深蓝学院
shenlanxueyuan.com

深度学习：从理论到实践

第一章：深度学习理论（二）
——卷积神经网络（上）

课程大纲

深蓝学院
shenlanxueyuan.com

- ✓ 基本概念
- ✓ 发展历程
- ✓ 网络特点

课程大纲

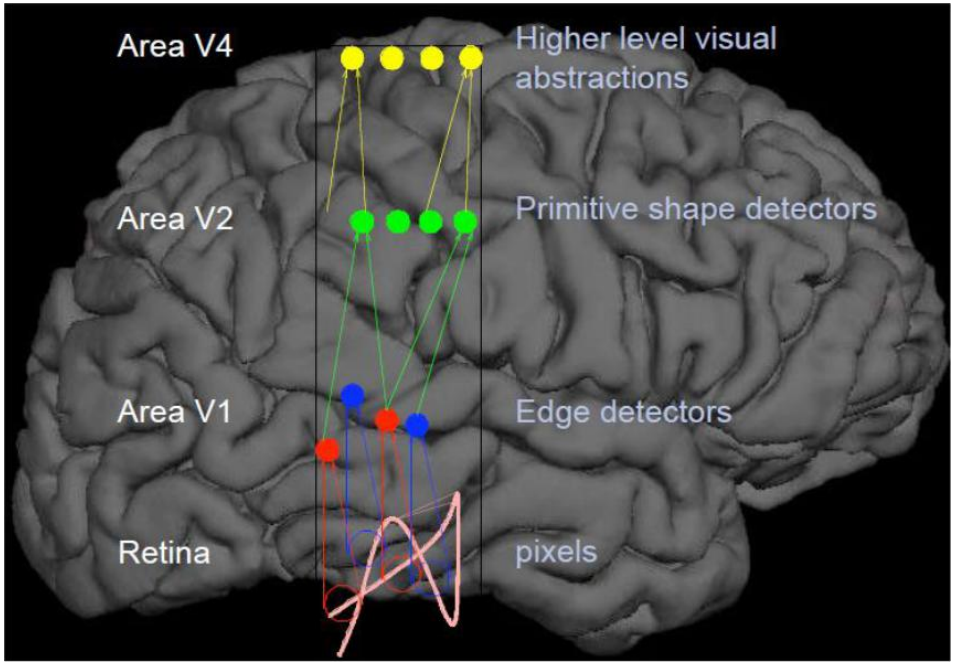
深蓝学院
shenlanxueyuan.com

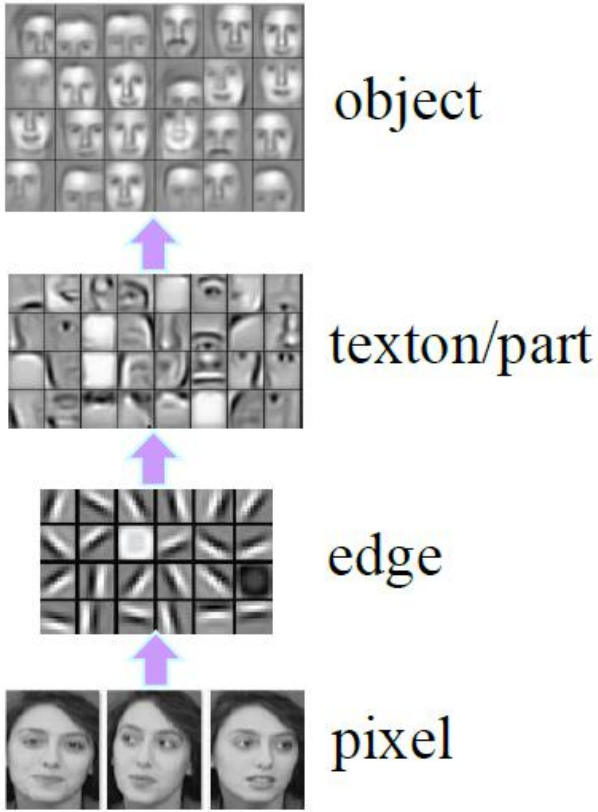
- ✓ 基本概念
- ✓ 发展历程
- ✓ 网络特点

基本概念

深蓝学院
shenlanxueyuan.com

人类视觉系统的层次结构





基本概念

深蓝学院 shenlanxueyuan.com

局部感受野

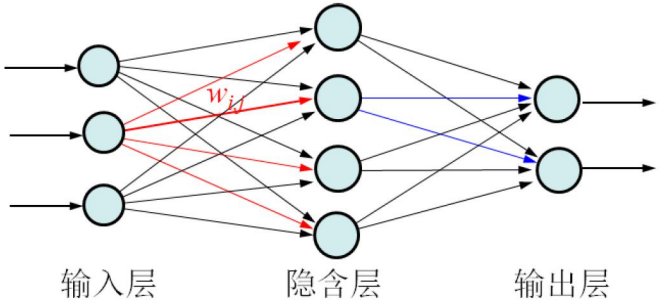
- 1962年，Hubel和Wiesel通过对猫视觉皮层细胞的研究（局部敏感和方向选择的神经元具有独特的网络结构），提出了**感受野**（receptive field）的概念。
- 1984年，日本学者Fukushima基于感受野概念提出的神经认知机，是感受野概念在人工神经网络领域的首次应用。
- 视觉系统**局部感受野**：**人对外界的认知是从局部到全局的**。视觉皮层的神经元就是局部接受信息的（即只响应某些特定区域的刺激）。

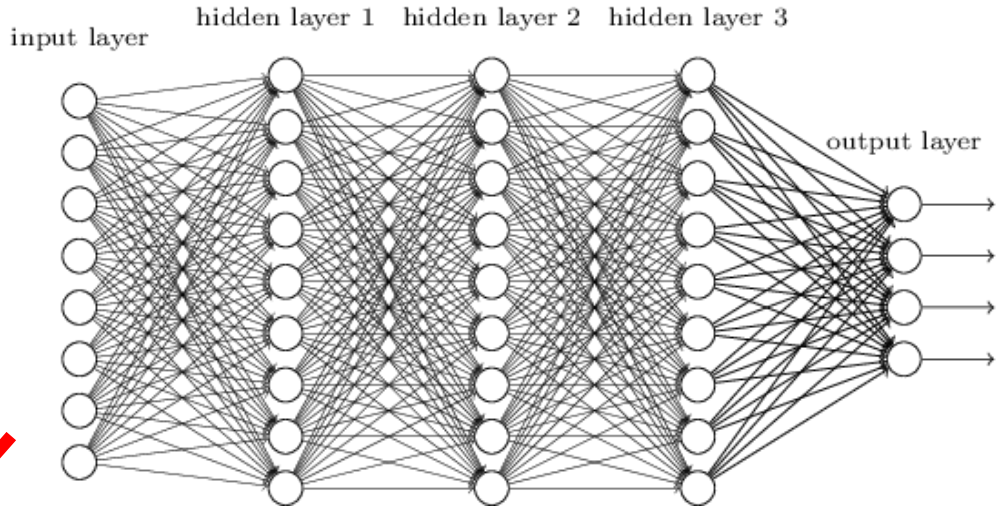
基本概念

深蓝学院
shenlanxueyuan.com

从神经网络到卷积神经网络

（convolutional neural network，CNN）





参数数目巨大

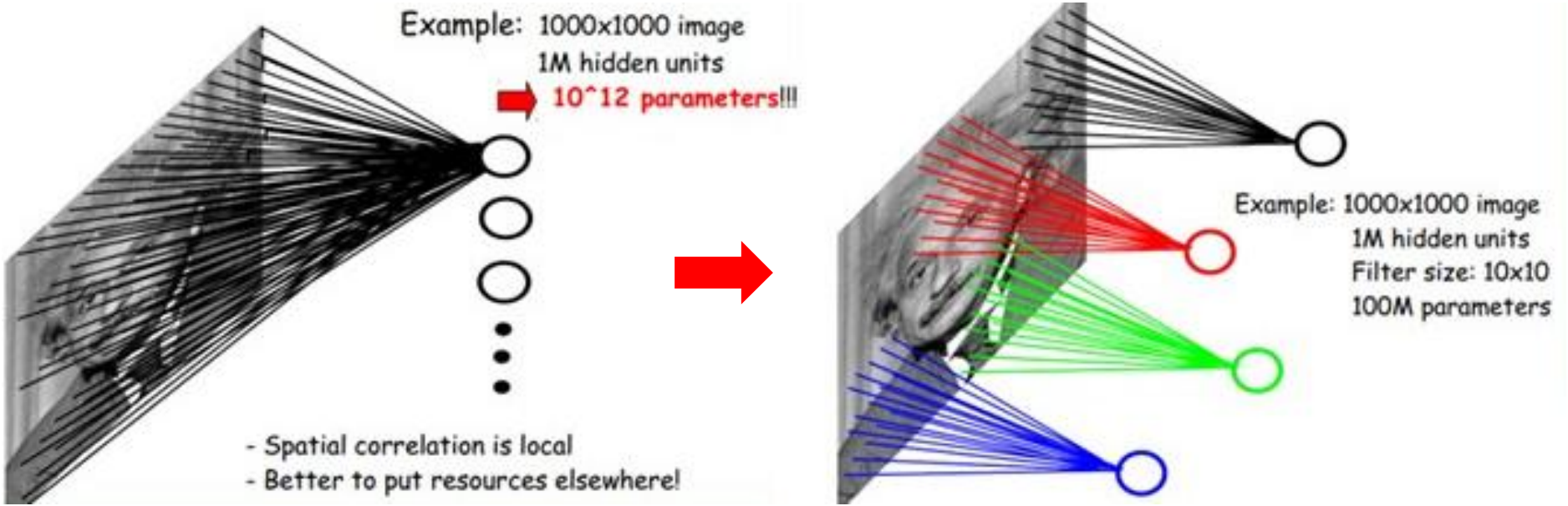
难以训练 — 梯度消失/爆炸

基本概念

从神经网络到卷积神经网络

全连接

局部连接



基本概念

深蓝学院 shenlanxueyuan.com

卷积的数学描述

一维卷积

$$\mathrm{w}(\mathrm{x}) * f(x) = \int_{-\infty}^{\infty} w(u) f(x-u)\, du$$

二维卷积

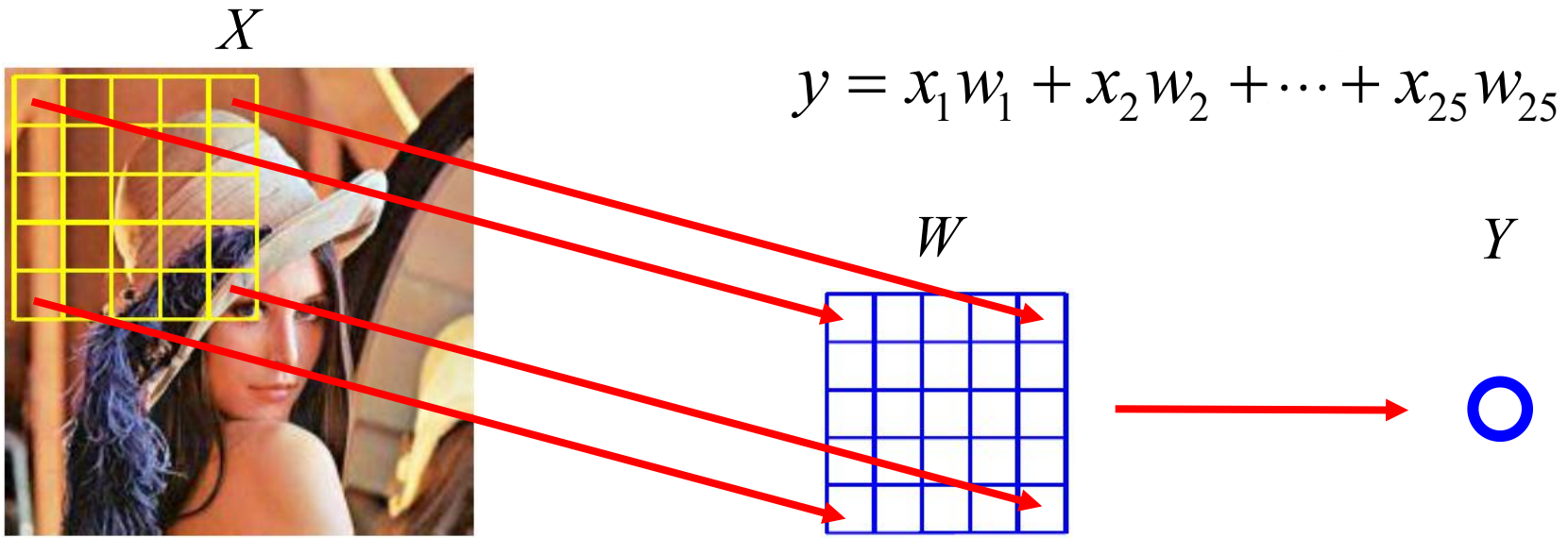
$$\mathrm{w}(\mathrm{x}, y) * f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} w(u, v) f(x-u, y-v)\, du\, dv$$

