



第六周 计算机视觉

李泽椿，复旦大学生物医学工程与技术创新学院



目录

1

计算机视觉简介

2

卷积神经网络

3

视觉模型演化历程

4

计算机视觉应用场景



计算机视觉 | 主要内容



分类任务有哪些

流程 ⊙ 6. 机器学习

性能指标

最近邻、Lp距离 ⊙ 7. 图像分类

神经网络与卷积神经网络 ⊙ 8. 深度学习

9. 目标检测、图像分割、实例分割、全景分割

10. 人脸检测、人脸对齐

11. 人脸识别、人脸验证、属性估计、人脸分割

生成对抗网络 ⊙ 12. 深度学习

13. 图像风格迁移、图像生成、图像补全

目标跟踪、行为识别、人体跟踪、行人重识别 ⊙ 14. 视频

15. 3D建模

文本检测、文本识别OCR、车牌检测、车牌识别

应用

公共安全、无人车、机器人、医疗

计算机视觉

高层视觉

图像生成

1. 摄像机几何模型

2. 颜色

早期视觉

3. 滤波与边缘

线性滤波

卷积：计算

边缘效应

平滑滤波

锐化滤波

边缘

卷积微分

Canny边缘检测

角点

Harris角点检测

4. 底层特征

SIFT

多尺度表征

尺度不变特征变换 SIFT

HOG

梯度方向直方图

5. 纹理

| 主要内容 |

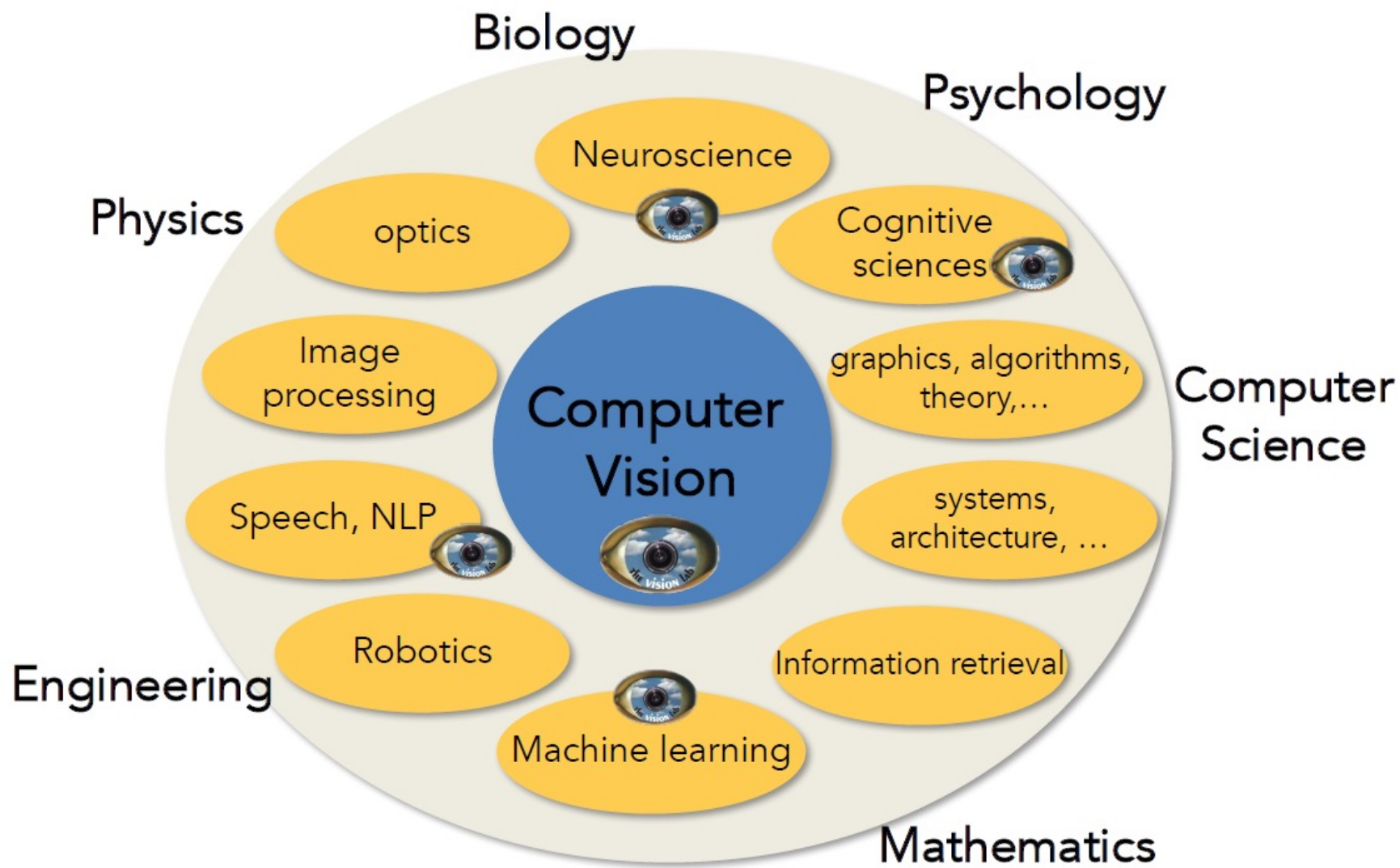
第一部分 初级视觉	计算机视觉面临的挑战 图像与图像处理 图像滤波和形态学 网络的作用 图像分割
第二部分 中级视觉	二值化形状分析 数学形态分析 直线、圆和椭圆的检测 广义霍夫变换 物体分割与形状模型
第三部分 机器学习和深度学习网络	基本分类概念 机器学习：概率方法 深度学习网络
第四部分 三维视觉和运动	三维世界 解决非正交性问题 不变性与透视 图像变换和摄像机校准 运动
第五部分 计算机视觉的应用	人脸识别与识别：深度学习带来的影响 监控 车载视觉系统 结语——计算机视觉展望

计算机视觉：原理、算法、应用及学习



● 应用领域

- 安防监控
- 图像搜索
- 工业视觉
- 人机交互
- 虚拟现实
- 生物医学
- 遥感测绘
- ...





计算机视觉

(Computer Vision, CV)

核心任务

图像分类
(Image Classification)

物体检测
(Object Recognition)

语义分割
(Semantic Segmentation)

视频分析
(Video Analysis)

其他任务



人体姿态识别



目标跟踪



SLAM



OCR

→ .txt 边缘检测
细粒度识别

.....

落地应用



人脸识别



自动驾驶

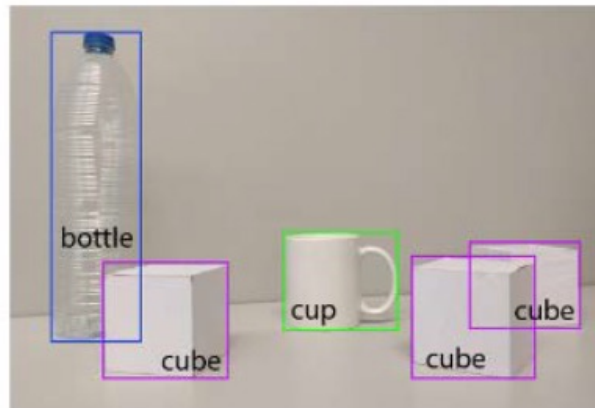


医疗影像
辅助诊断

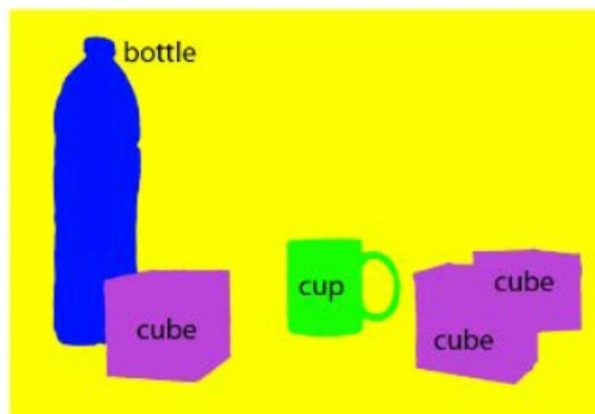




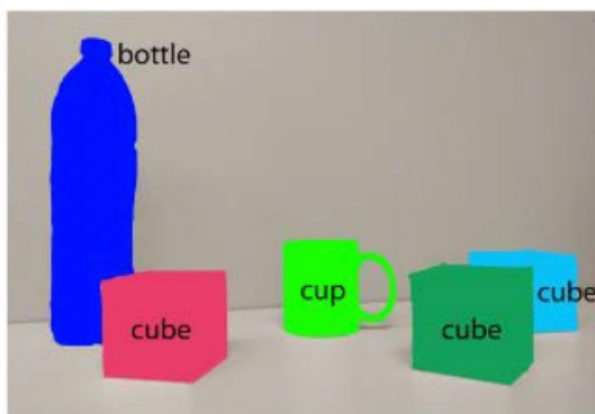
(a) Image classification



(b) Object localization



(c) Semantic segmentation



(d) Instance segmentation

左图展示了四种主要的计算机视觉任务（由粗粒度到精细粒度）

1、图像分类 (what)

为图像赋予一个或多个语义标签

2、目标检测 (what & where)

找到图像中物体的类别及所在位置

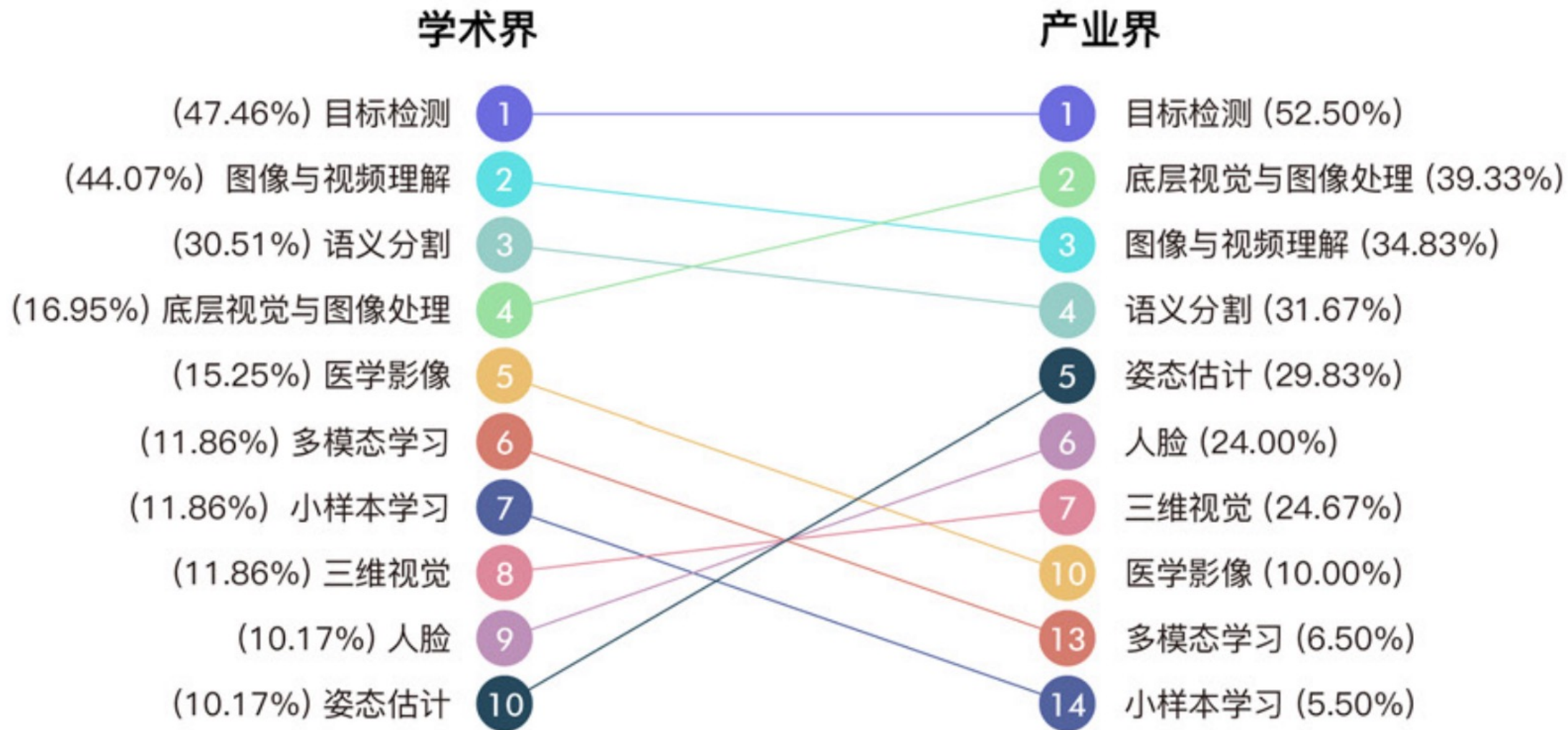
3、图像语义分割 (what & where)

确定图像中物体的类别并精确勾勒出其所在位置

4、图像实例分割 (what & where)

多个同类物体存在时，将它们一一区分出来

计算机视觉 | 研究领域对比



目录

1

计算机视觉简介

2

卷积神经网络

3

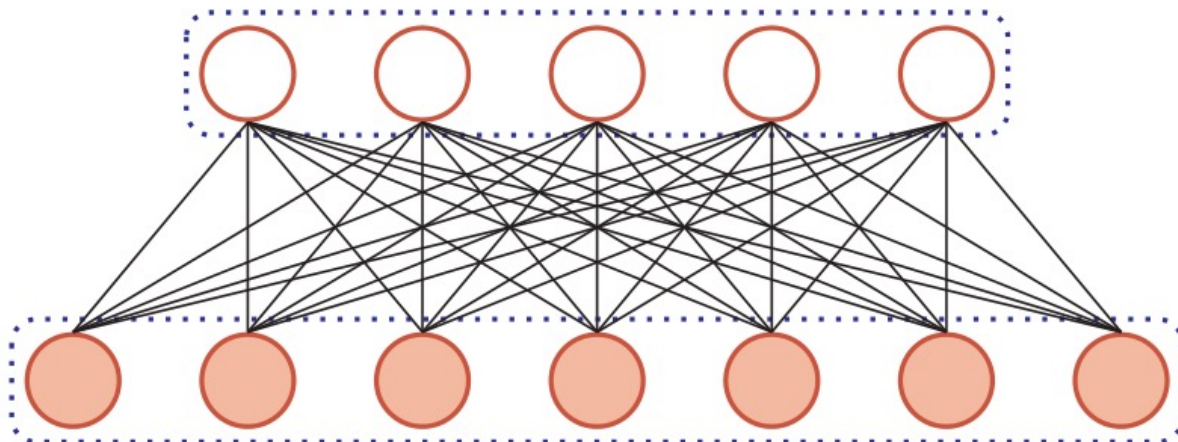
视觉模型演化历程

4

计算机视觉应用场景



- 权重矩阵的参数非常多



- 局部不变性特征

- 自然图像中的物体都具有局部不变性特征，比如尺度缩放、平移、旋转等操作不影响其语义信息。
- 而全连接前馈网络很难提取这些局部不变特征。

计算机视觉 | 卷积神经网络 (CNN)

