





- 三维点云、图、网格