二维离散卷积

$$W(x, y) * F(x, y) = \sum_{u} \sum_{v} W(u, v) F(x-u, y-v)$$

基本概念

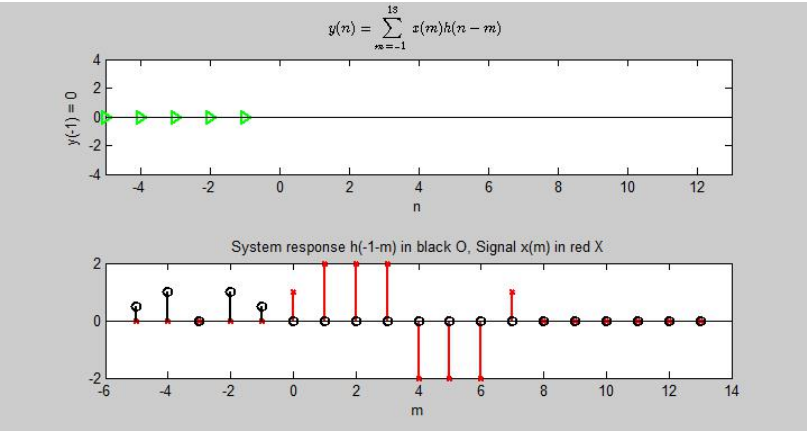
局部连接 — 通过卷积操作实现

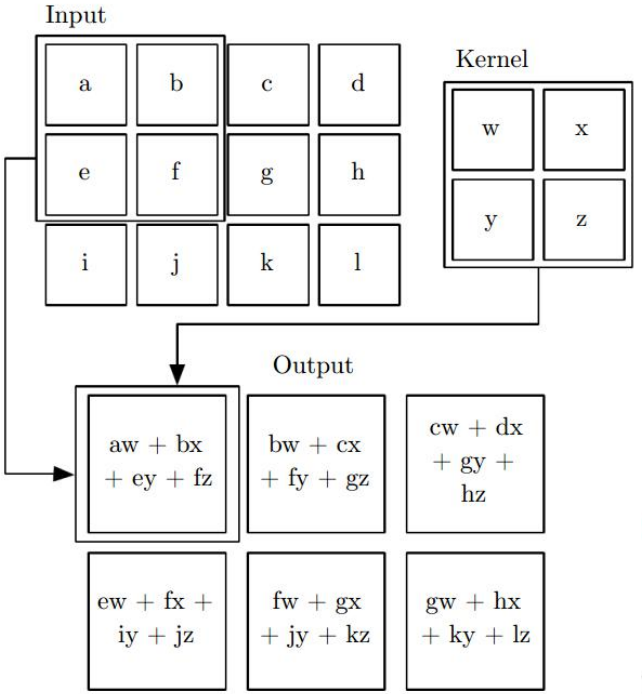


关于卷积运算与相关运算的对比可参考https://www.zhihu.com/question/32067344?sort=created

基本概念

卷积的示意图





本质：计算信号（如图像）
对某种特定模式的响应

关于二维卷积的动态演示可参考http://blog.csdn.net/liyuan123zhouhui/article/details/60139790
关于卷积的具体介绍可参考http://blog.sina.com.cn/s/blog_7445c2940102wmrp.html

基本概念

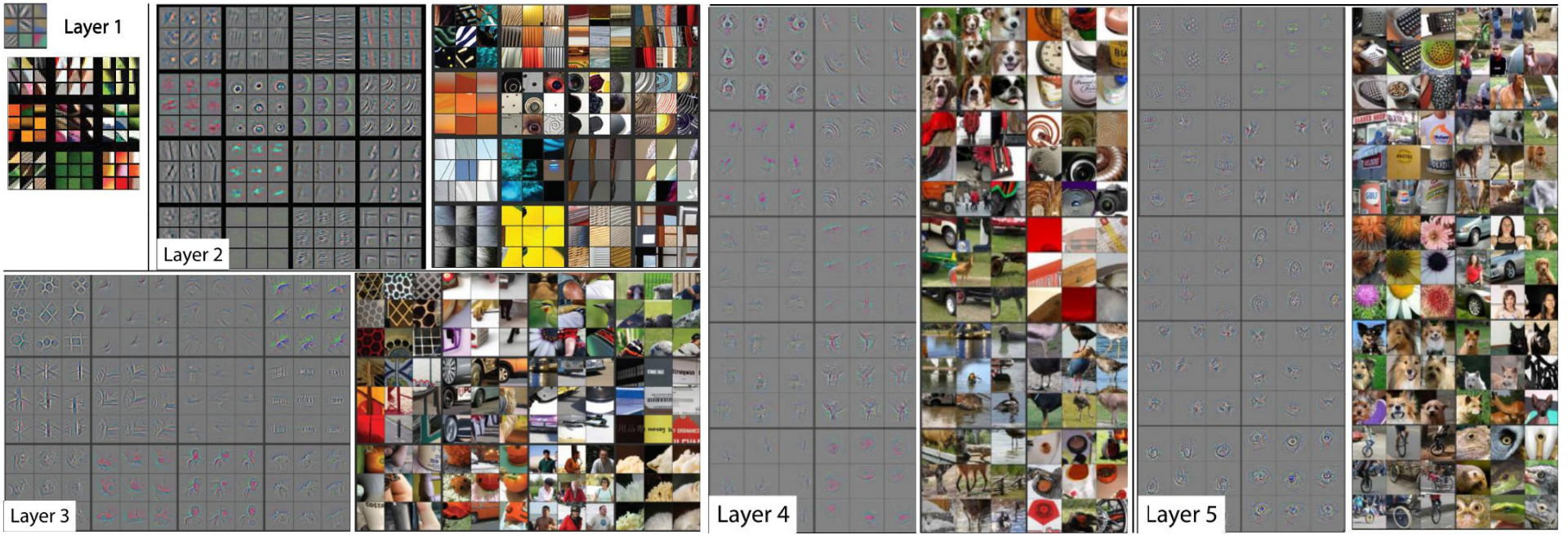
卷积的作用

- 类似于数字图像处理中的各类算子（Sobel，Robert，Laplacian）。
- 用于提取特征，因此将得到的结果称为特征图（feature map）。

基本概念

深蓝学院
shenlanxueyuan.com

卷积的作用

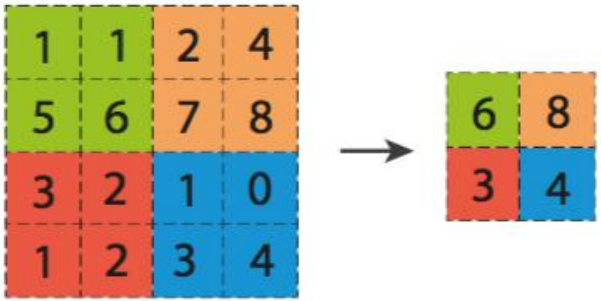


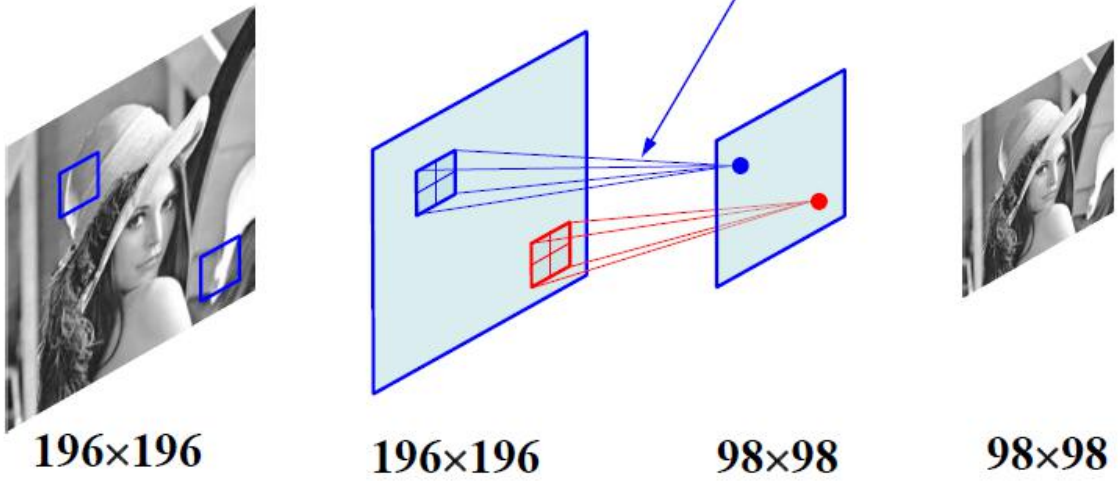
基本概念

深蓝学院 shenlanxueyuan.com

池化操作（pooling）

- 图像区域内关于某个特征的统计聚合。
- 两种方式：区域内取最大值或平均。
- 是一种下采样操作。





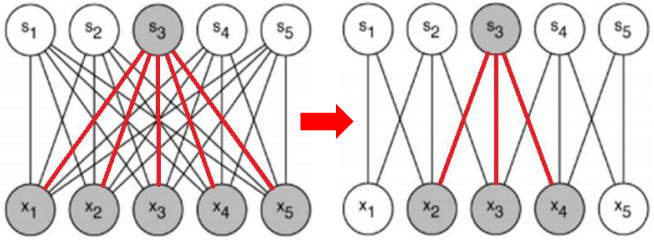
基本概念

深蓝学院 shenlanxueyuan.com

卷积神经网络与多层感知器的对比

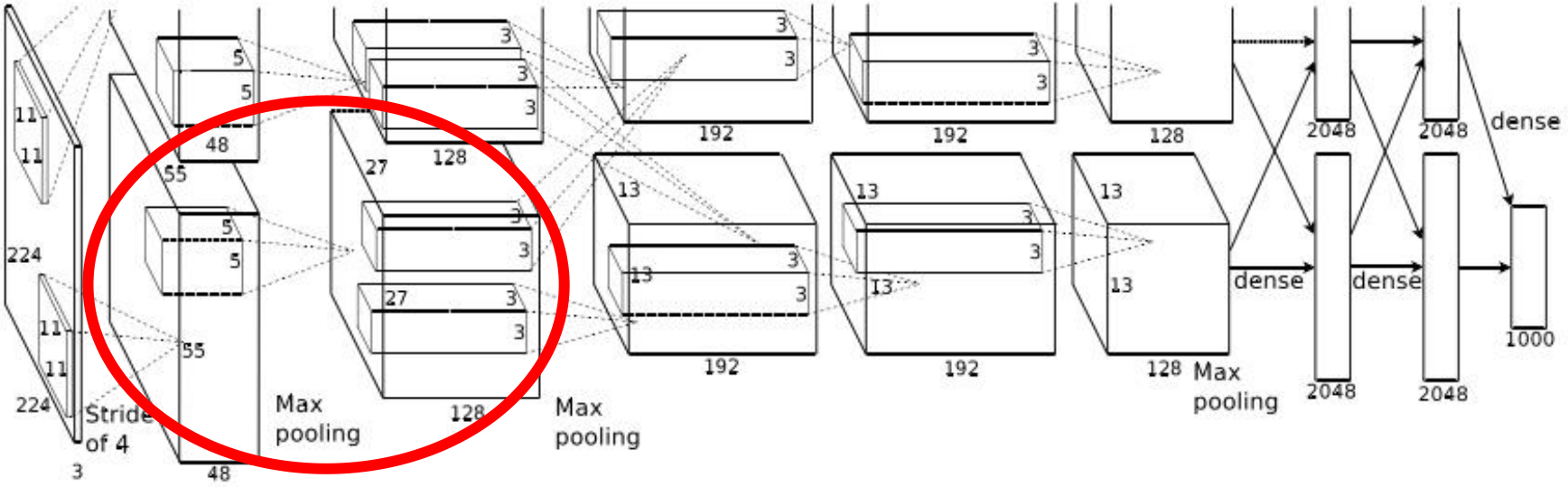
	多层感知器	卷积神经网络
隐含层数目	通常只有一层	多层
连接方式	全连接	局部连接、稀疏连接
基本操作	乘法、加法、非线性映射	乘法、加法、非线性映射、池化
相邻两层之间的参数数目	前一层点数*后一层节点数	前一层特征图数目*卷积核尺寸*后一层特征图数目
计算负担	一般较高	卷积操作容易高效实现