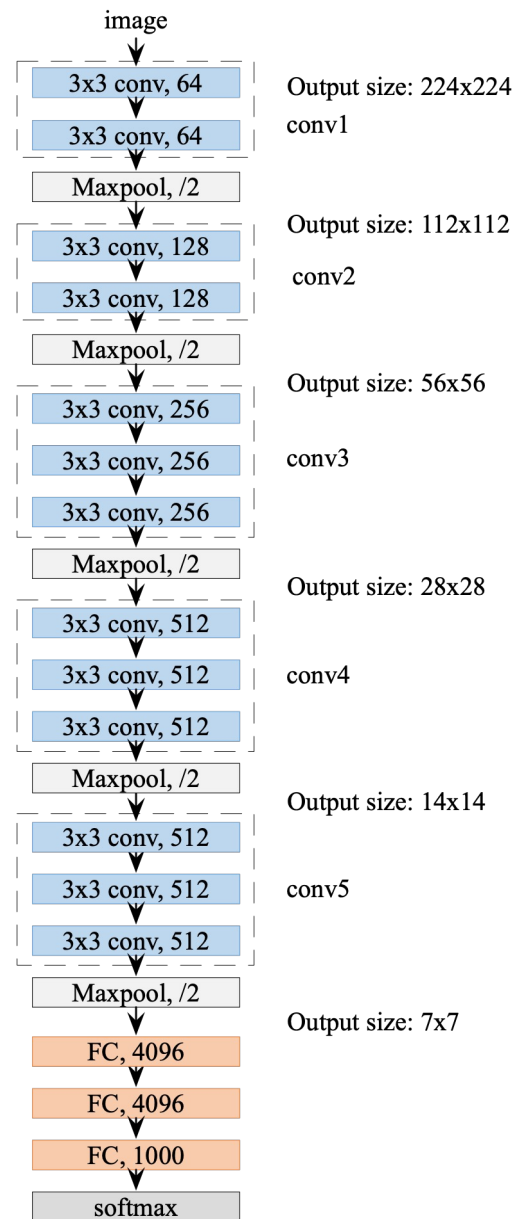


	全连接	卷积
局部连接		
权重共享	所有神经元之间的连接都使用不同权重。	输出层神经元共用同一组权重，进一步减少权重数量。
权重数量	$w_i \times h_i \times w_o \times h_o$	$f \times f$



• VGG16

- “ 卷积层 (conv)
- “ 池化层 (max pool)
- “ 全连接层 (FC)
- “ Softmax



• 卷积层如何检测特征

- 检测复杂边缘
- 将权重作为参数，在训练中学习。



w0	w1	w2
w3	w4	w5
w6	w7	w8

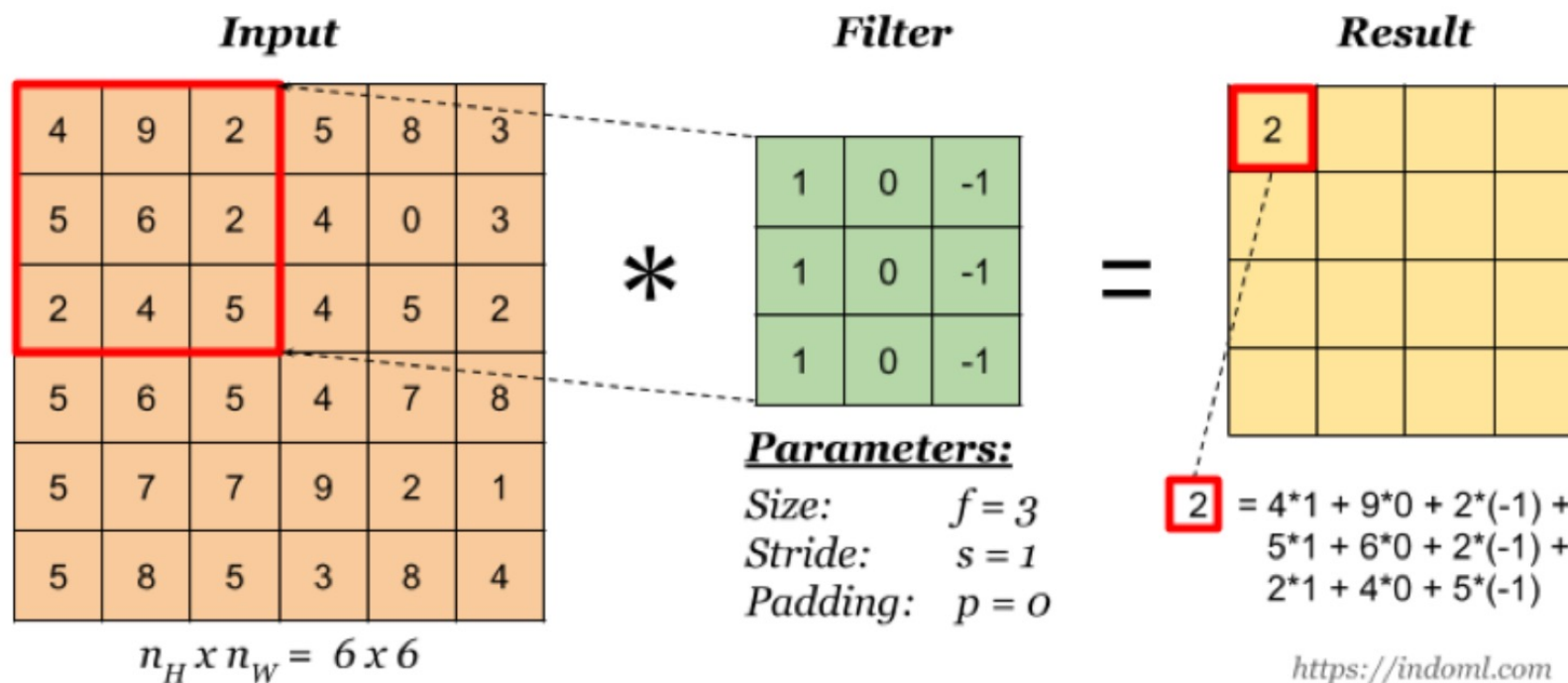
filter/kernel

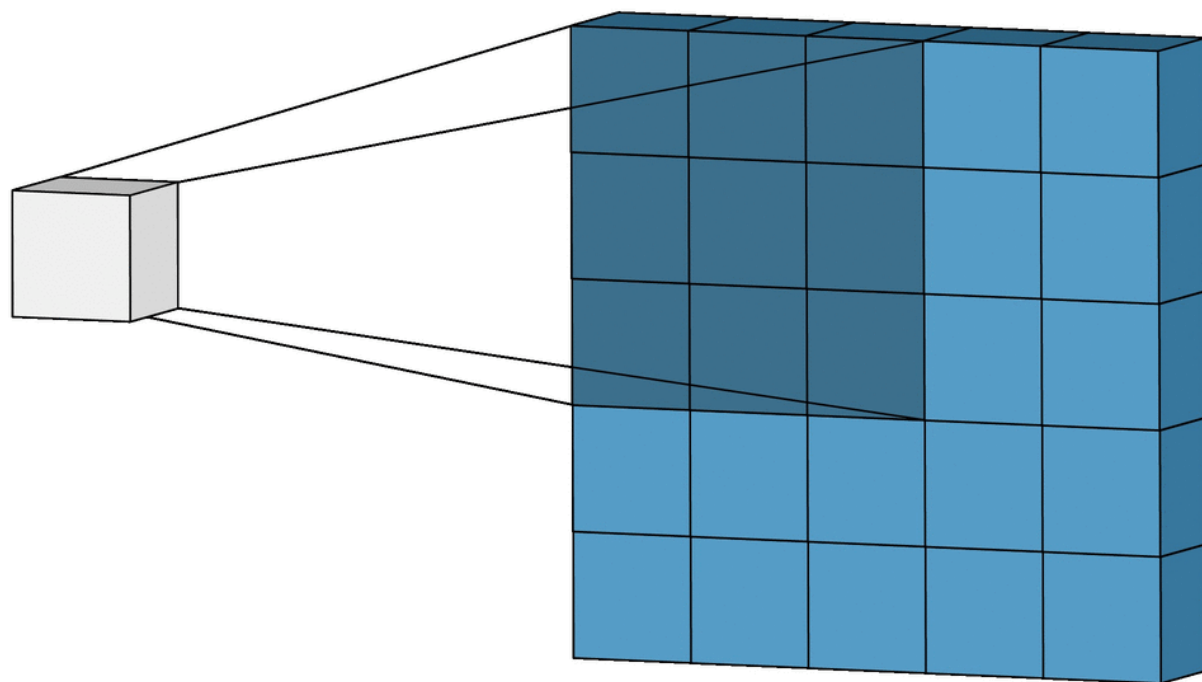
- 卷积神经网络的两个重要特征：**局部连接、权重共享**
可有效减少权重参数，避免过拟合，为增加卷积层数提供可能。



- 基本操作单元：卷积层

卷积核 (或称滤波器, filter/kernel)





- 移动的阴影：卷积核 (convolution kernel)
- 阴影大小：3x3的矩阵
- 移动顺序：左上角 -> 右下角
- 每次移动步长 (stride)：1个小格
- 卷积操作：对位相乘后加和

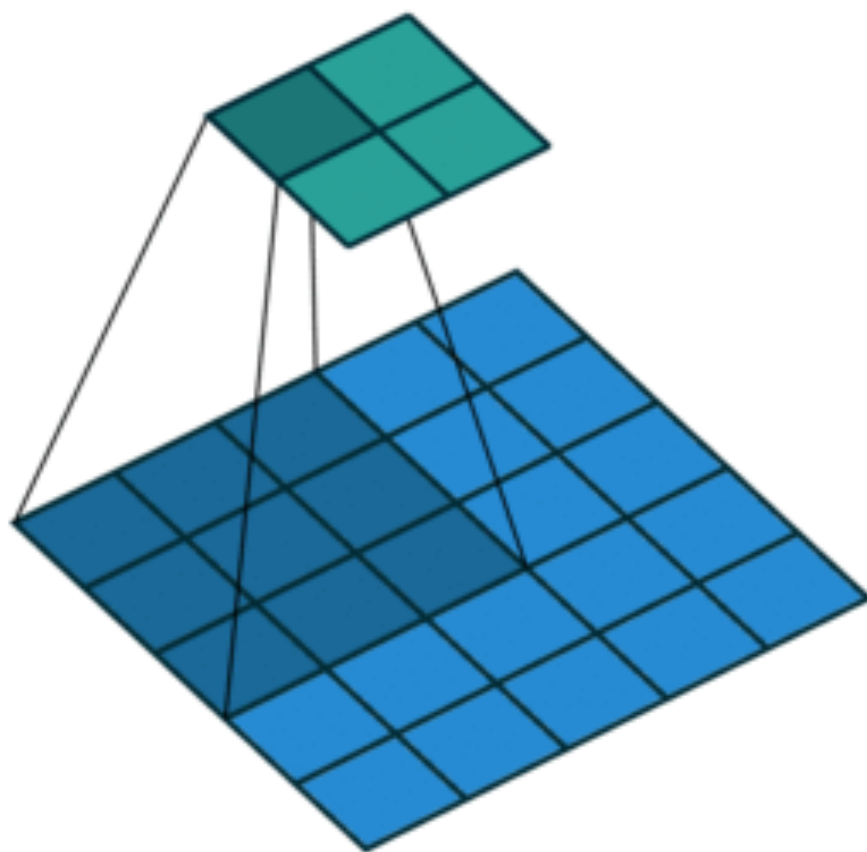




3_0	3_1	2_2	1	0
0_2	0_2	1_0	3	1
3_0	1_1	2_2	2	3
2	0	0	2	2
2	0	0	0	1

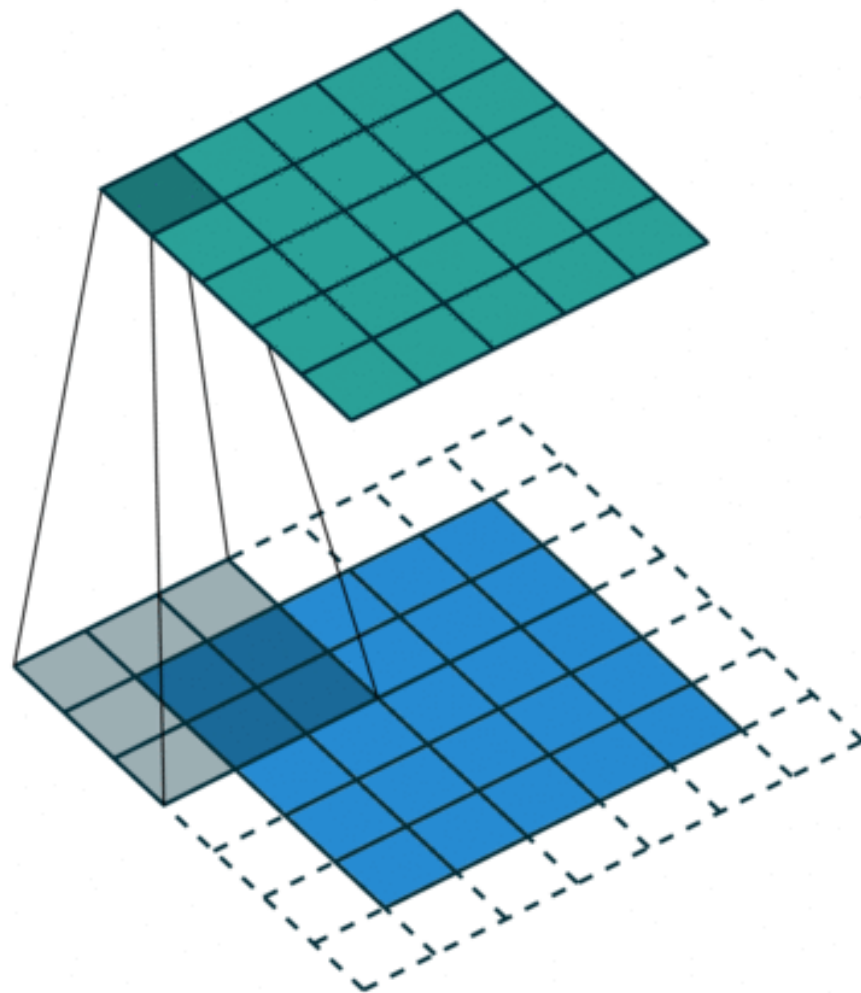
12.0	12.0	17.0
10.0	17.0	19.0
9.0	6.0	14.0





- 卷积操作具有降维效果
- 卷积后的大小与卷积核大小和步长 (stride) 有关



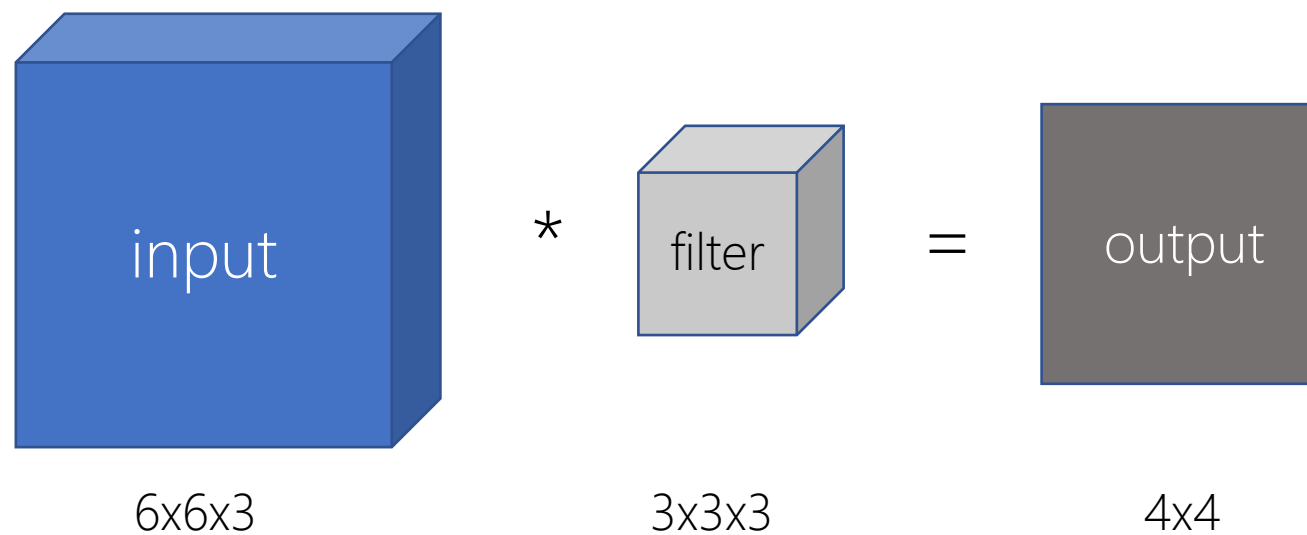


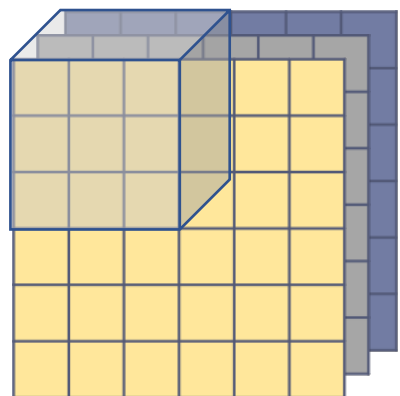
□ Padding: 使卷积后的大小跟原图大小一致 (要保持维度不变)