与图像本身的尺寸无关
参数数目大幅度降低



基本概念

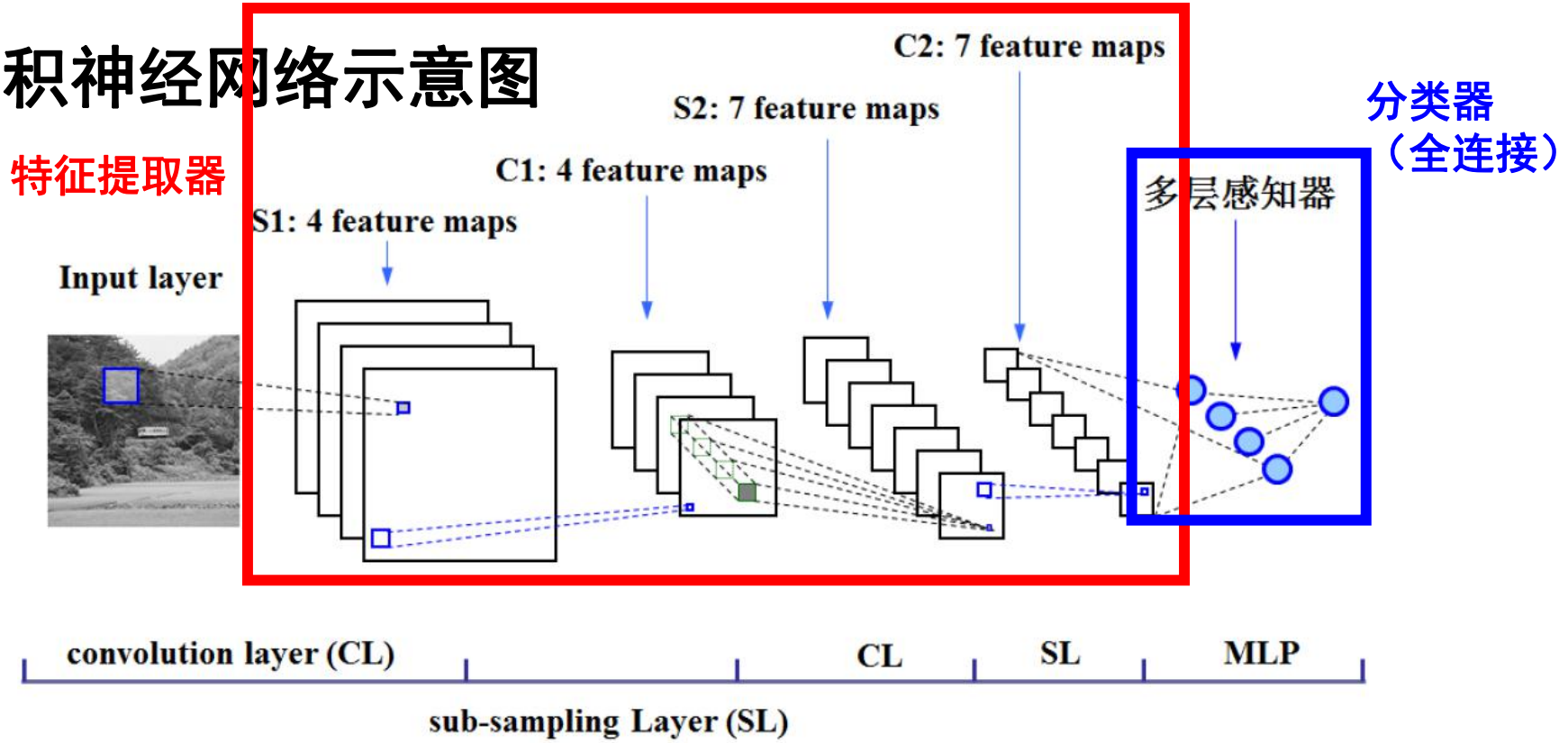
卷积神经网络与多层感知器的对比



$$48\times5\times5\times128=153600$$

基本概念

卷积神经网络示意图



课程大纲

深蓝学院
shenlanxueyuan.com

- ✓ 基本概念
- ✓ 发展历程
- ✓ 网络特点

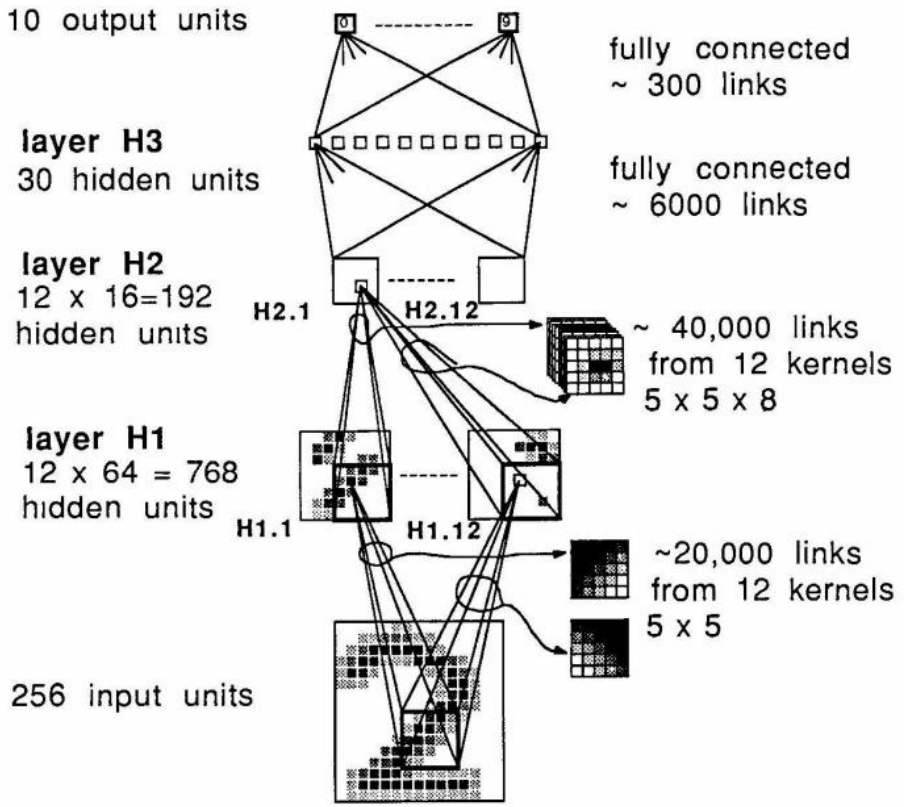
发展历程

深蓝学院 shenlanxueyuan.com

LeNet — 图像分类

- 1989年由Yann LeCun提出。
- 应用于手写数字（邮政编码）识别。
- 卷积神经网络最初的形式，

  也是最基本的形式，目前仍在沿用。



Y. LeCun et al., Backpropagation applied to handwritten zip code recognition. Neural Computation 1989.

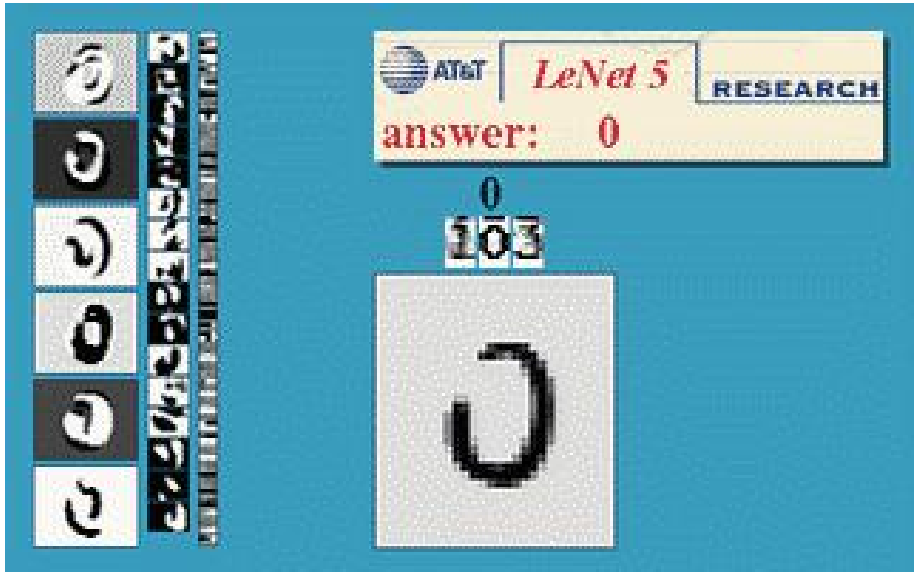
发展历程

深蓝学院
shenlanxueyuan.com

LeNet — 图像分类

- 1989年由Yann LeCun提出。
- 应用于手写数字（邮政编码）识别。
- 卷积神经网络最初的形式，

  也是最基本的形式，目前仍在沿用。



Y. LeCun et al., Backpropagation applied to handwritten zip code recognition. Neural Computation 1989.

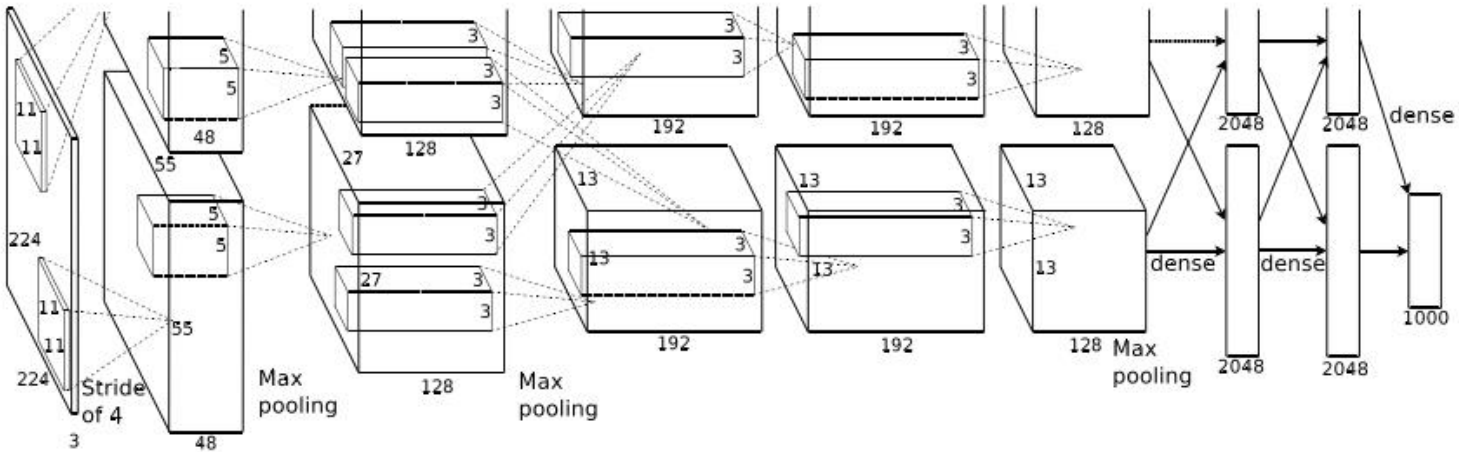
发展历程

深蓝学院 shenlanxueyuan.com

AlexNet — 图像分类

Hinton是DBN的发明者，
但却凭借CNN获奖。

- LeNet虽然取得了一定的效果，但并未受到广泛关注。
- 2012年的ImageNet比赛中，Alex Krizhevsky等对其进行了成功的应用。



A. Krizhevsk et al., ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012.

发展历程

深蓝学院 shenlanxueyuan.com

AlexNet的重要意义

- 证实了卷积神经网络可以利用大规模训练数据得到优秀的分类效果，从此以后卷积神经网络在计算机视觉领域便得到飞速发展和应用。
- 奠定了GPU在深度学习中的重要作用（同时也使NVIDIA公司得到了巨大的发展）。
- 多种网络设置和训练的技巧，如ReLU 、Dropout等，使深层网络的训练成为可能。
- 网络学习到的特征对分类效果具有一定的解释性。

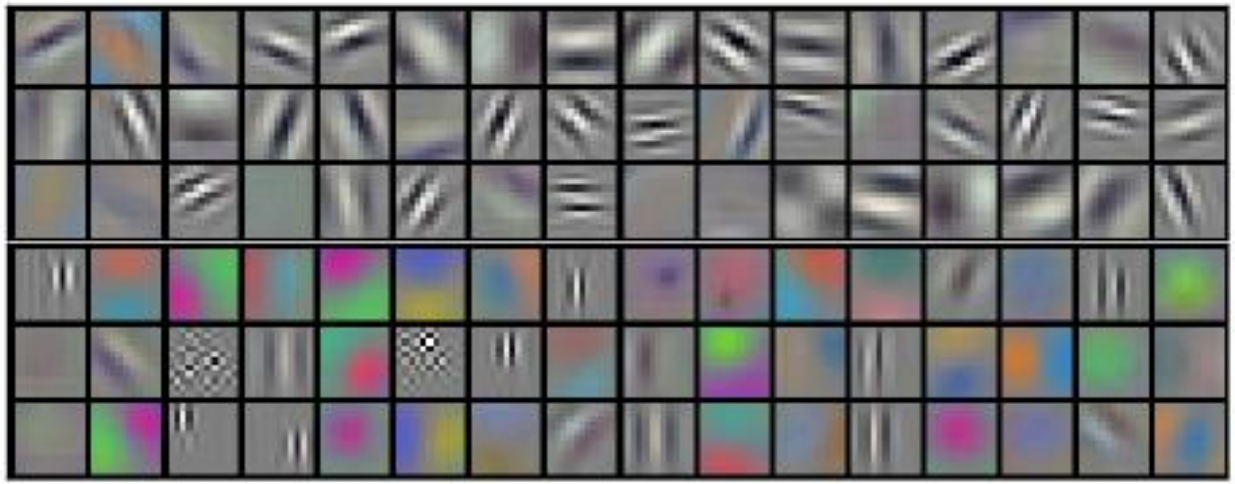
A. Krizhevsk et al., ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012.

发展历程

深蓝学院
shenlanxueyuan.com

AlexNet的重要意义 — 卷积神经网络用于特征提取

学习到的卷积核

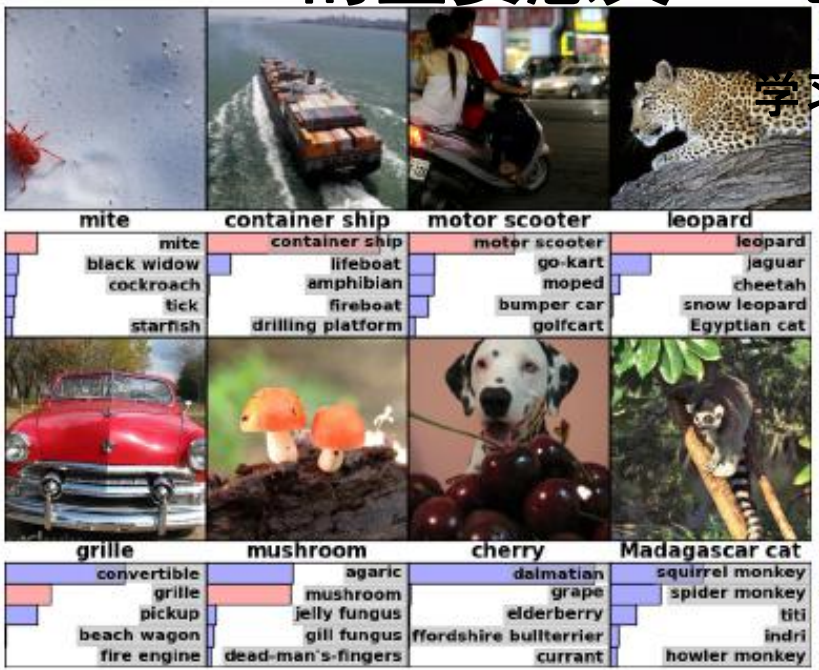


A. Krizhevsk et al., ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012.

发展历程

深蓝学院 shenlanxueyuan.com

AlexNet的重要意义 — 卷积神经网络用于特征提取



学习到的卷积核

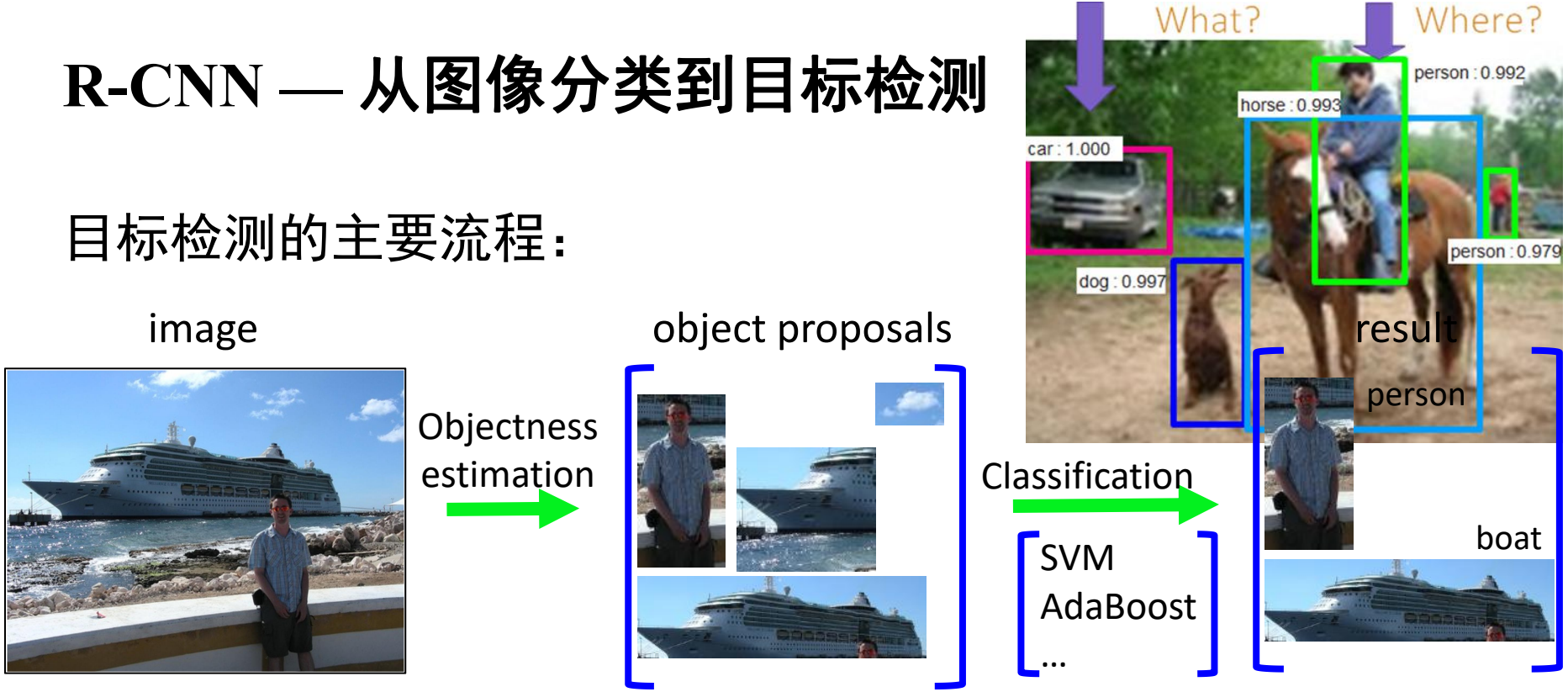
A. Krizhevsk et al., ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012.

发展历程

深蓝学院 shenlanxueyuan.com

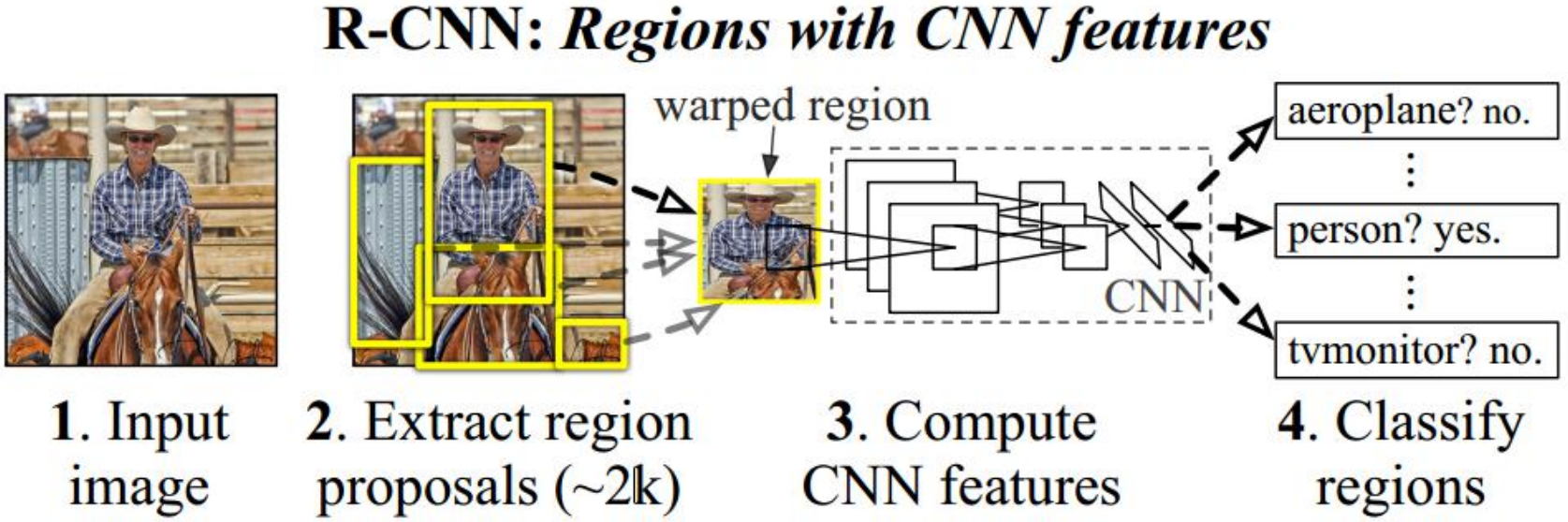
R-CNN — 从图像分类到目标检测

目标检测的主要流程：



发展历程

R-CNN — 从图像分类到目标检测



R. Girshick et al., Rich feature hierarchies for accurate object detection and semantic segmentation. CVPR 2014.

发展历程

R-CNN的重要意义

- 证实了可以将图像的局部作为基本研究对象，将原问题转化为图像分类问题，利用卷积神经网络以图像分类的方式进行解决。
- 引出了卷积神经网络在计算机视觉中极为丰富的各类应用：
  - fast R-CNN，faster R-CNN，SSD, Mask R-CNN……
  - 图像分割、图像描述、目标解析……