即在原图周围添加空白格 (same padding)



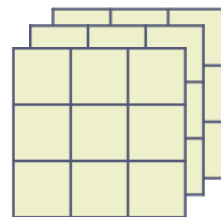
- 多输入特征图单输出特征图卷积运算





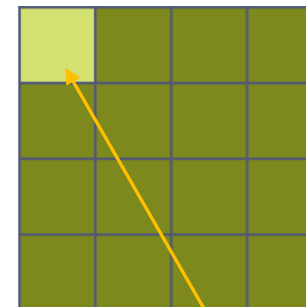
6x6x3

*



3x3x3

=



4x4

C = 0

0	0	0
0	2	2
0	1	2

C = 1

0	0	0
0	0	2
0	1	2

C = 2

0	0	0
0	1	1
0	0	2

*

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

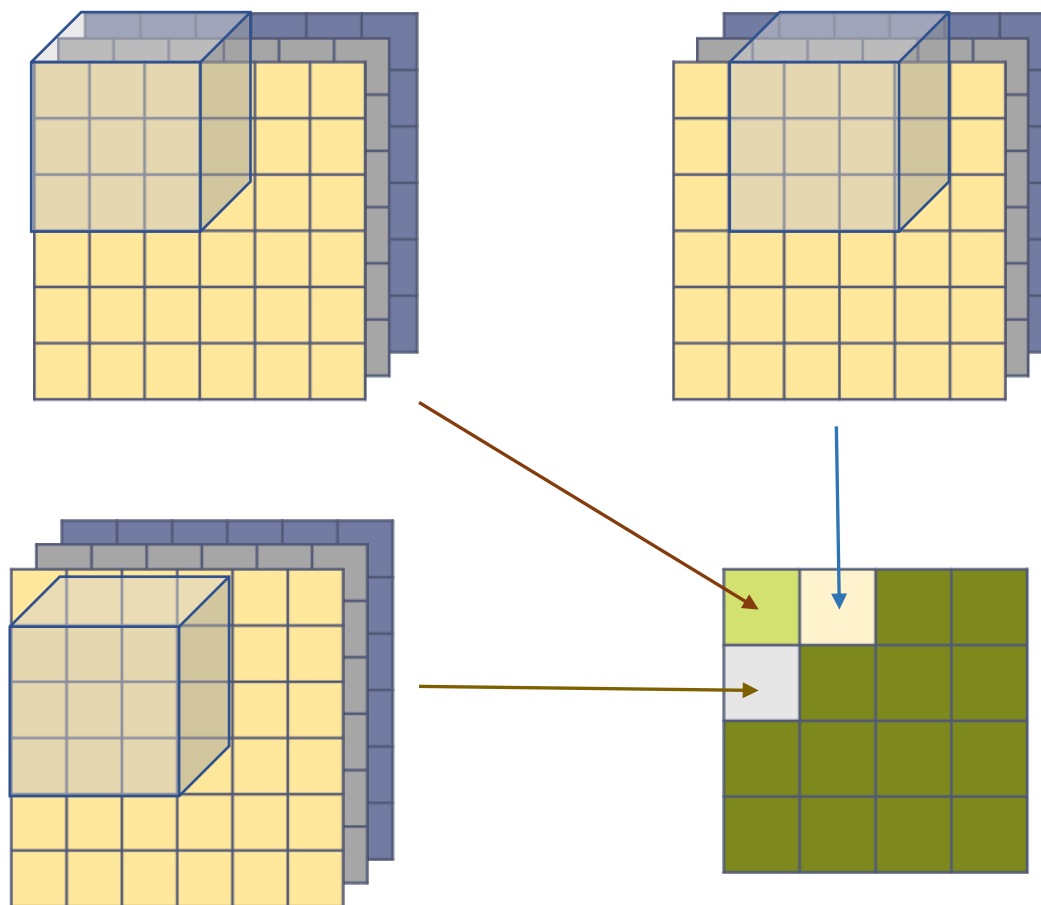
1	-1	-1
-1	0	-1
-1	0	1

1	-1	-1
-1	-1	0
-1	1	1

=

$$\begin{aligned}
 & 2-2-1+2 \\
 & \quad + \\
 & 0-2+0+2 \\
 & \quad + \\
 & (-1)+0+0+2 \\
 & = 2
 \end{aligned}$$