R. Girshick et al., Rich feature hierarchies for accurate object detection and semantic segmentation. CVPR 2014.

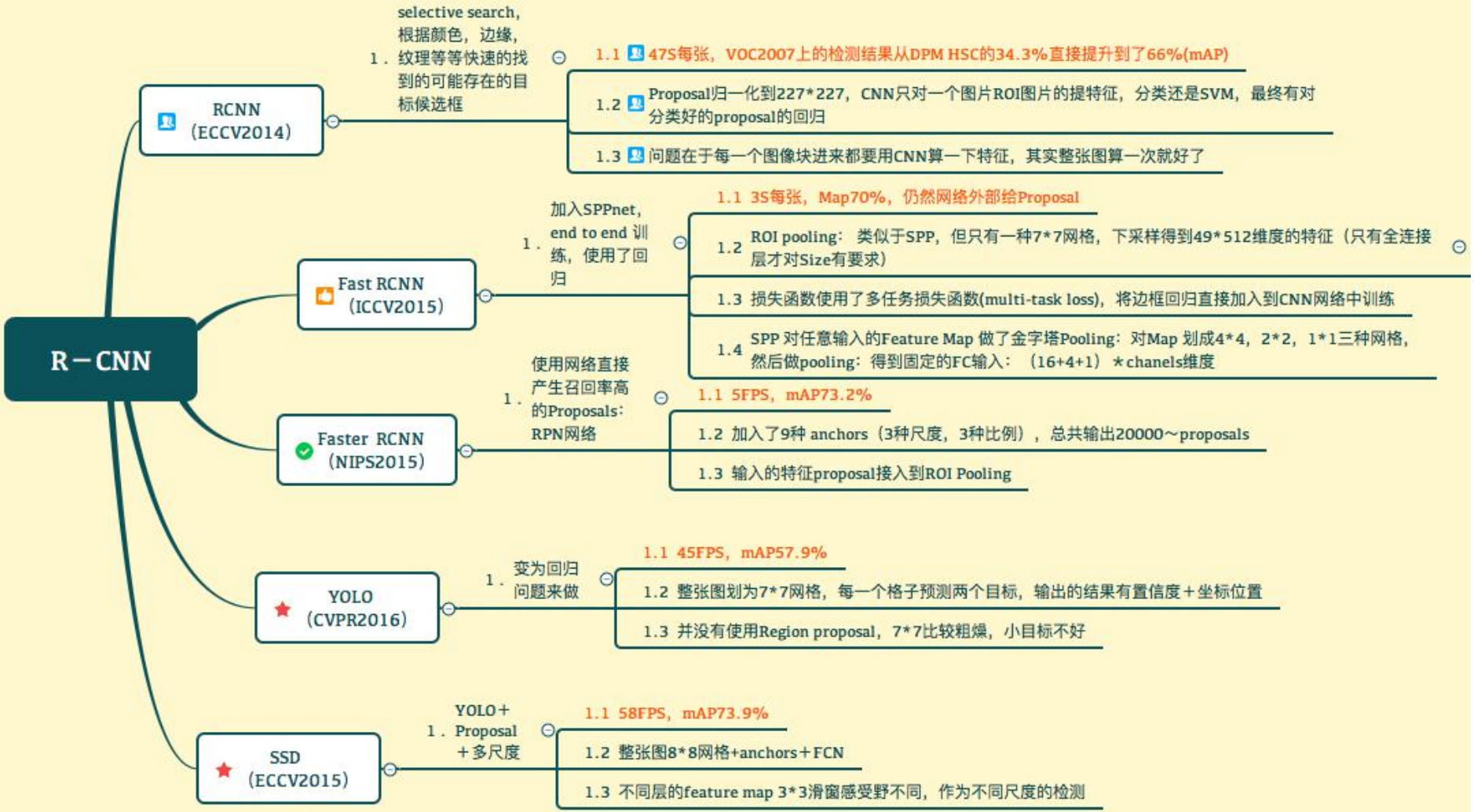
发展历程

深蓝学院 shenlanxueyuan.com

R-CNN的重要意义

Ross B. Girshick et al. "Deformable Part Models are Convolutional Neural Networks." Techical report.

- 在使用CNN之前，目标检测主要采用的方法是DPM (Deformable parts models)，精度只能到43%。
- Ross B. Girshick et al. “Rich feature hierarchies for accurate object detection and semantic segmentation.” CVPR 2014. (66.0%, 47s)
- Ross B. Girshick. “Fast r-cnn.” CVPR 2015. (66.9%, 0.22s)
- Shaoqing Ren, Kaiming He, Ross B. Girshick et al. “Faster R-CNN: Towards real-time object detection with region proposal networks.” NIPS 2015. (73.2%, 0.196s including all steps)


R－CNN
RCNN
(ECCV2014)
1. selective search，根据颜色，边缘，纹理等等快速的找到的可能存在的目标候选框
1.1 47S每张，VOC2007上的检测结果从DPM HSC的34.3%直接提升到了66%(mAP)
1.2 Proposal归一化到227*227，CNN只对一个图片ROI图片的提特征，分类还是SVM，最终有对分类好的proposal的回归
1.3 问题在于每一个图像块进来都要用CNN算一下特征，其实整张图算一次就好了
Fast RCNN
(ICCV2015)
1. 加入SPPnet，end to end 训练，使用了回归
1.1 3S每张，Map70%，仍然网络外部给Proposal
1.2 ROI pooling：类似于SPP，但只有一种7*7网格，下采样得到49*512维度的特征（只有全连接层才对Size有要求）
1.3 损失函数使用了多任务损失函数(multi-task loss)，将边框回归直接加入到CNN网络中训练
1.4 SPP 对任意输入的Feature Map 做了金字塔Pooling：对Map 划成4*4，2*2，1*1三种网格，然后做pooling：得到固定的FC输入：（16+4+1）＊chanels维度
Faster RCNN
(NIPS2015)
1. 使用网络直接产生召回率高的Proposals：RPN网络
1.1 5FPS，mAP73.2%
1.2 加入了9种 anchors（3种尺度，3种比例），总共输出20000～proposals
1.3 输入的特征proposal接入到ROI Pooling
YOLO
(CVPR2016)
1. 变为回归问题来做
1.1 45FPS，mAP57.9%
1.2 整张图划为7*7网格，每一个格子预测两个目标，输出的结果有置信度＋坐标位置
1.3 并没有使用Region proposal，7*7比较粗燥，小目标不好
SSD
(ECCV2015)
1. YOLO＋Proposal＋多尺度
1.1 58FPS，mAP73.9%
1.2 整张图8*8网格+anchors＋FCN
1.3 不同层的feature map 3*3滑窗感受野不同，作为不同尺度的检测

发展历程

深蓝学院 shenlanxueyuan.com

FCN

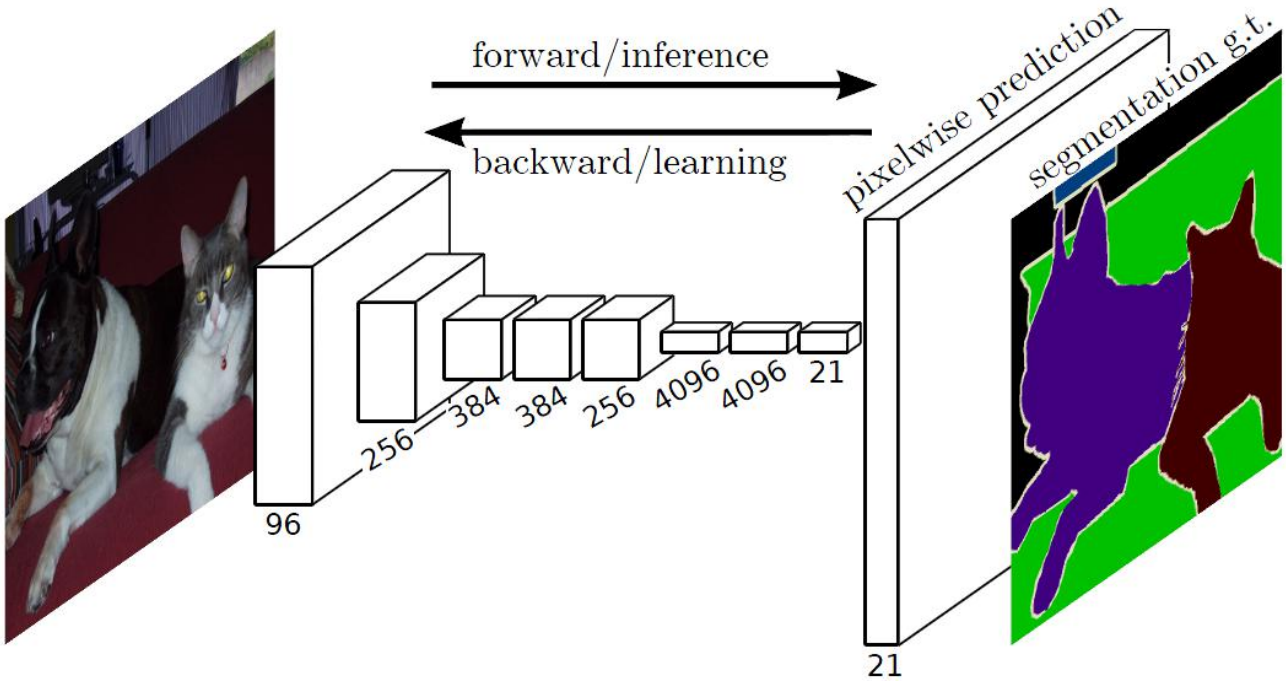
- 在FCN出现之前的图像分割，一般将图像的局部作为基本研究对象，如像素、超像素、图像块（矩形区域或者不规则区域）等，通过对它们进行分类来实现图像分割。
- FCN则采用“整幅图像输入，整幅分割结果输出”这种“端到端”（end-to-end）式的结构，能够接受任意尺寸的输入图像。
- 第一个“端到端”式的图像分割模型。

J. Long et al., Fully convolutional networks for semantic segmentation. CVPR 2015.

发展历程

深蓝学院
shenlanxueyuan.com

FCN

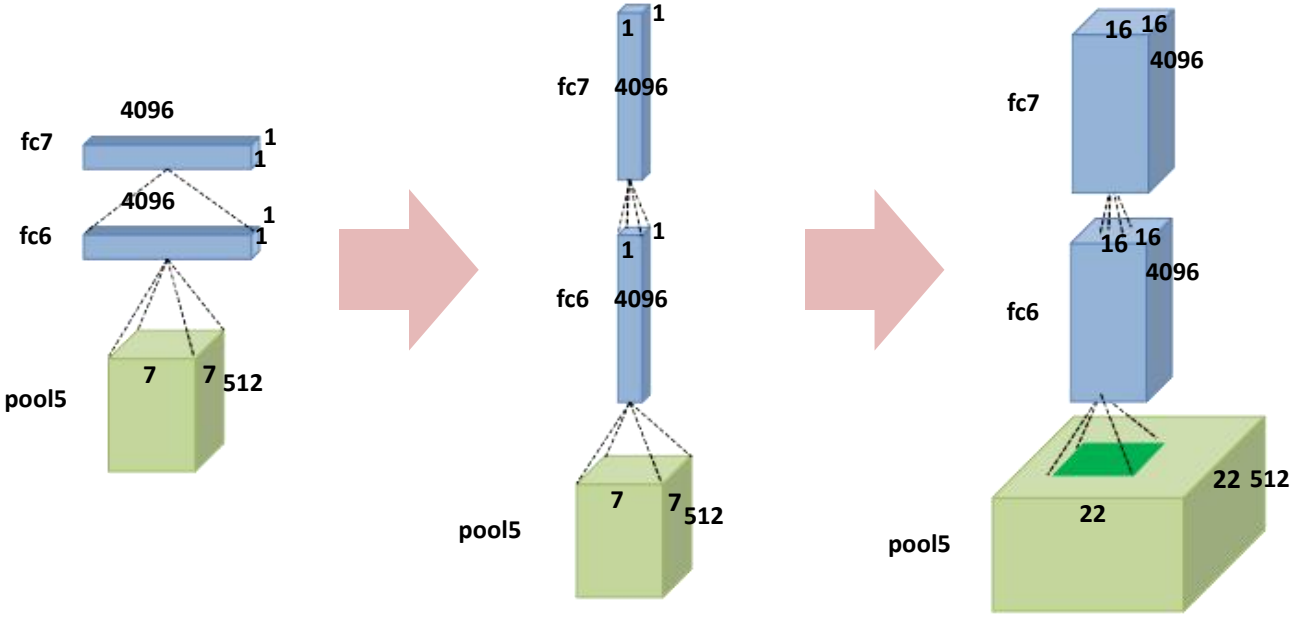


J. Long et al., Fully convolutional networks for semantic segmentation. CVPR 2015.

发展历程

深蓝学院 shenlanxueyuan.com

FCN



Fully connected layers

Convolution layers

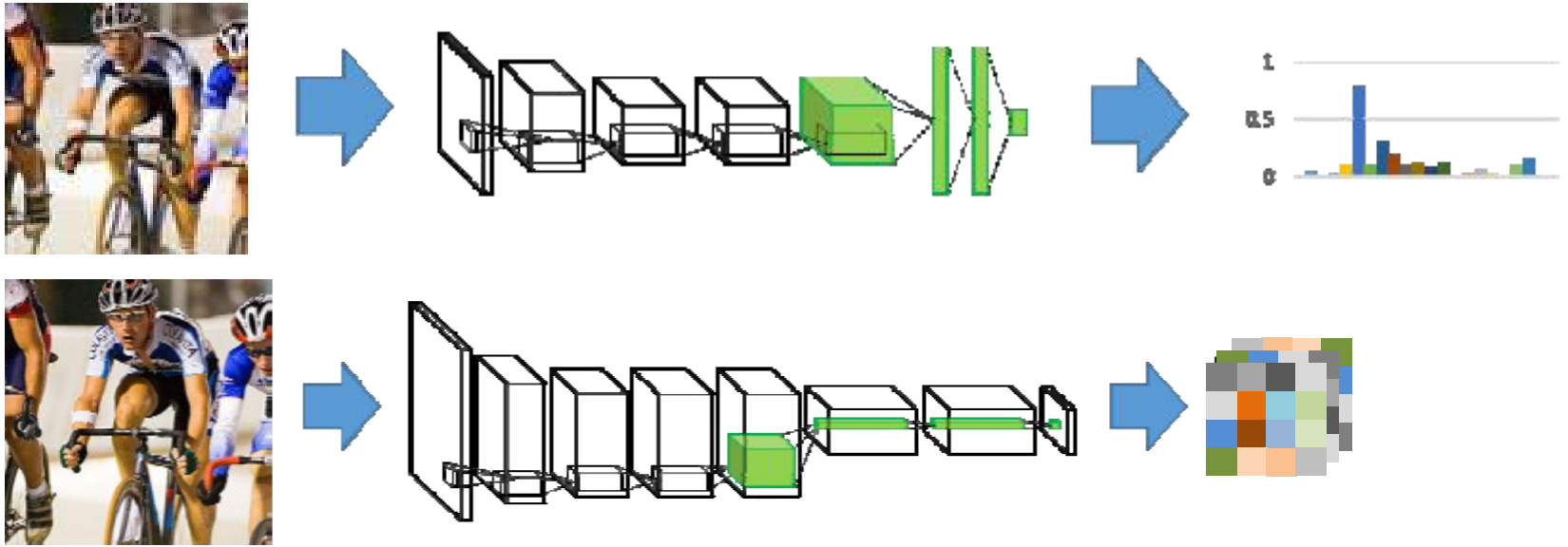
For the larger Input field

J. Long et al., Fully convolutional networks for semantic segmentation. CVPR 2015.

发展历程

深蓝学院
shenlanxueyuan.com

FCN



J. Long et al., Fully convolutional networks for semantic segmentation. CVPR 2015.

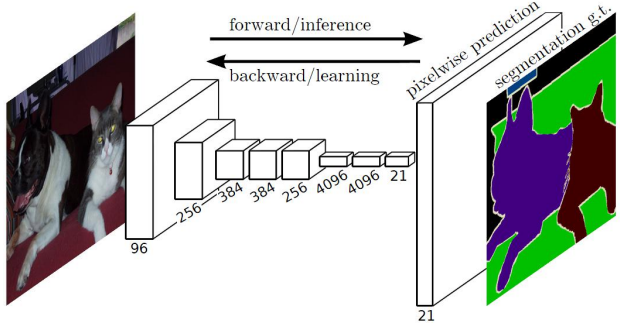
发展历程

FCN的重要意义

- 充分保留了全局的语义信息。
- 减少了重复计算（如果采用图像块，图像块之间会有overlap），提高了训练和测试的效率。
- 去除了对输入图像尺寸的限制（因为没有全连接层，每层的节点数目于参数数目不再直接相关）。
- 全卷积网络成为许多后续网络模型的重要基础：Deeplab，CRFRNN…
- “端到端”式结构成为后续网络设计的一项重要标准。

发展历程

Conv-Deconv Net



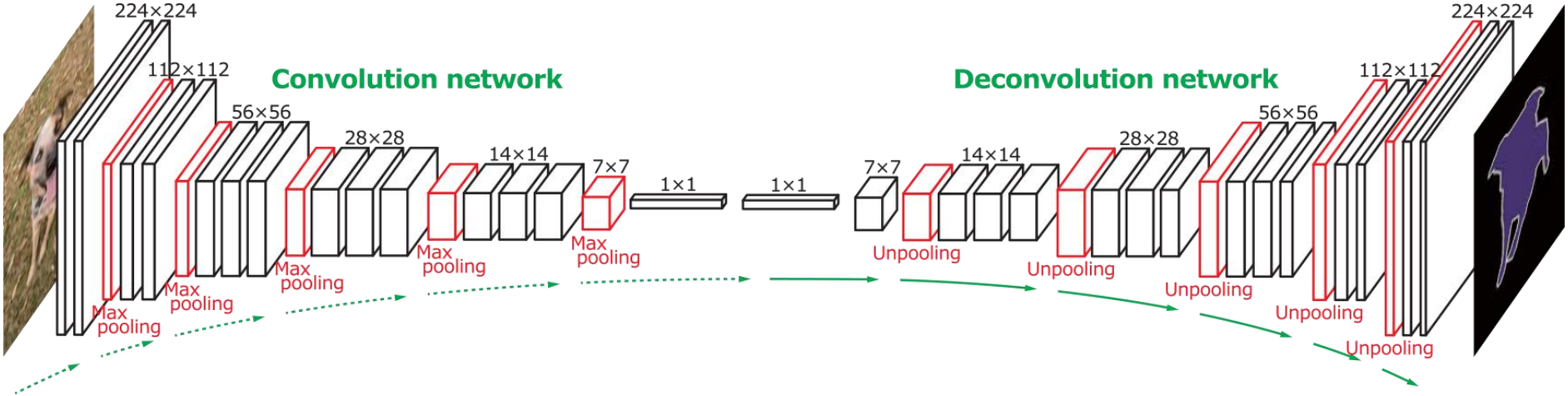
- FCN中特征图经过逐级下采样，尺寸大大减小，最后经过一次性的上采样恢复输入尺寸。
- 整个过程损失了太多的细节信息。
- Conv-Deconv Net采用与卷积神经网络对称的结构，通过逐级上采样来恢复特征图的尺寸。

H. Noh et al., Learning deconvolution network for semantic segmentation. ICCV 2015.

发展历程

深蓝学院 shenlanxueyuan.com

Conv-Deconv Net

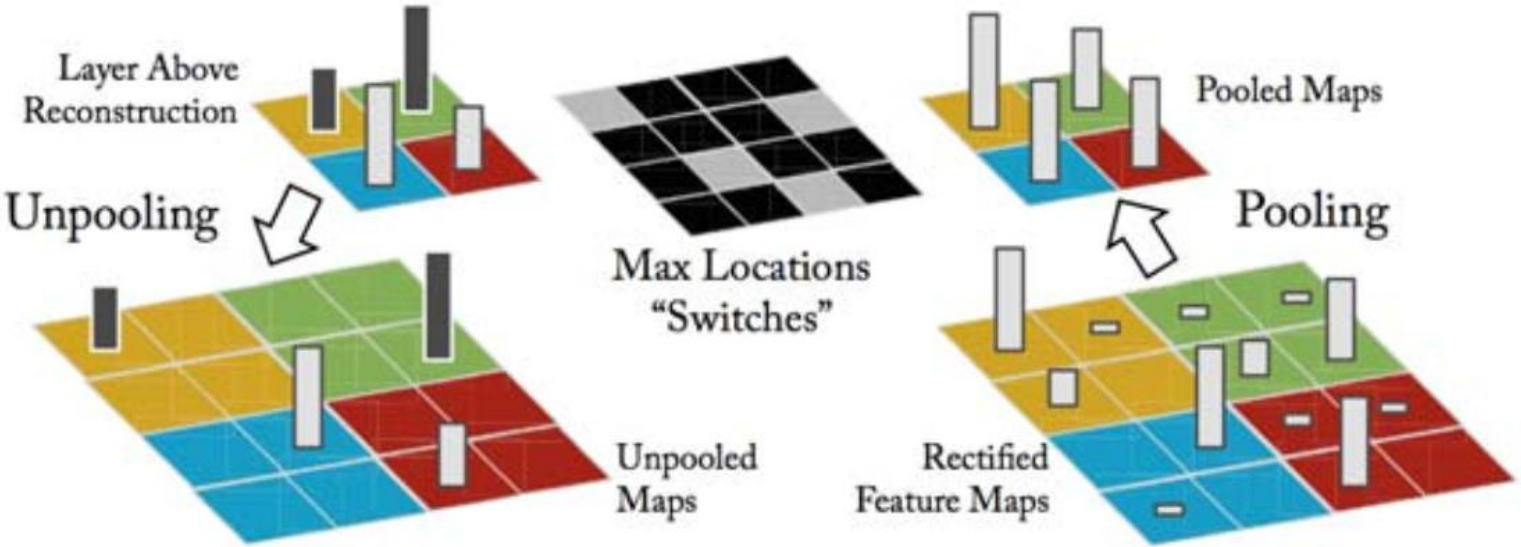


H. Noh et al., Learning deconvolution network for semantic segmentation. ICCV 2015.

发展历程

深蓝学院
shenlanxueyuan.com

Conv-Deconv Net中的下采样与上采样



H. Noh et al., Learning deconvolution network for semantic segmentation. ICCV 2015.

发展历程

深蓝学院
shenlanxueyuan.com

Conv-Deconv Net的重要意义

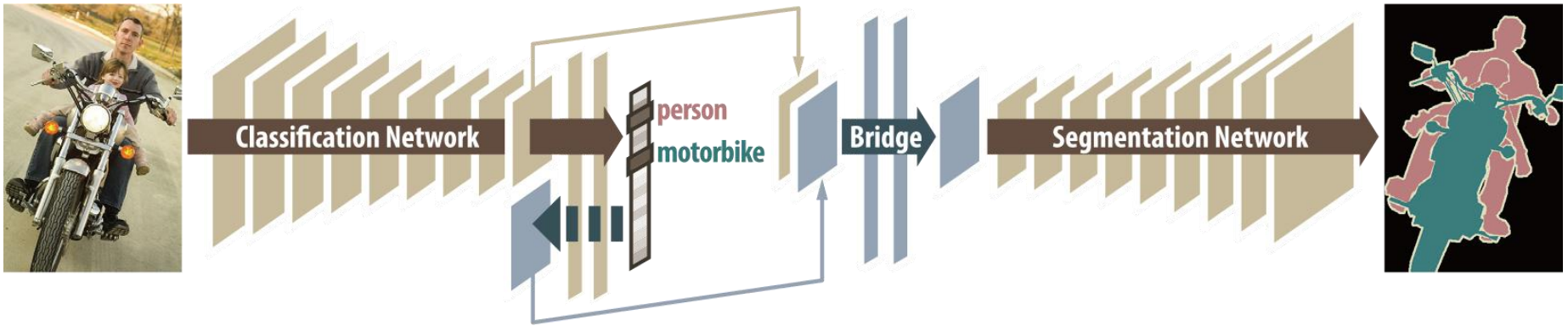
- 逐级的上采样，实现了以“由粗到精”的方式恢复细节信息。
- 弥补了全卷积网络的不足（一次性高倍数上采样）。
- 这种“沙漏形”的对称网络结构成为众多像素级标注问题的基础。