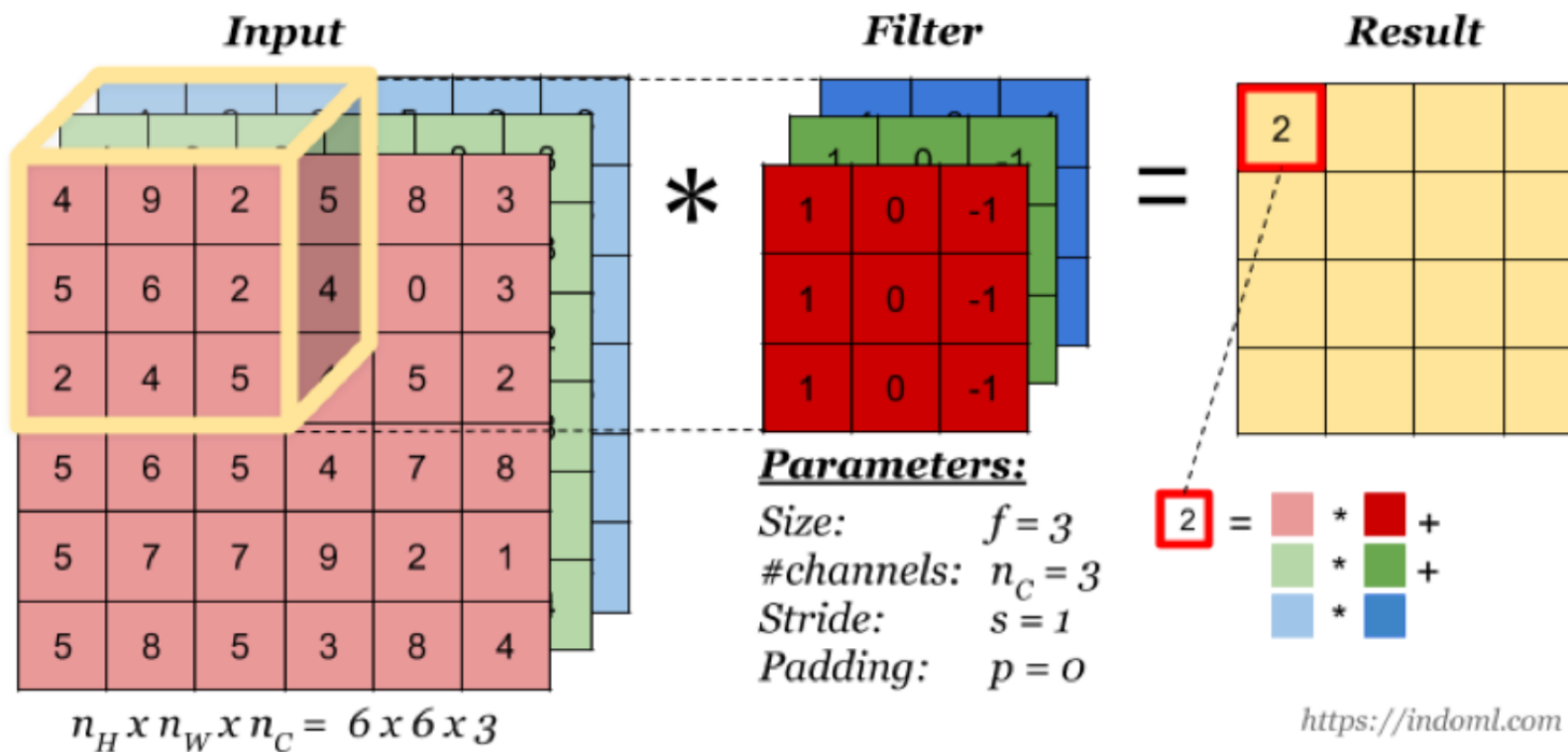






- 基本操作单元：卷积层

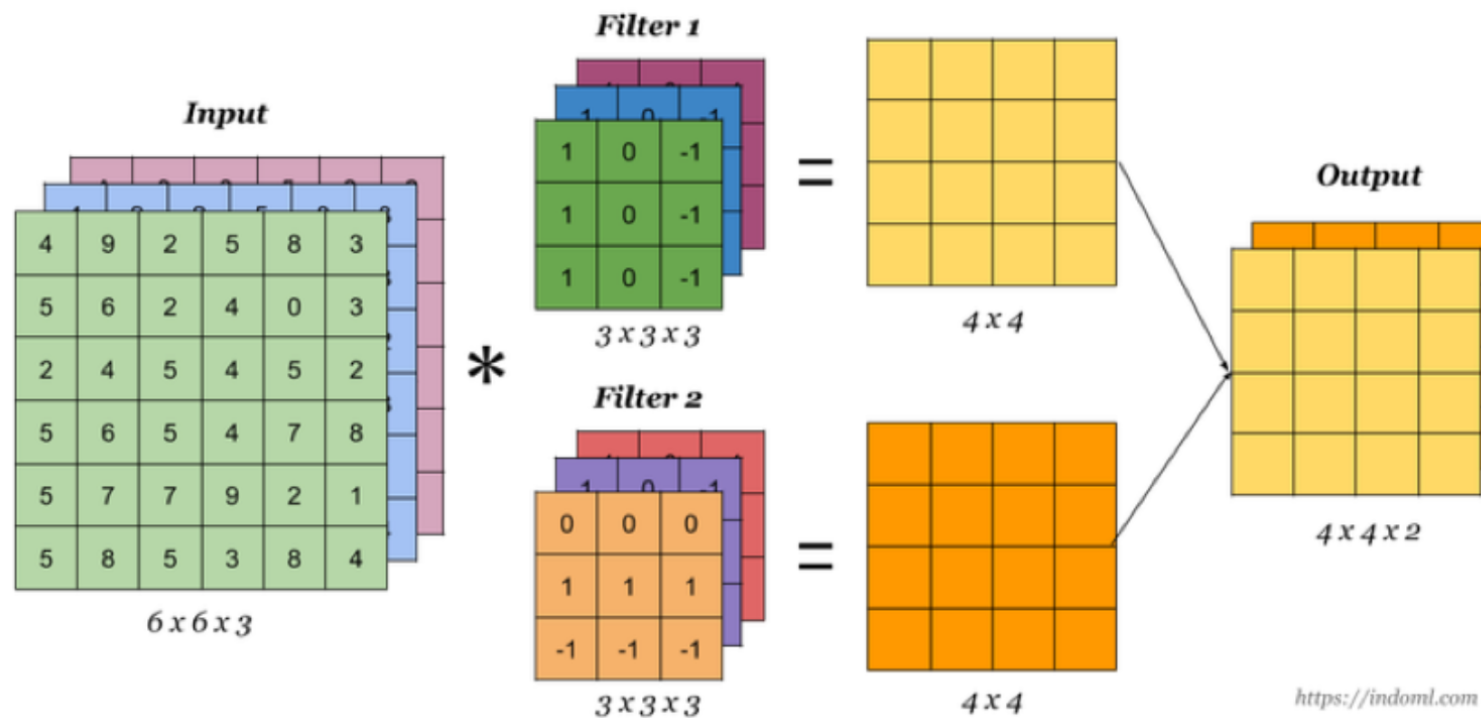
多通道卷积





• 基本操作单元：卷积层

多卷积核

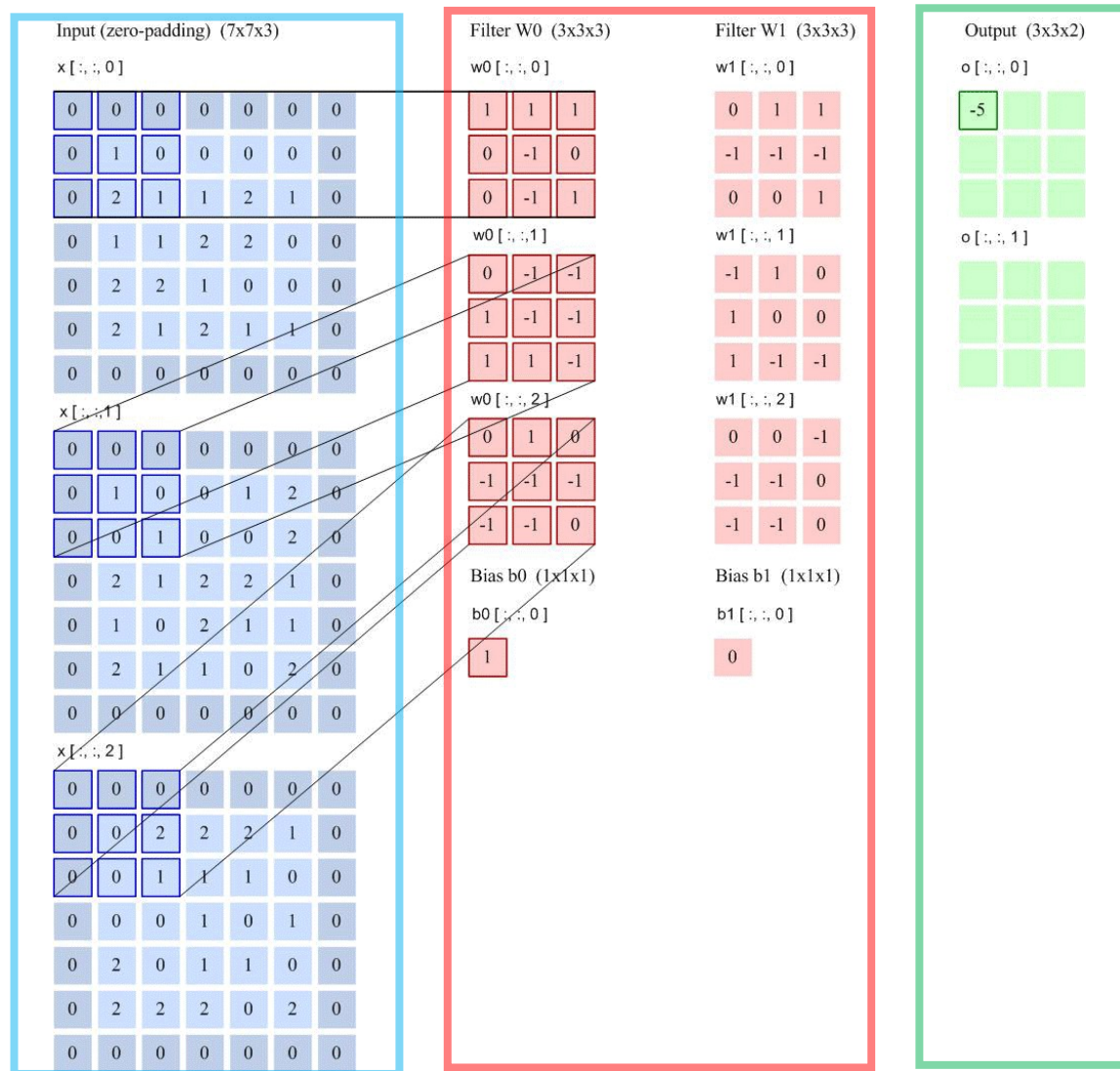


- $size = 3$
- $c_{in} = 3$
- $c_{out} = 2$
- $stride = 1$
- $padding = 0$

<https://indoml.com>



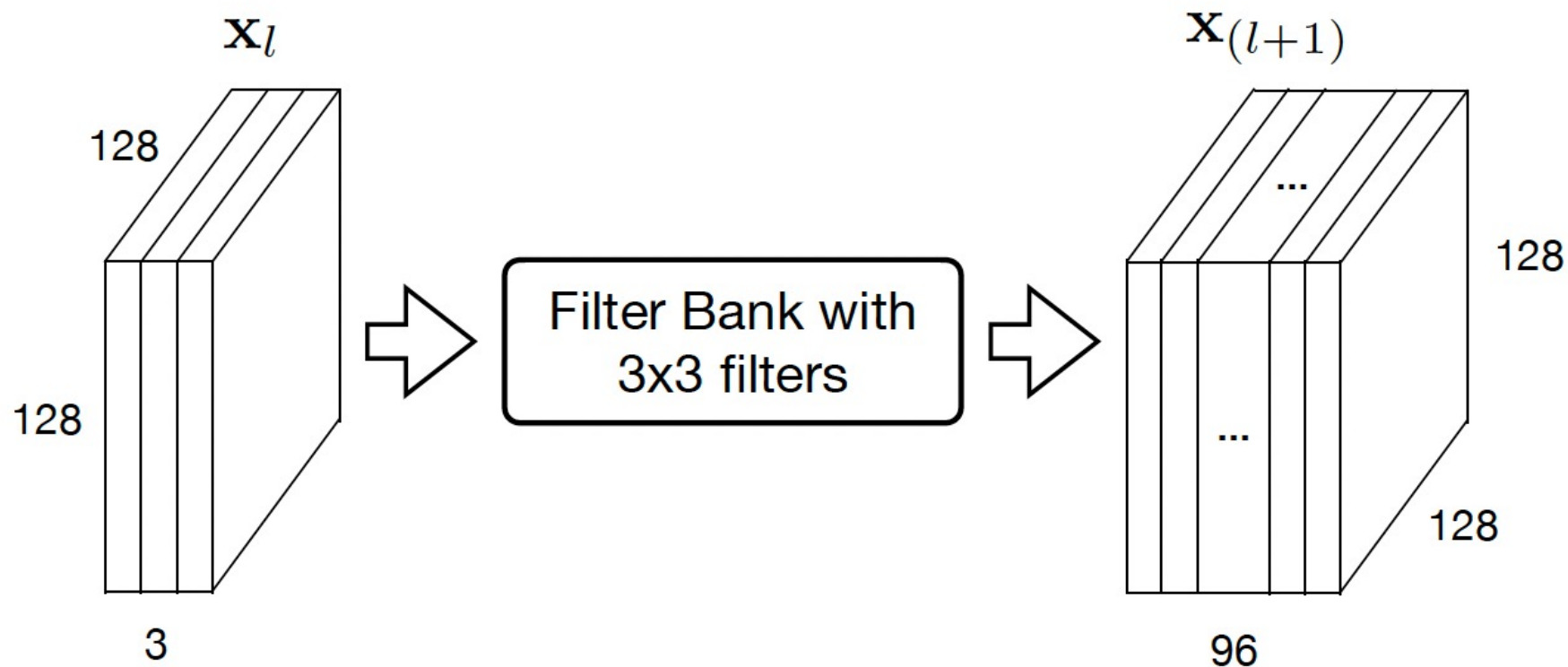
计算机视觉 | 步长2 filter个数2 3*3 填充





	Batch		Height		Width		In Depth		Out Depth
Input	256	×	227	×	227	×	3		
Kernel			11	×	11	×	3	×	96
Output	256	×	55	×	55			×	96

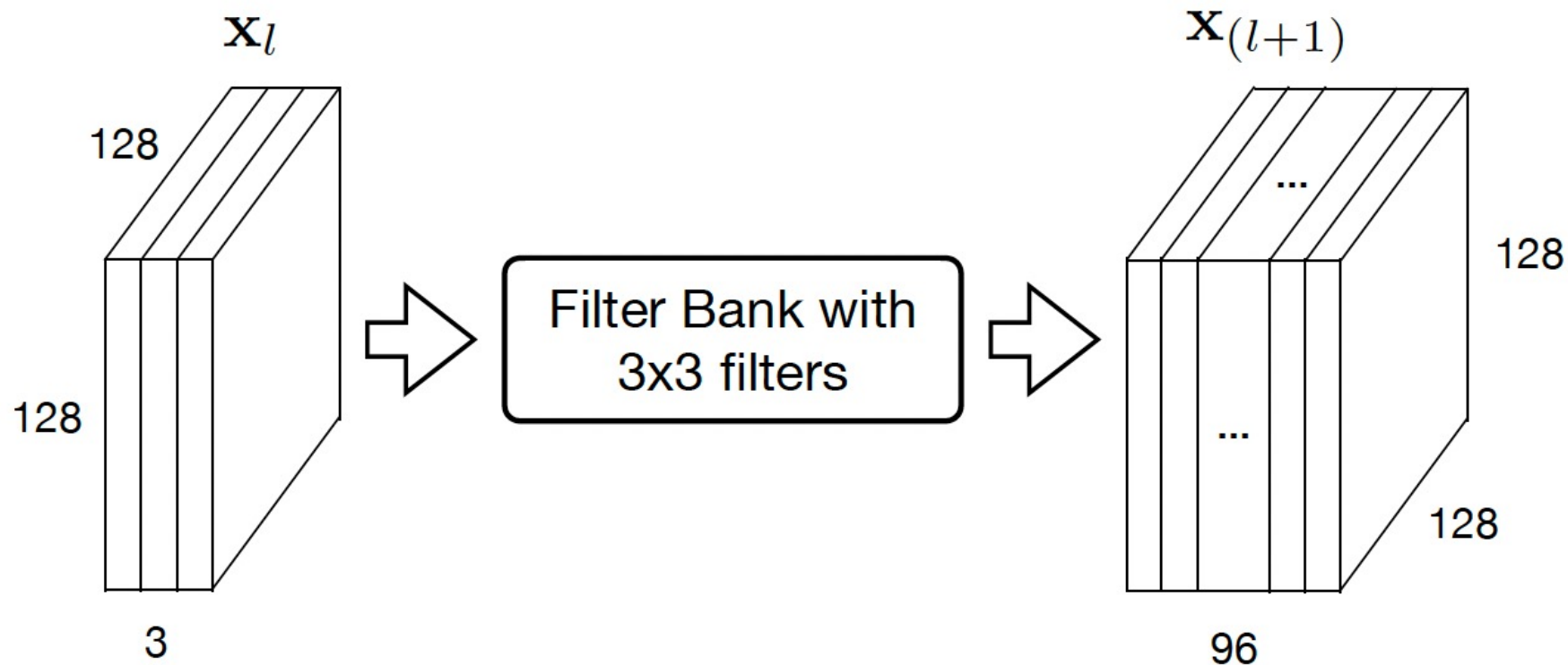




How many filters are in the bank?

- (a) 3 (b) 27 (c) 96 (d) can't say

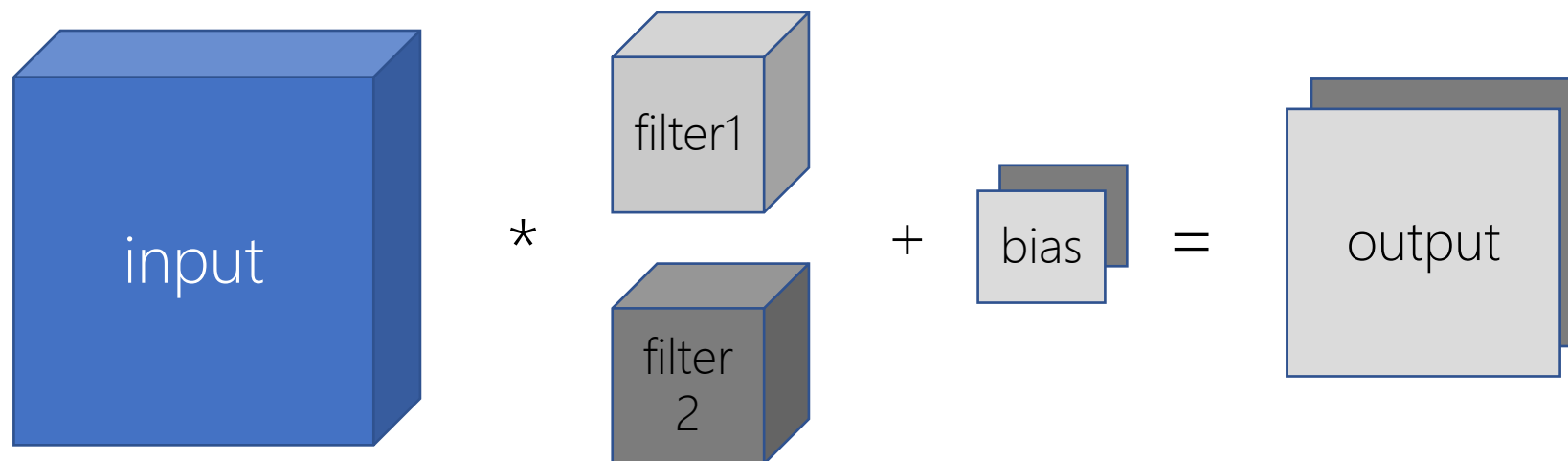




How many filters are in the bank?

- (a) 3 (b) 27 (c) 96 (d) can't say

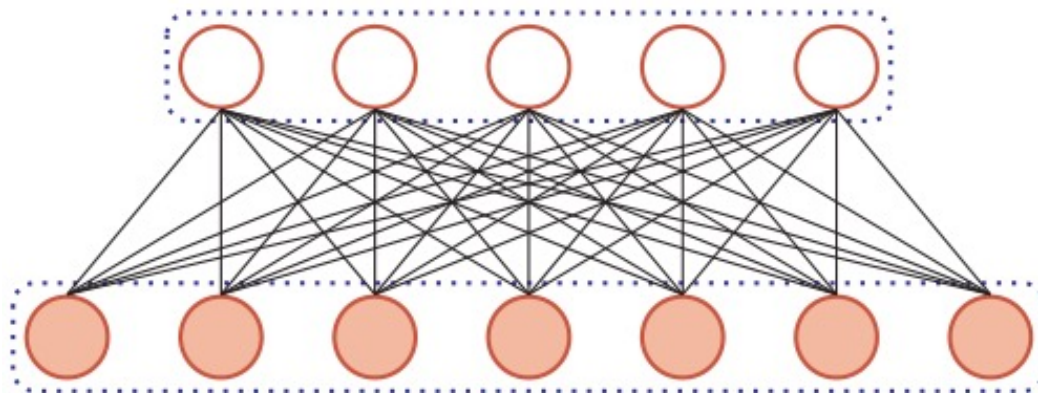




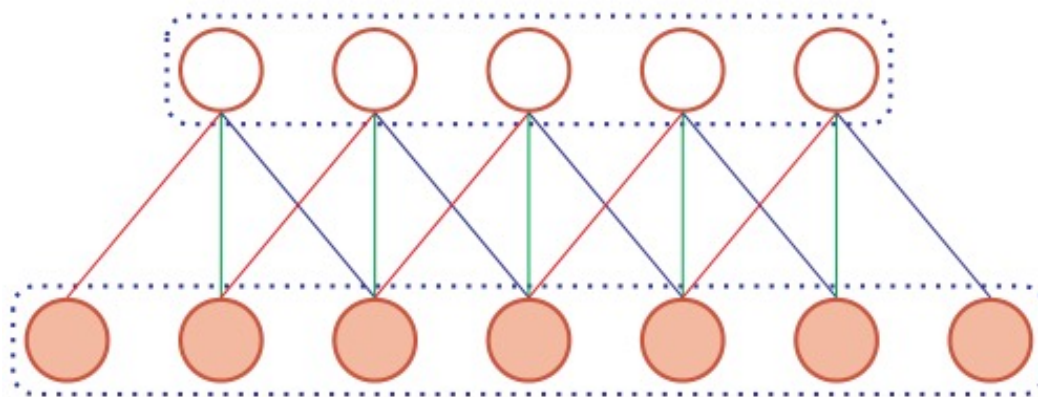
- filter: 可训练
- bias: 可训练, 使分类器偏离激活函数原点, 更灵活;
- activation



- 卷积层相较于全连接层的参数优势



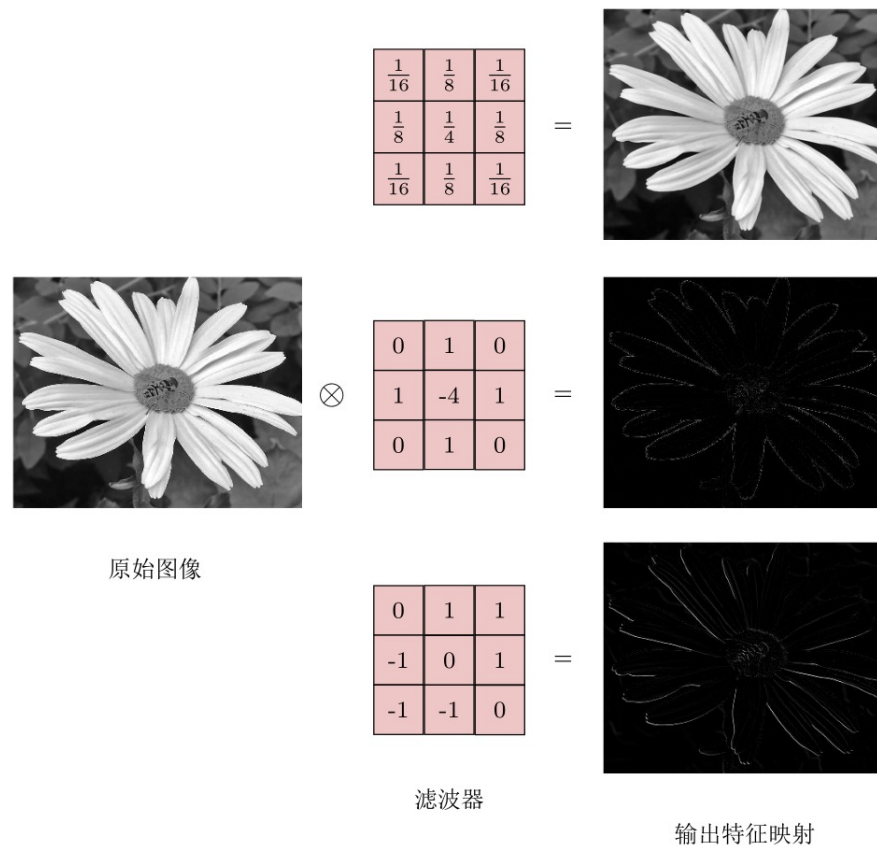
(a) 全连接层



(b) 卷积层

- 特征映射 (Feature Map) : 图像经过卷积后得到的特征。

- 卷积核看成一个特征提取器



- 卷积层通过卷积核的参数分布来提取对应特征

- 检测垂直边缘

10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0
10	10	10	0	0	0

 *

1	0	-1
1	0	-1
1	0	-1

 =

0	30	30	0
0	30	30	0
0	30	30	0
0	30	30	0

- 检测对角线边缘

10	10	10	10	10	0
10	10	10	10	0	0
10	10	10	0	0	0
10	10	0	0	0	0
10	0	0	0	0	0
0	0	0	0	0	0

 *

1	1	0
1	0	-1
0	-1	-1

 =

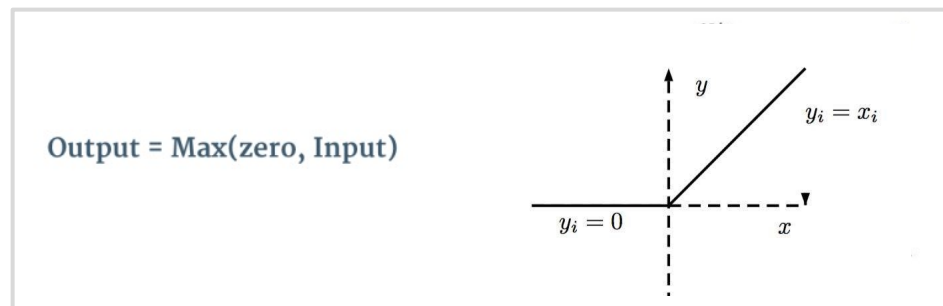
0	10	30	30
10	30	30	10
30	30	10	0
30	10	0	0

- 卷积层扫描一遍图像得到的输出Feature Map

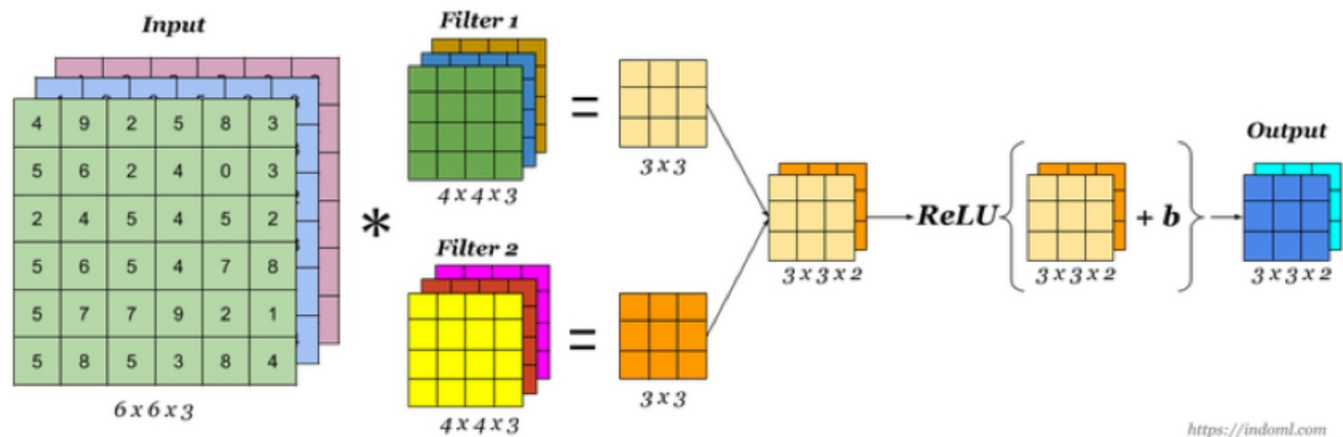


Input

- 基本操作单元：激活层



A Convolution Layer



<https://indoml.com>

- 基本操作单元：激活层

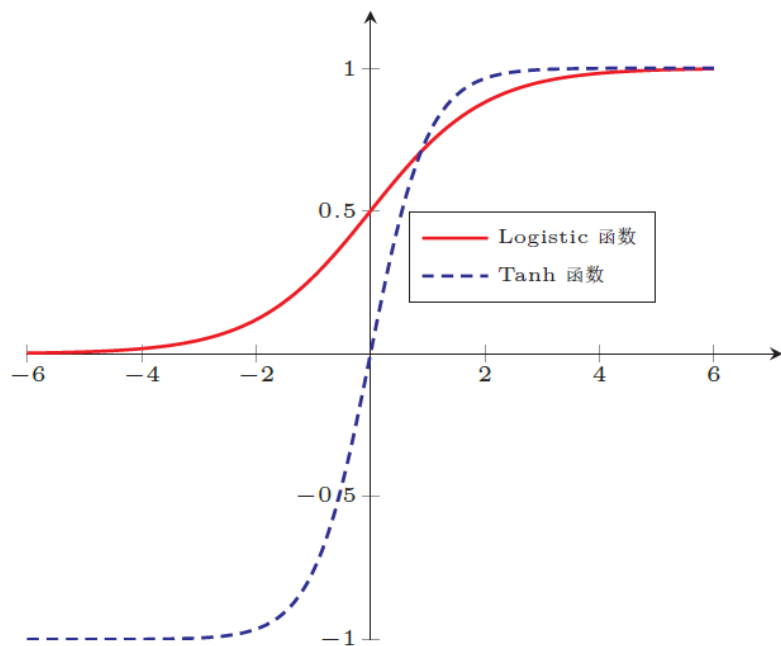


图 4.2 Logistic 函数和 Tanh 函数

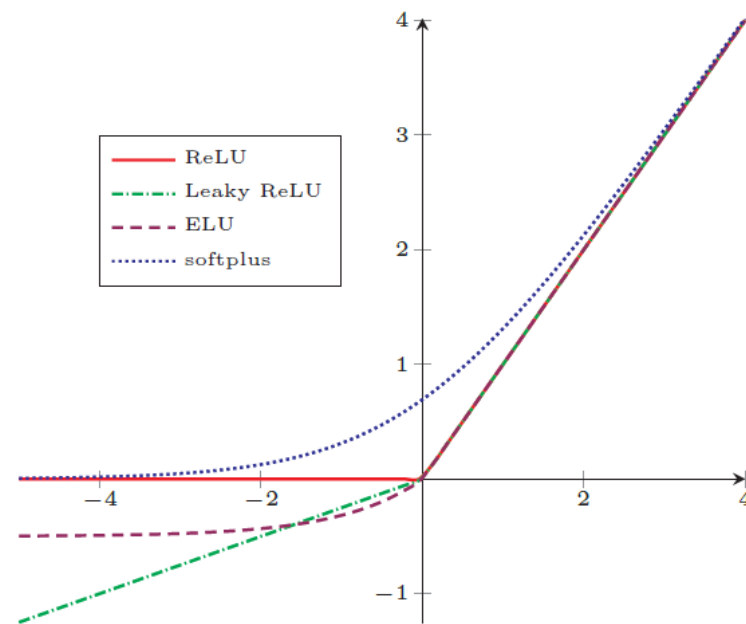
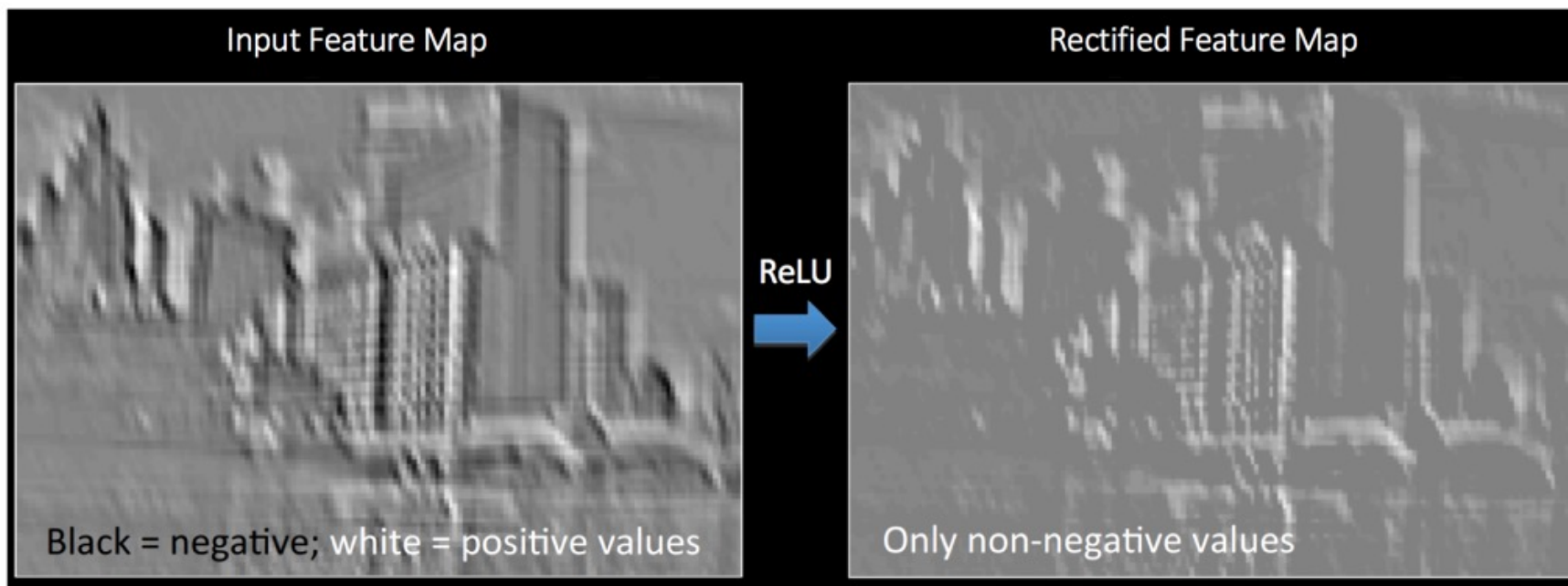


图 4.4 ReLU、Leaky ReLU、ELU 以及 Softplus 函数

- 基本操作单元：激活层





Max Pooling

29	15	28	184
0	100	70	38
12	12	7	2
12	12	45	6

2 x 2
pool size

100	184
12	45

Average Pooling

31	15	28	184
0	100	70	38
12	12	7	2
12	12	45	6

2 x 2
pool size

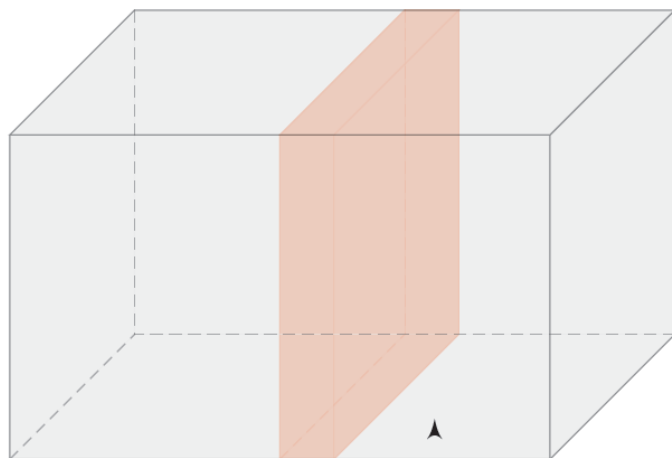
36	80
12	15

Down sampling effect





池化 Pooling / 降采样 层

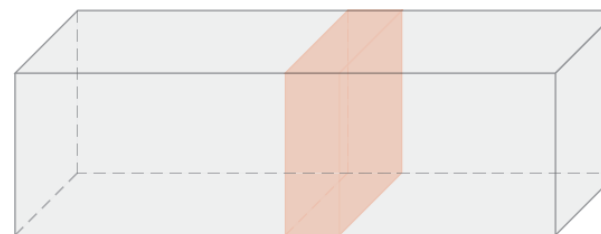


输入特征映射组 \mathbf{X}

1	1	2	0
0	1	1	2
0	0	1	3
0	0	1	1

输入特征映射 X^d

max pooling



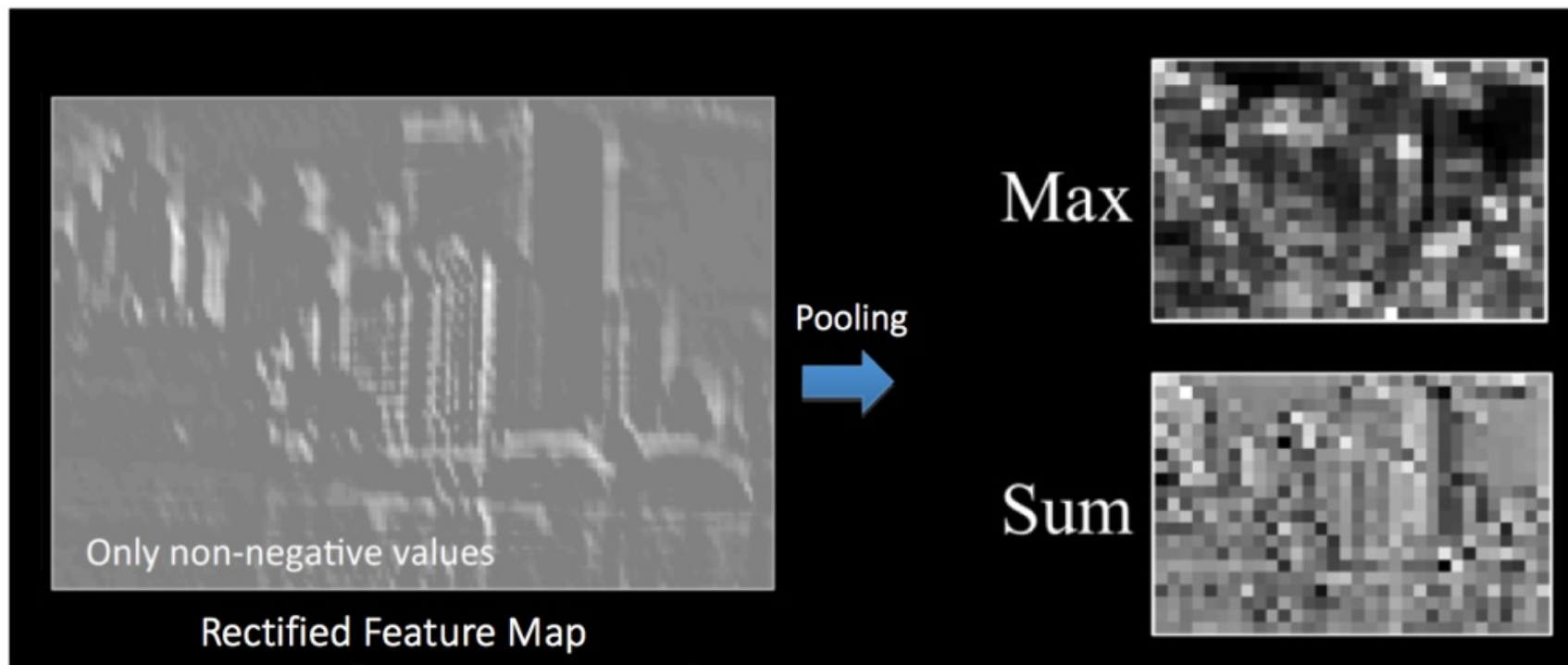
输出特征映射组 \mathbf{Y}

1	2
0	3

输出特征映射 Y^d



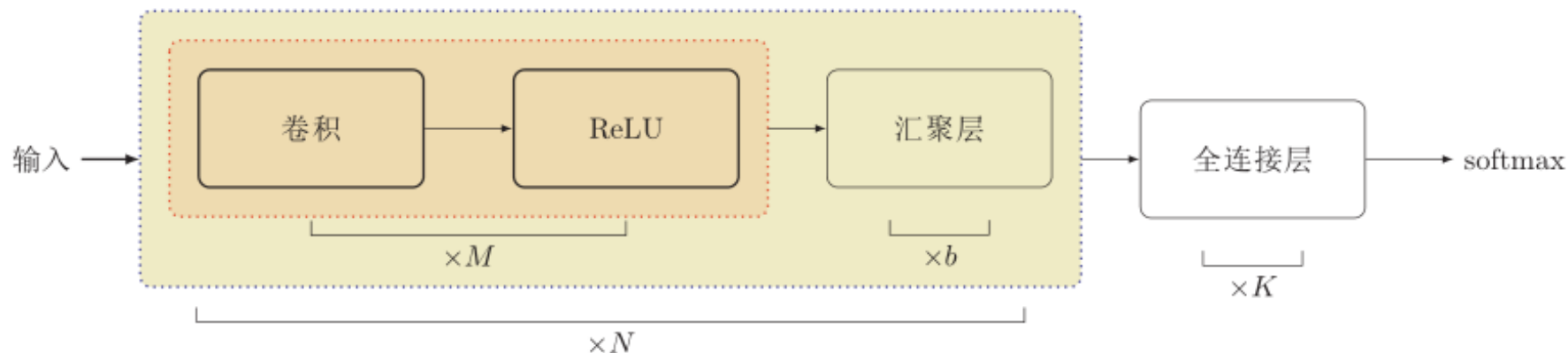
- 基本操作单元：池化Pooling / 降采样 层



- 卷积网络是由卷积层、汇聚层、全连接层交叉堆叠而成。

- 趋向于小卷积、大深度
- 趋向于全卷积

- 典型结构



- 一个卷积块为连续 M 个卷积层和 b 个汇聚层 (M 通常设置为 $2 \sim 5$, b 为 0 或 1)。一个卷积网络中可以堆叠 N 个连续的卷积块, 然后在接着 K 个全连接层 (N 的取值区间比较大, 比如 $1 \sim 100$ 或者更大; K 一般为 $0 \sim 2$)。