H. Noh et al., Learning deconvolution network for semantic segmentation. ICCV 2015.

发展历程

深蓝学院
shenlanxueyuan.com

Conv-Deconv形式的各种模型



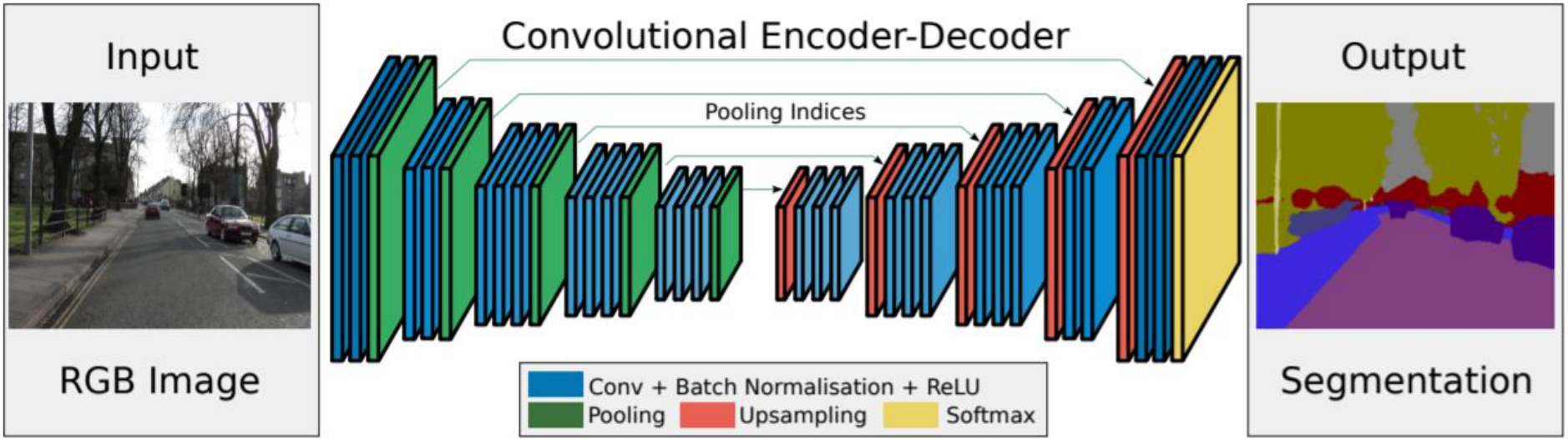
S. Hong et al., Decoupled Deep Neural Network for Semi-supervised Semantic Segmentation. NIPS 2015.

发展历程

深蓝学院 shenlanxueyuan.com

Conv-Deconv形式的各种模型（SegNet）

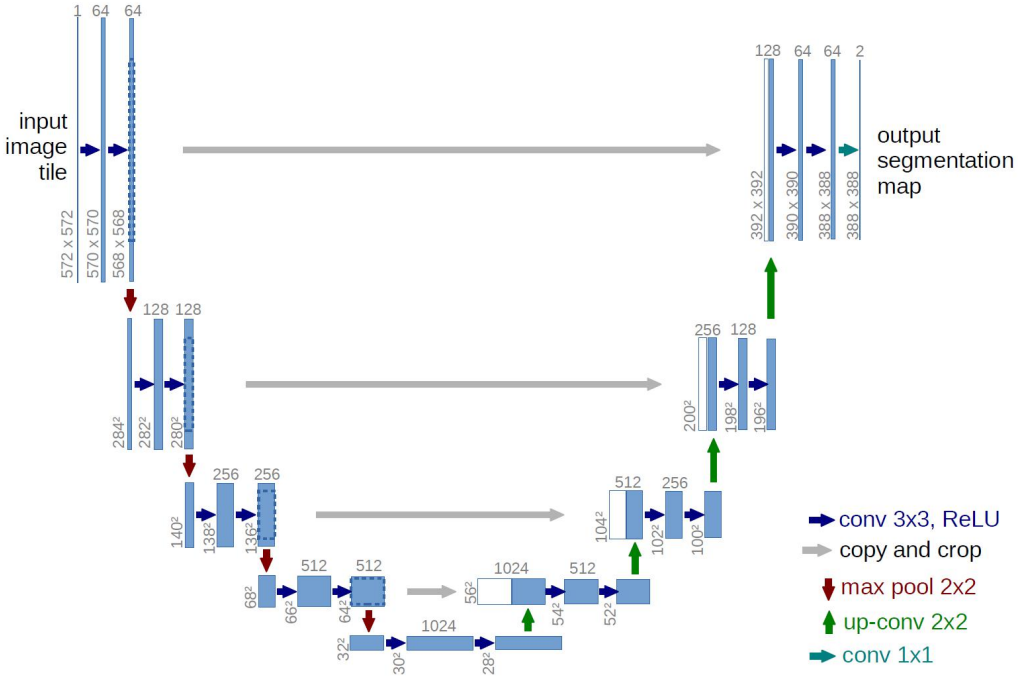
相比于Conv-Deconv模型，
最大区别在于去掉了
中间7*7的池化



V. Badrinarayanan et al., SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation, arXiv 2016.

发展历程

Conv-Deconv形式的各种模型（U-Net）



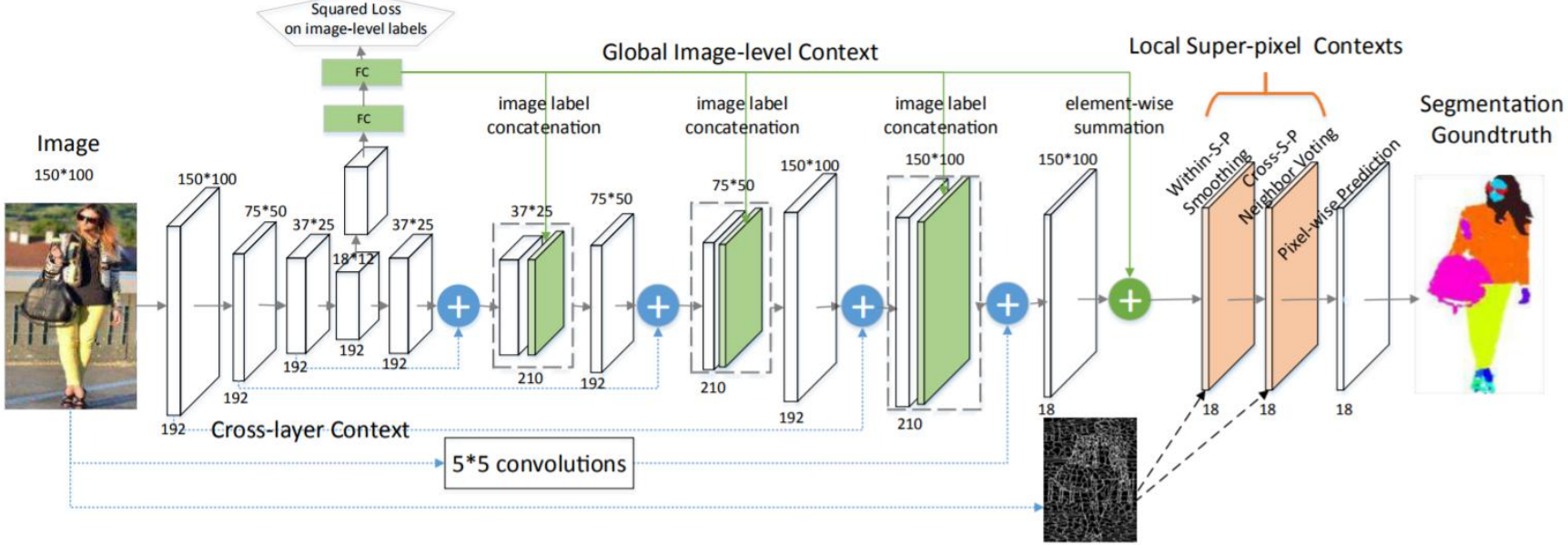
两阶段的feature map结合起来，更好地保留了细节信息

O. Ronneberger et al., U-Net: Convolutional Networks for Biomedical Image Segmentation. MICCAI 2015.

发展历程

深蓝学院 shenlanxueyuan.com

Conv-Deconv形式的各种模型

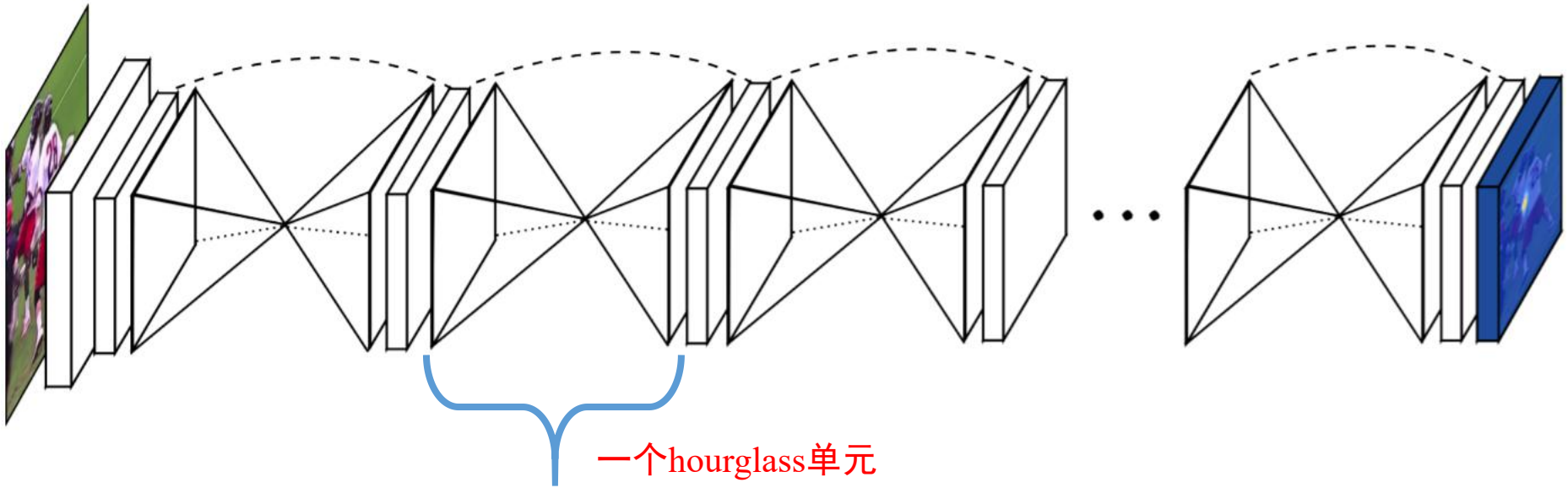


X. Liang et al., Human Parsing with Contextualized Convolutional Neural Network. ICCV 2015.

发展历程

深蓝学院 shenlanxueyuan.com

Conv-Deconv形式的各种模型



A. Newell et al., Stacked Hourglass Networks for Human Pose Estimation. arXiv 2016.

课程大纲

深蓝学院
shenlanxueyuan.com

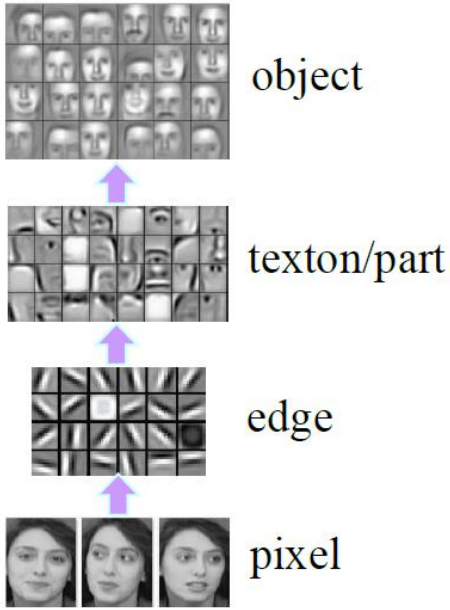
- ✓ 基本概念
- ✓ 发展历程
- ✓ **网络特点**

网络特点

深蓝学院 shenlanxueyuan.com

层次化的特征学习

- 直接从图像底层开始进行学习。随着网络层数的加深，能够学习到更高层的语义信息。
- 不必事先提取人为设计的特征，比如Gabor纹理特征、多尺度小波特征、 SIFT特征、 HOG特征等。
- 人为设计特征的缺点：这些特征具有一些参数，如尺度、梯度方向、频域划分等，其泛化能力不强。