- 深度特征学习

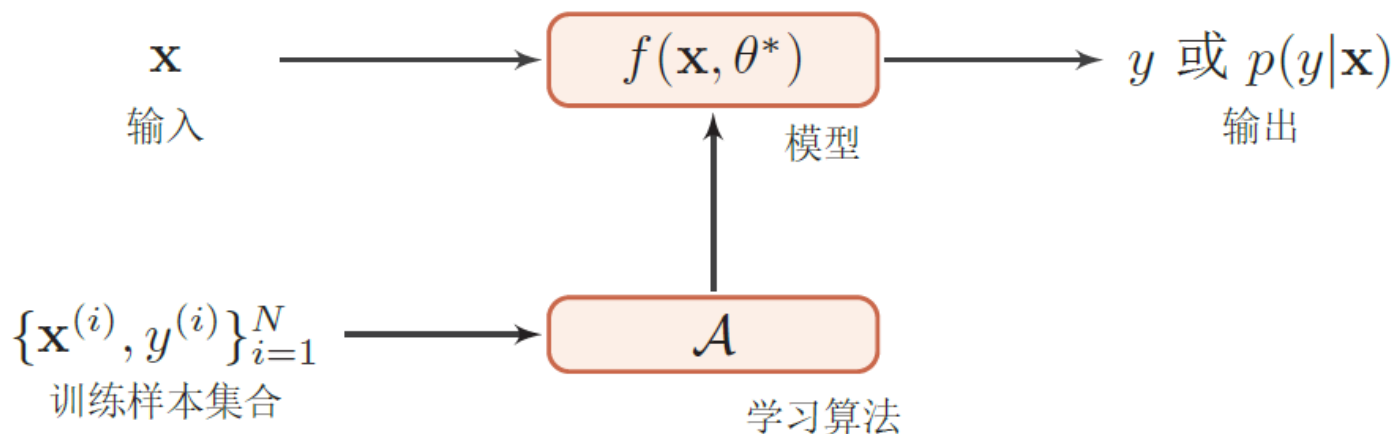
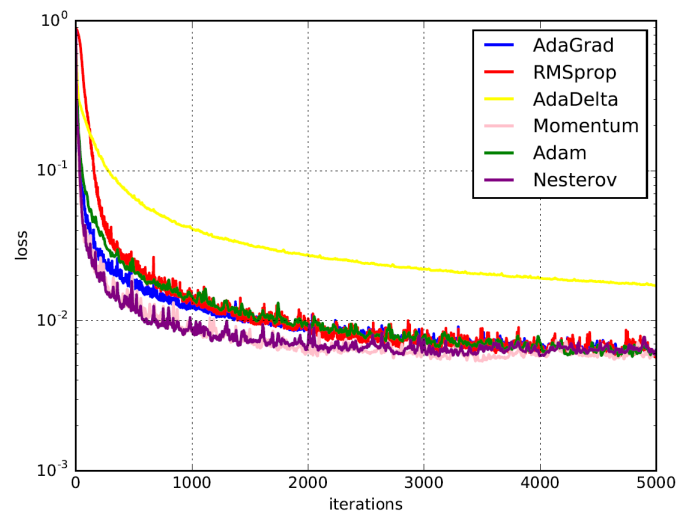


图 2.2 机器学习系统示例

$$L = 1, \text{ if } y \neq f(x)$$

$$L = CE(y, f(x)) \text{ 交叉熵}$$

$$L = ||y - f(x)||_1$$



- 深度特征学习

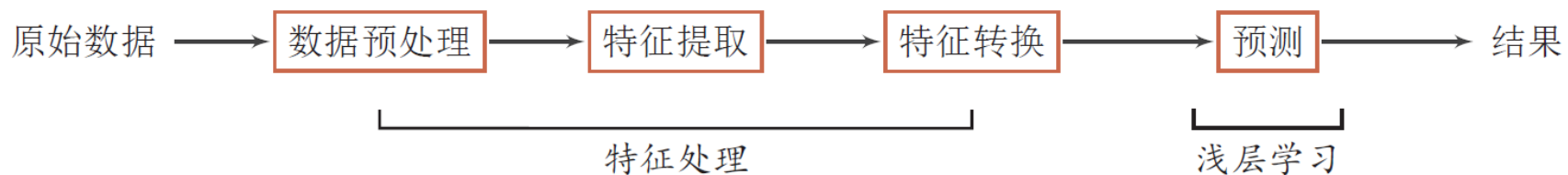


图 1.3 传统机器学习的数据处理流程

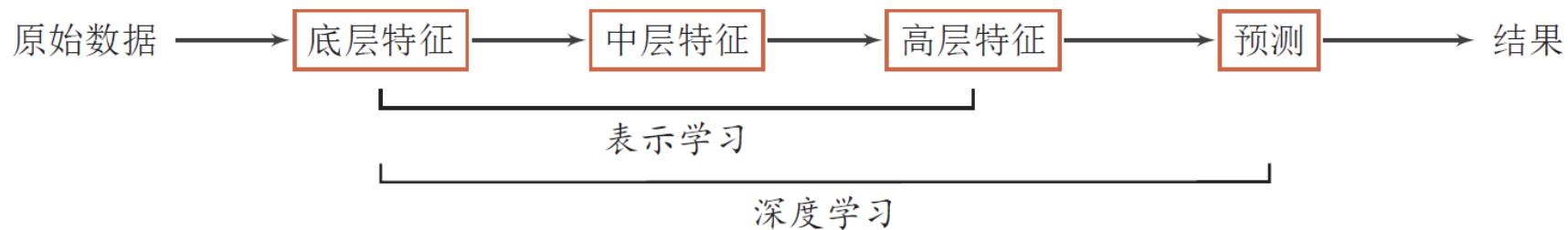
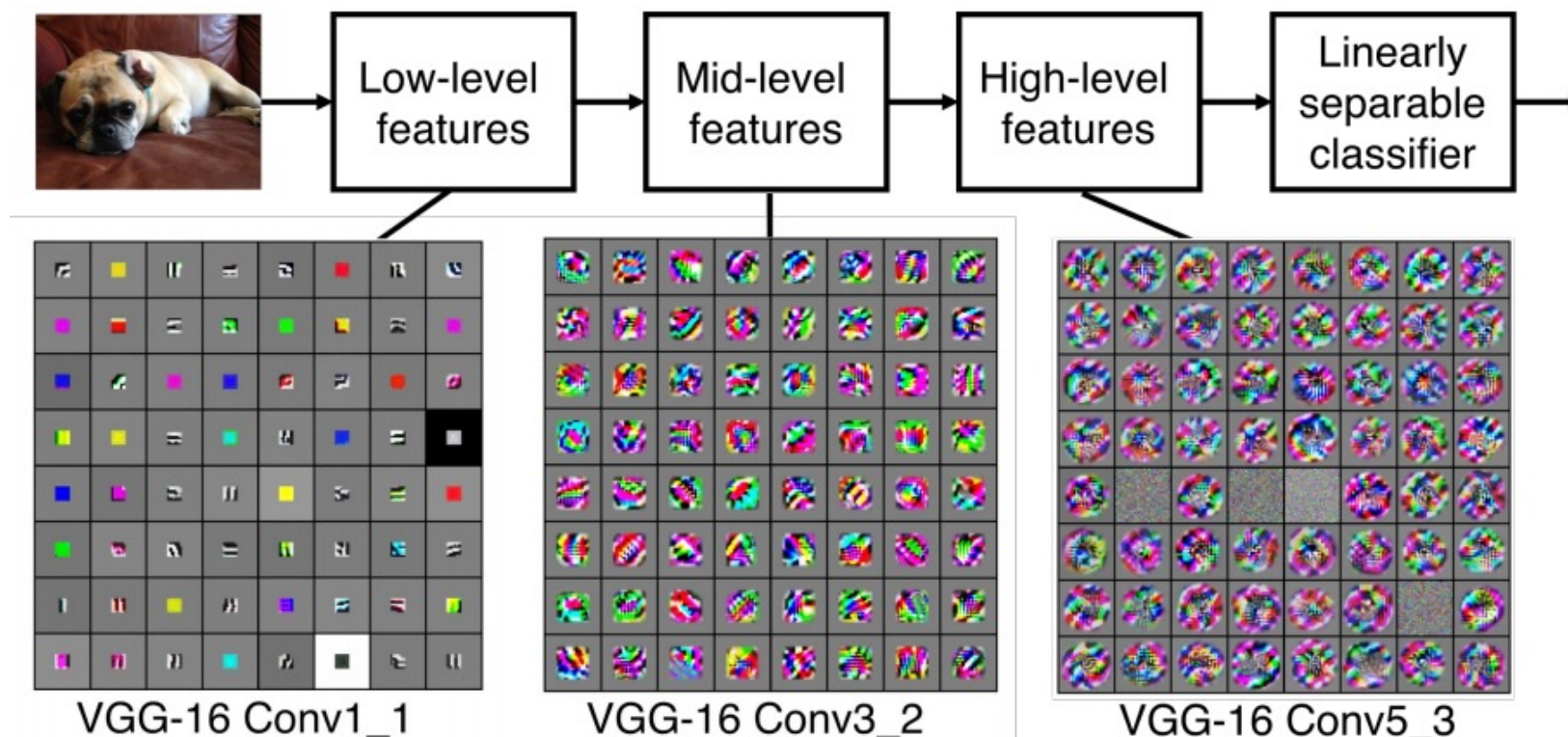
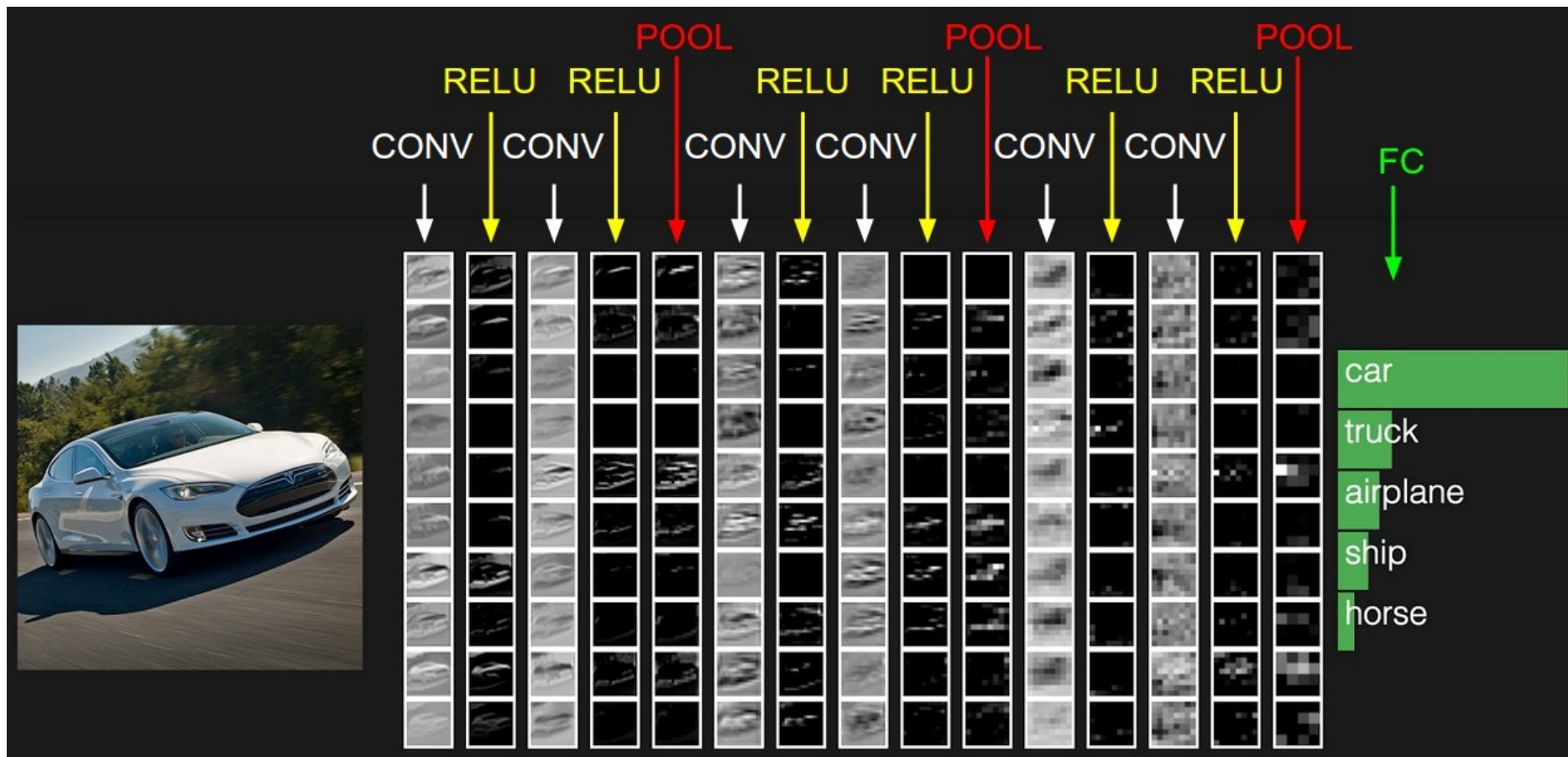


图 1.5 深度学习的数据处理流程

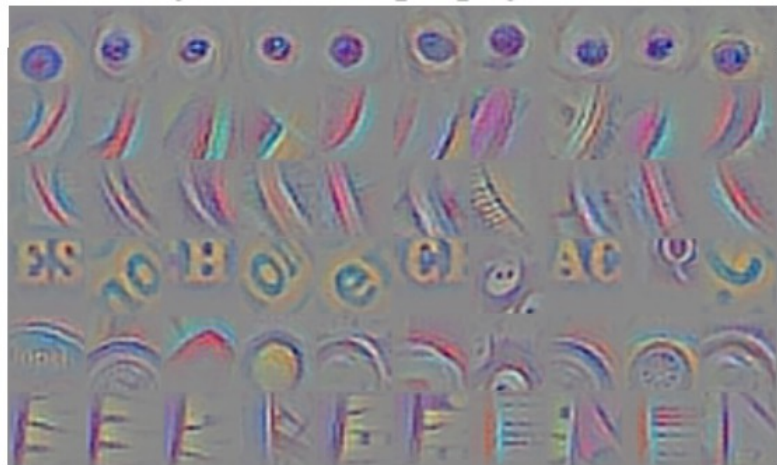
- CNN以图像的原始像素作为输入，基于输出层定义的损失函数使用反向传播算法端到端(End-to-end)学习，从而自动学习得到图像底层到高层的层次化语义表达



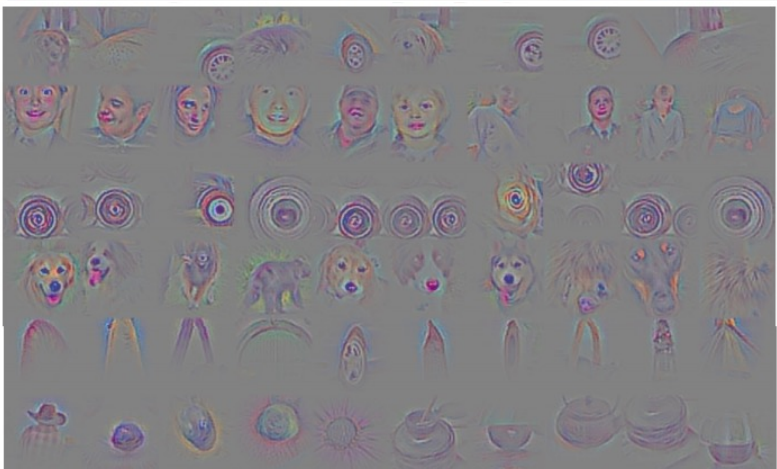


- 神经网络可视化:

conv6



conv9





● 为何选择 “深” 而非 “广” 的网络结构

➤ 即使只有一层隐层，只要有足够的神经元，神经网络理论上可以拟合任意连续函数。为什么还要使用深层网络结构？

■ 深度网络可从局部到整体 “理解图像”