思考：自动学习的特征优于人为设计的特征 ？

网络特点

深蓝学院 shenlanxueyuan.com

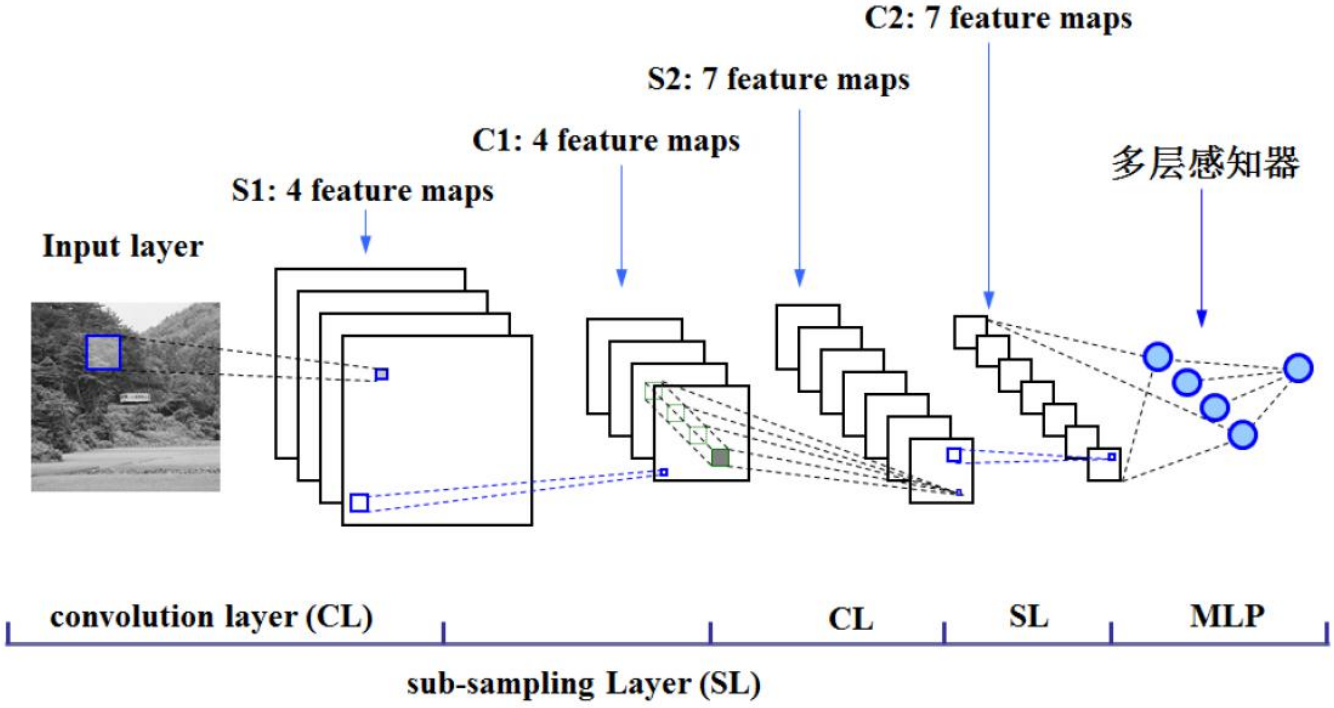
局部感受野（局部连接）

- 图像的空间相关性： 对图像而言，局部邻域内的像素联系较紧密，距离较远的像素相关性较弱。
- 每个神经元没必要对全局图像进行感知，只需要对局部区域进行感知。
- 在更高层将局部信息综合起来，以获得全局信息。
- 极大地降低了网络参数的数目。
- 增加网络层数能够保证节点具有更大范围的实际感受野（相对于输入图像而言）。

思考：实际感受野如何计算？

网络特点

局部感受野（局部连接）



思考：实际感受野如何计算？

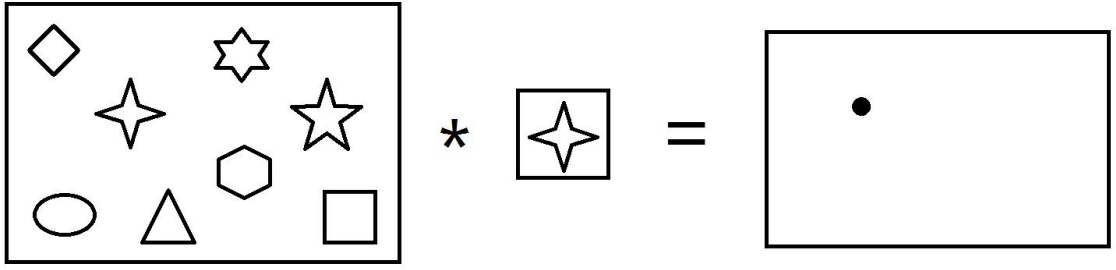
网络特点

深蓝学院 shenlanxueyuan.com

权值共享（卷积）

- 从图像任意局部区域内连接到同一类型的隐含结点，其权重保持不变。
- 实现：卷积（卷积核在图像或特征图上滑动）。
- 每个卷积核可以看成学习一种特征的滤波器。
- 也是一种降低网络参数数目的手段。

思考：【面试真题】卷积层和全连接层的参数数目及计算负担的对比。（网络配置信息见下页）



包含多种模式的合成图像　卷积核　　响应结果

每种卷积核负责寻找图像中的某种特定模式

网络特点

深蓝学院 shenlanxueyuan.com

池化（pooling）

- 降低特征图的尺寸，突出更强烈的响应。
- 使卷积神经网络具有平移不变性和畸变不变性。
- 每次进行下采样时，特征图数目相应增加

  => 牺牲空间细节信息，获得更多的语义信息。
- 增加高层特征图中节点的实际感受野。
- 也是一种降低网络参数数目的手段。

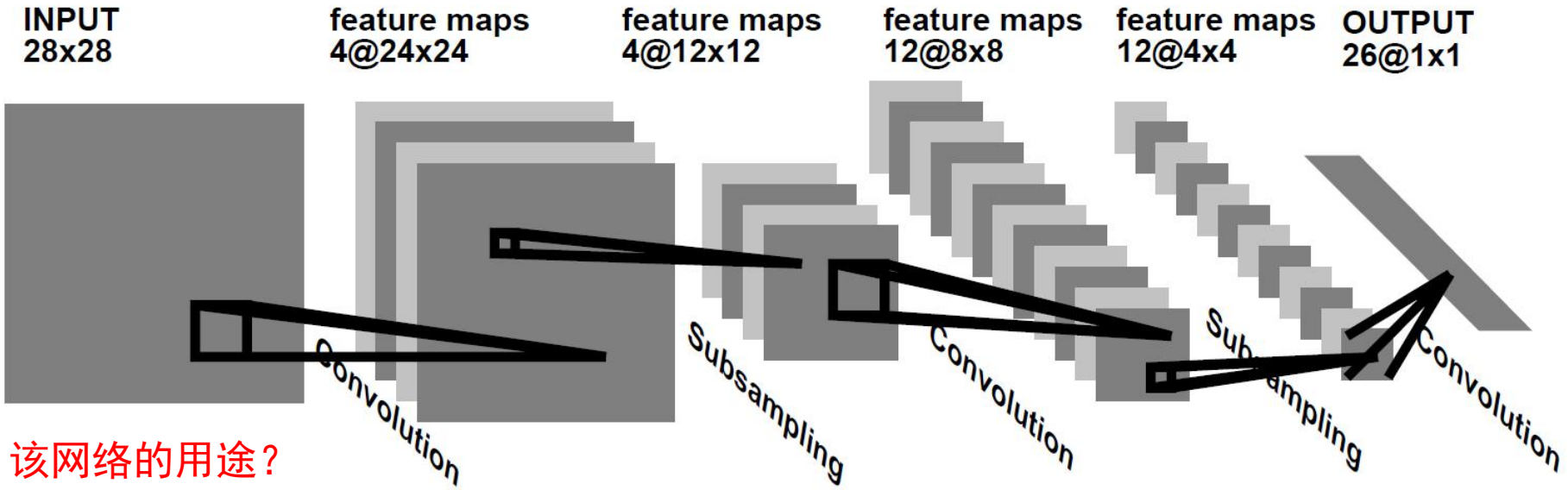
name	kernel size	output size
input	-	224 × 224 × 3
conv1-1	3 × 3	224 × 224 × 64
conv1-2	3 × 3	224 × 224 × 64
pool1	2 × 2	112 × 112 × 64
conv2-1	3 × 3	112 × 112 × 128
conv2-2	3 × 3	112 × 112 × 128
pool2	2 × 2	56 × 56 × 128
conv3-1	3 × 3	56 × 56 × 256
conv3-2	3 × 3	56 × 56 × 256
conv3-3	3 × 3	56 × 56 × 256
pool3	2 × 2	28 × 28 × 256
conv4-1	3 × 3	28 × 28 × 512
conv4-2	3 × 3	28 × 28 × 512
conv4-3	3 × 3	28 × 28 × 512
pool4	2 × 2	14 × 14 × 512
conv5-1	3 × 3	14 × 14 × 512
conv5-2	3 × 3	14 × 14 × 512
conv5-3	3 × 3	14 × 14 × 512
pool5	2 × 2	7 × 7 × 512
fc6	7 × 7	1 × 1 × 4096
fc7	1 × 1	1 × 1 × 4096

网络特点

深蓝学院
shenlanxueyuan.com

池化（pooling）与参数数目

思考：【面试真题】如果去掉池化层，将会带来什么后果？



该网络的用途？

网络特点

深蓝学院
shenlanxueyuan.com

池化（pooling）与平移不变性和畸变不变性

- 在图像识别和图像分类任务中，当图像中的某种特征被检测出（即该特征对卷积核具有较高响应），其具体位置便变得不再重要，只需要大致保持与其他特征的相对位置即可。
- 经过池化之后，能够更好地进行“非极大值抑制”（即抑制其他不明显特征响应的干扰），更加凸显该特征。
- 降低特征图的尺寸后，网络对平移和畸变的敏感性也相应降低。

思考：池化操作带来的平移不变性 vs. 卷积操作中卷积核滑动的“平移不变性”。

网络特点

深蓝学院 shenlanxueyuan.com

池化（pooling）与平移不变性和畸变不变性

- 有利于应对实际问题中各种不同类型、不同质量的图像。
- 通常还需专门对输入图像进行平移、畸变等处理来提高网络的泛化性能 — 数据增广

关于卷积神经网络的平移不变性和畸变不变性的具体介绍可以参考技术报告
Y. LeCun and Y. Bengio, Convolutional Networks for Images, Speech and Time-Series.
以及知乎回答https://www.zhihu.com/question/36980971/answer/94840350

参考资料

1. Y. LeCun et al., Backpropagation applied to handwritten zip code recognition. Neural Computation 1989.

2. Y. LeCun et al., Gradient-Based Learning Applied to Document Recognition. Proceedings of the IEEE 1998.

3. A. Krizhevsk et al., ImageNet Classification with Deep Convolutional Neural Networks. NIPS 2012.

4. J. Long et al., Fully convolutional networks for semantic segmentation. CVPR 2015.

5. Y. LeCun and Y. Bengio, Convolutional Networks for Images, Speech and Time-Series. Technical report.

6. Y. Bengio et al, Deep learning. https://github.com/HFTrader/DeepLearningBook .

在线问答

深蓝学院
shenlanxueyuan.com

Q&A

深蓝学院
shenlanxueyuan.com

深度学习理论
课程地址

感谢各位聆听！

Thanks for Listening