学习复杂特征时（例如人脸识别），浅层的卷积层感受野小，学习到局部特征，深层的卷积层感受野大，学习到整体特征。

■ 深度网络可减少权重数量

以宽度换深度，用多个小卷积替代一个大卷积，在获得更多样特征的同时所需权重数量也更少。





神经网络实现图像分类的具体流程如下：

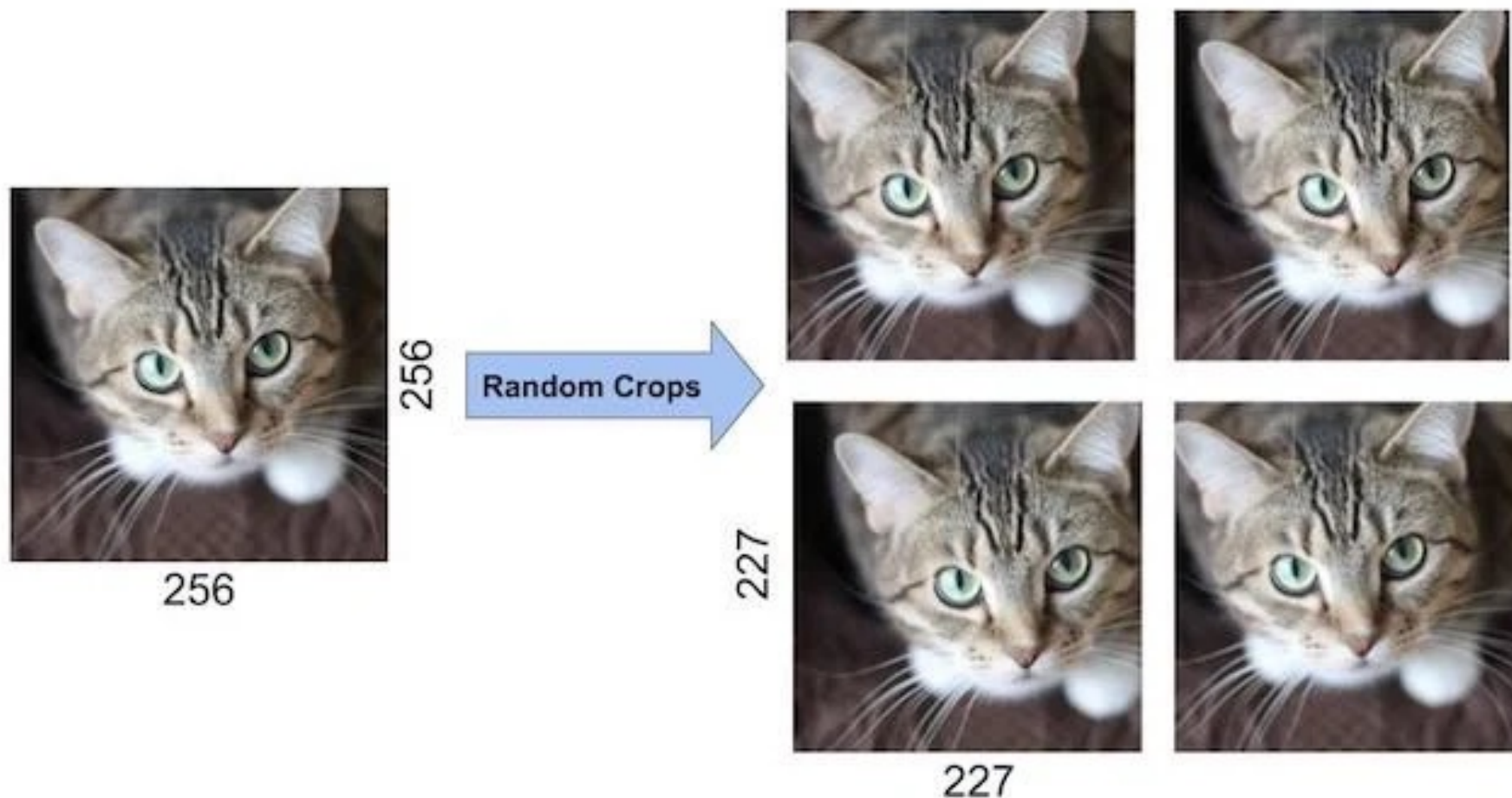
1. 数据预处理：将图像数据进行预处理，包括图像归一化、图像增强、数据增强等，以提高模型的鲁棒性和泛化能力。
2. 模型设计：选择适当的神经网络模型，包括卷积神经网络（CNN）、残差神经网络（ResNet）、Inception 等，设计并搭建对应的模型结构。
3. 模型训练：使用训练数据对神经网络进行训练，通过反向传播算法来调整网络中的权重和偏置，使其逐渐适应图像分类任务。
4. 模型评估：使用测试数据对训练好的模型进行评估，计算模型的准确率、精度、召回率等指标，以评估模型的性能。
5. 模型优化：对模型进行优化，包括超参数调整、正则化、优化器选择等，以提高模型的性能和泛化能力。
6. 模型调用：将训练好的模型部署到实际应用场景中，对输入的图像数据进行分类预测。



- 减少过拟合的最简单方法是增加训练样本。

- 图像增强方法

- 旋转 (rotating)
- 翻转 (flipping)
- 放缩 (scaling)
- 平移 (shifting)
- ...



4. (5分) 在神经网络中, 非线性激活函数 (如 Sigmoid, ReLU, Tanh 等) 的主要作用是什么?
- A. 确保网络的输出总是线性的, 便于解释。
 - B. 使网络能够学习和表示非线性复杂模式。
 - C. 显著加快模型的训练速度, 减少计算量。
 - D. 将每个神经元的输入数据进行归一化处理, 防止梯度消失。

4. (5分) 在神经网络中, 非线性激活函数 (如 Sigmoid, ReLU, Tanh 等) 的主要作用是什么?
- A. 确保网络的输出总是线性的, 便于解释。
 - B. 使网络能够学习和表示非线性复杂模式。
 - C. 显著加快模型的训练速度, 减少计算量。
 - D. 将每个神经元的输入数据进行归一化处理, 防止梯度消失。

8. 构建图像识别 AI 的基础模块 (假设基于一个典型的图像分类项目经验)你回顾了你的初始训练配置, 其核心参数如下:

- (1) (3 分) 在搭建一个基础的 AI 模型来识别图片中的物体 (例如, 猫或狗) 时, 通常会使用卷积神经网络 (CNN)。CNN 中负责从图片中提取初步视觉特征 (如边缘、角点) 的是什么类型的层? 它是如何工作的?
- (2) (3 分) 在上述特征提取层之后, 通常会紧跟一个什么组件来为网络引入非线性能力? 为什么这种非线性能力对图像识别 AI 如此重要?
- (3) (3 分) 为了降低模型复杂度、减少计算量, 并使模型对物体在图片中的微小位置变化不那么敏感, CNN 中经常会用到什么技术或类型的层?
- (4) (3 分) 在网络经过多层特征提取和处理后, 最终负责根据学习到的高级特征做出分类判断 (例如, 判断图片是猫还是狗) 的是什么类型的层?

8. 构建图像识别 AI 的基础模块 (假设基于一个典型的图像分类项目经验)你回顾了你的初始训练配置, 其核心参数如下:

- (1) (3 分) 在搭建一个基础的 AI 模型来识别图片中的物体 (例如, 猫或狗) 时, 通常会使用卷积神经网络 (CNN)。CNN 中负责从图片中提取初步视觉特征 (如边缘、角点) 的是什么类型的层? 它是如何工作的?
- (2) (3 分) 在上述特征提取层之后, 通常会紧跟一个什么组件来为网络引入非线性能力? 为什么这种非线性能力对图像识别 AI 如此重要?
- (3) (3 分) 为了降低模型复杂度、减少计算量, 并使模型对物体在图片中的微小位置变化不那么敏感, CNN 中经常会用到什么技术或类型的层?
- (4) (3 分) 在网络经过多层特征提取和处理后, 最终负责根据学习到的高级特征做出分类判断 (例如, 判断图片是猫还是狗) 的是什么类型的层?

答案:

- (1) CNN 中负责提取初步视觉特征的是卷积层 (Convolutional Layer)。它通过使用多个可学习的卷积核 (也叫滤波器) 在输入图像上进行滑动 (卷积) 运算, 每个卷积核对图像的局部区域进行加权求和, 从而检测出边缘、角点、颜色块等局部模式。
- (2) 卷积层之后通常会紧跟一个激活函数 (Activation Function), 如 ReLU。引入非线性能力至关重要, 因为现实世界中的图像内容和物体间的关系通常是高度非线性的。如果没有非线性激活, 多层神经网络的表达能力将等同于单层网络, 无法学习复杂的视觉模式。
- (3) CNN 中经常使用池化层 (Pooling Layer) (如最大池化 Max Pooling)。它的作用是对特征图进行下采样, 减少特征图的尺寸, 从而降低后续计算的复杂度, 同时也有助于使模型对物体在图像中的小幅平移、旋转或缩放具有一定的不变性。
- (4) 在网络尾部, 负责根据学习到的高级抽象特征做出最终分类判断的是全连接层 (Fully Connected Layer / Dense Layer), 其最后一层通常会配合 Softmax 激活函数 (对于多分类任务) 输出每个类别的概率。

10. 图像分类 AI 的输出层 (Softmax 计算)

一个用于图像分类的人工智能模型（例如，识别图片是“猫”、“狗”还是“鸟”）在最后输出层之前，会为每个类别计算一个原始得分 (logits)。为了将这些原始得分转换成表示每个类别概率的、和为 1 的数值，通常会使用 Softmax 函数。

给定以下模型的原始输出得分 (logits):

* 类别“猫” (Cat) 的得分: $z_{cat} = 2.0$

* 类别“狗” (Dog) 的得分: $z_{dog} = 1.0$

* 类别“鸟” (Bird) 的得分: $z_{bird} = 0.5$

已知近似值 (如果需要)

* $e^{2.0} \approx 7.39$

* $e^{1.0} \approx 2.72$

* $e^{0.5} \approx 1.65$

(1) (4分) 计算所有类别得分的指数值之和 (即 Softmax 公式中的分母 $\sum_{j=1}^K e^{z_j}$)。

(2) (4分) 分别计算图片被识别为“猫”、“狗”、“鸟”的概率 (即 $S(z_{cat})$, $S(z_{dog})$, $S(z_{bird})$)。请将结果保留到小数点后三位。

(3) (6分) 这三个计算出的概率值加起来大约等于多少? 这个结果符合 Softmax 函数的特性吗?

10. 图像分类 AI 的输出层 (Softmax 计算)

一个用于图像分类的人工智能模型（例如，识别图片是“猫”、“狗”还是“鸟”）在最后输出层之前，会为每个类别计算一个原始得分 (logits)。为了将这些原始得分转换成表示每个类别概率的、和为 1 的数值，通常会使用 Softmax 函数。

给定以下模型的原始输出得分 (logits):

- * 类别“猫” (Cat) 的得分: $z_{cat} = 2.0$
- * 类别“狗” (Dog) 的得分: $z_{dog} = 1.0$
- * 类别“鸟” (Bird) 的得分: $z_{bird} = 0.5$

已知近似值 (如果需要)

- * $e^{2.0} \approx 7.39$
- * $e^{1.0} \approx 2.72$
- * $e^{0.5} \approx 1.65$

- (1) (4分) 计算所有类别得分的指数值之和 (即 Softmax 公式中的分母 $\sum_{j=1}^K e^{z_j}$)。
- (2) (4分) 分别计算图片被识别为“猫”、“狗”、“鸟”的概率 (即 $S(z_{cat})$, $S(z_{dog})$, $S(z_{bird})$)。请将结果保留到小数点后三位。
- (3) (6分) 这三个计算出的概率值加起来大约等于多少? 这个结果符合 Softmax 函数的特性吗?

答案:

- (1) 所有类别得分的指数值之和为: $A = e^2 + e^1 + e^{0.5} \approx 11.76$
- (2) 分别计算各类别概率

$$S(z_{cat}) \approx 0.628$$

$$S(z_{dog}) \approx 0.231$$

$$S(z_{bird}) \approx 0.140$$

- (3) 概率值加和验证: $0.628 + 0.231 + 0.140 \approx 0.999$, 非常接近于 1, 符合 Softmax 的特性

目录

1

计算机视觉简介

2

卷积神经网络

3

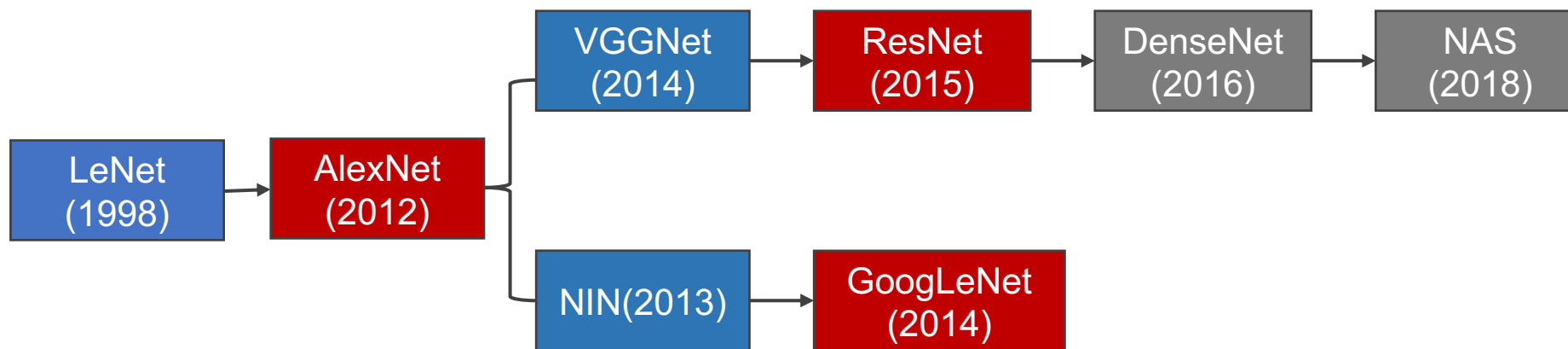
视觉模型演化历程

4

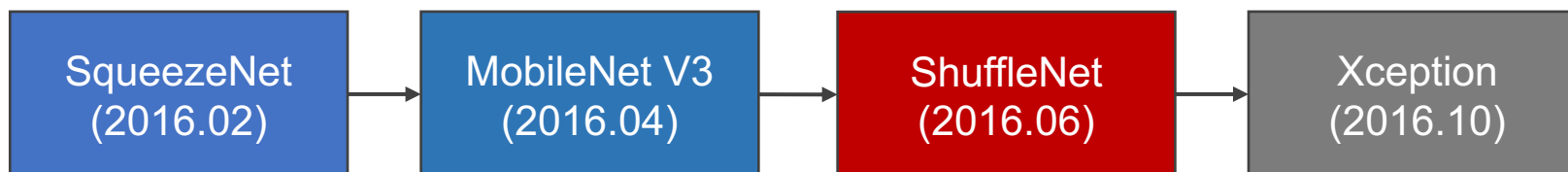
计算机视觉应用场景



• CNN模型结构变迁



■ 针对移动端的轻量型网络



MNIST数据集：经典中的经典



MNIST数据集

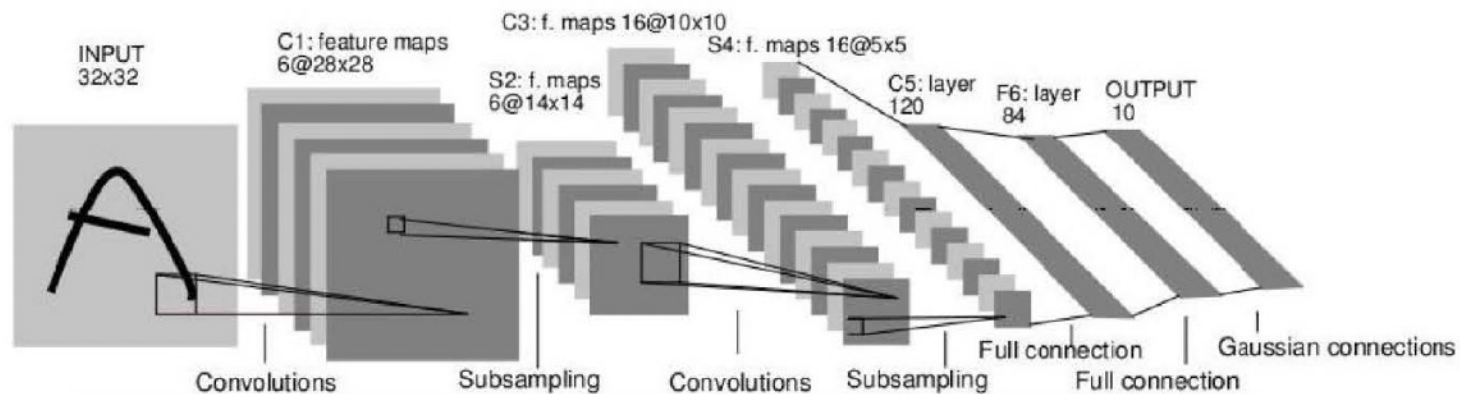
<http://yann.lecun.com/exdb/mnist/>

Yann LeCun et al. 1998

LeNet: 最经典的卷积神经网络

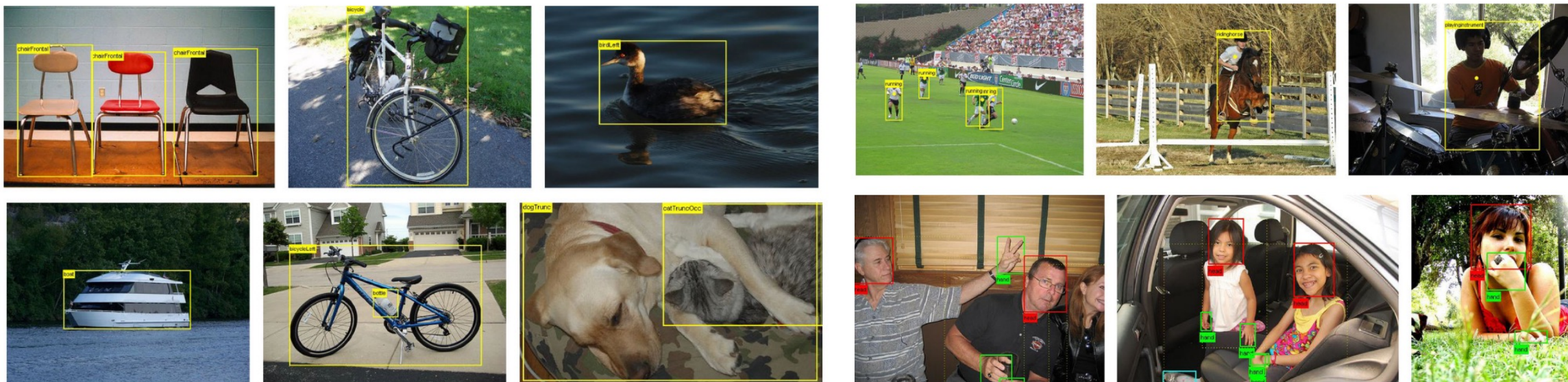


Yann LeCun
纽约大学教授



Layer	Kernels	Stride	Padding	Output	Parameters
Input				32*32*1	
C1	5*5*1 filters * 6	1	Valid	28*28*6	(5*5*1+1)*6
S2	2*2 filter	2		14*14*6	
C3	5*5*6 filters * 16	1	Valid	10*10*16	(5*5*6+1)*16
S4	2*2 filter	2		5*5*16	
C5	5*5*16 filters * 120	1		120	(5*5*16)*120
F6				84	120*84
Output				10	84*10

进入构建大数据时代: Pascal VOC

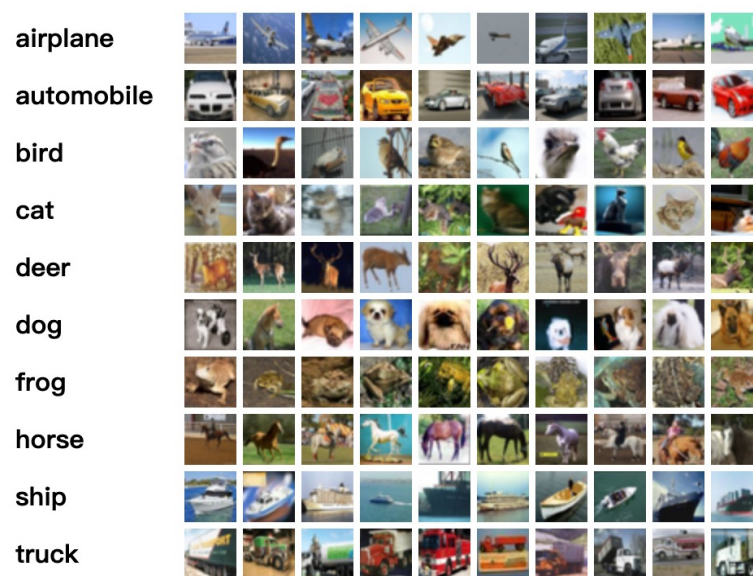


物体检测数据集 (20个类)

Pascal VOC Challenges 2005-2012

<http://host.robots.ox.ac.uk/pascal/VOC/>

CIFAR-10数据集



CIFAR-10数据集

<https://www.cs.toronto.edu/~kriz/cifar.html>

Alex Krizhevsky, Geoffrey Hinton, 2009

ImageNet数据集和ILSVRC



Fei-Fei Li

斯坦福大学教授

美国国家工程院院士

美国国家医学院院士

以人为本人工智能研究院院长



<https://www.image-net.org/>

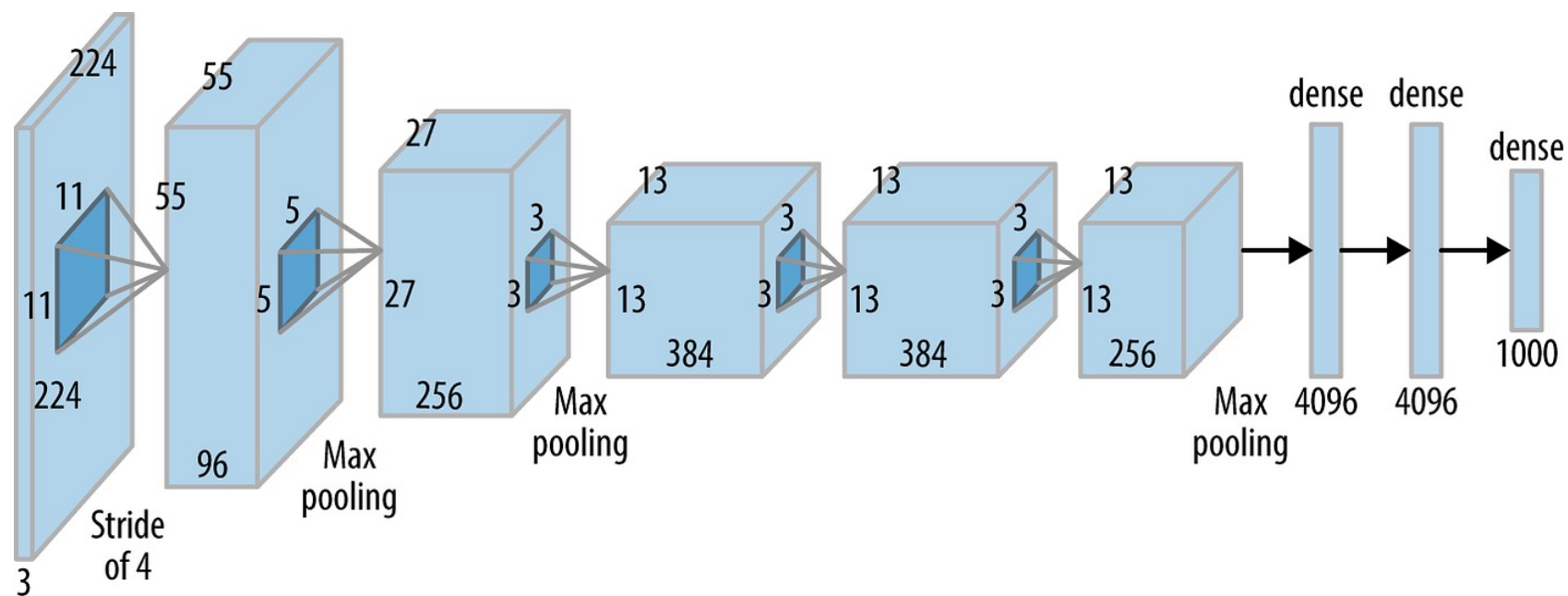
ILSVRC: ImageNet Large Scale Visual Recognition Competition,

<http://rogerioferis.com/VisualRecognitionAndSearch2014/material/papers/DPMCVPR98.pdf>

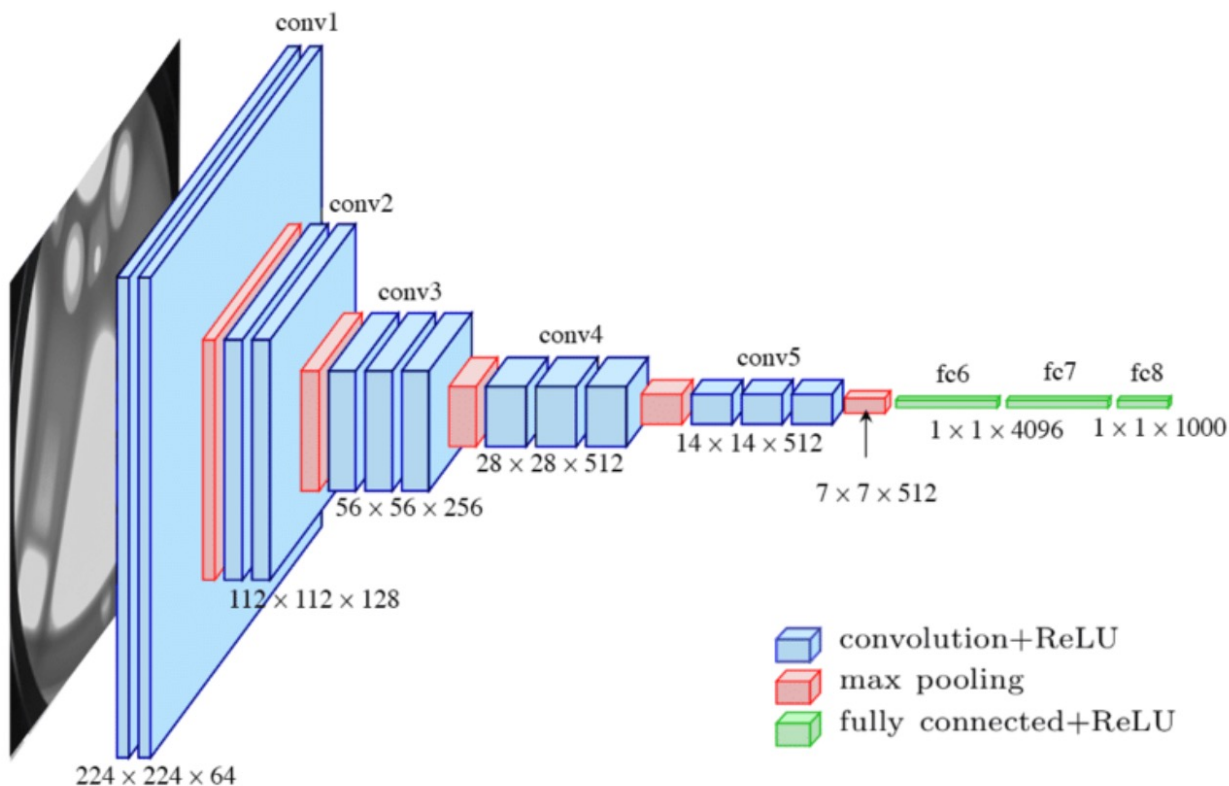
AlexNet: ImageNet 2012冠军，超出第二名10.9个百分点



Geoffrey Hinton



VGG-16: 2014年亚军(7.3%, 1亿3千8百万参数)



- ✓ 3x3卷积核（使用2个3*3的卷积核的组合比用1个5*5的卷积核 效果更佳 && 参数量降低）
- ✓ ReLU激活函数
- 3个全连接层导致参数爆炸，训练很慢，过拟合

✓ 参数量的评估:

以256通道的隐层数据为例:

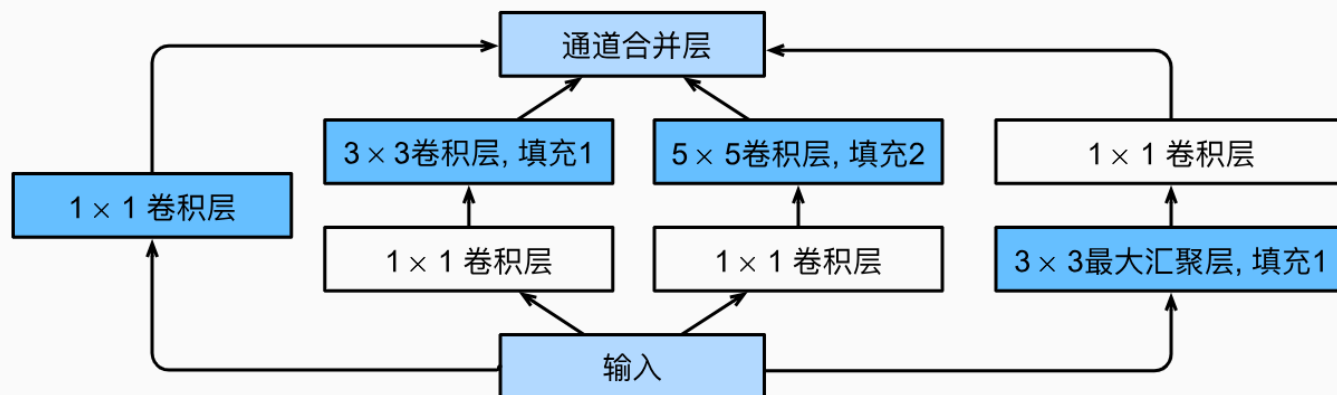
2个3*3的卷积核 == 1个5*5的卷积核

$2 \times 3 \times 3 \times 256$ params $5 \times 5 \times 256$ params

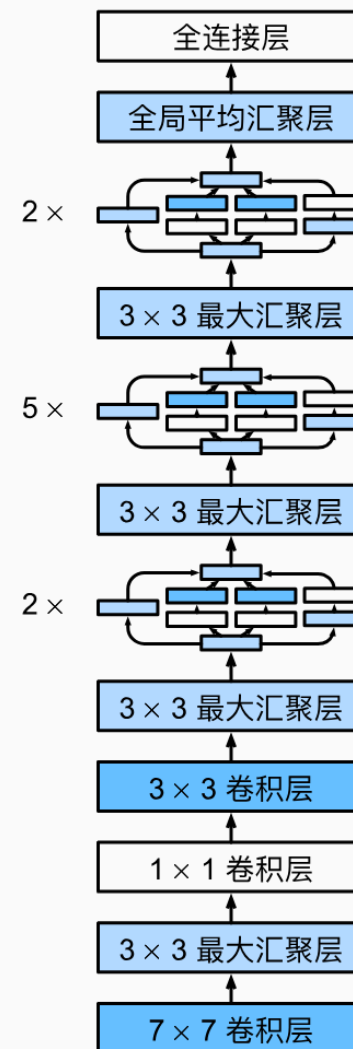
3个3*3的卷积核 == 1个7*7的卷积核

$3 \times 3 \times 3 \times 256$ params $7 \times 7 \times 256$ params

Inception V1 (GoogleNet) : 2014年冠军(6.67%, 400万参数)



Inception块的架构



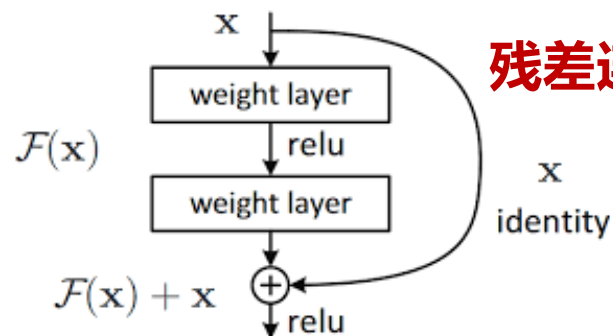
计算机视觉 | 2015: ResNet – “大道至简”



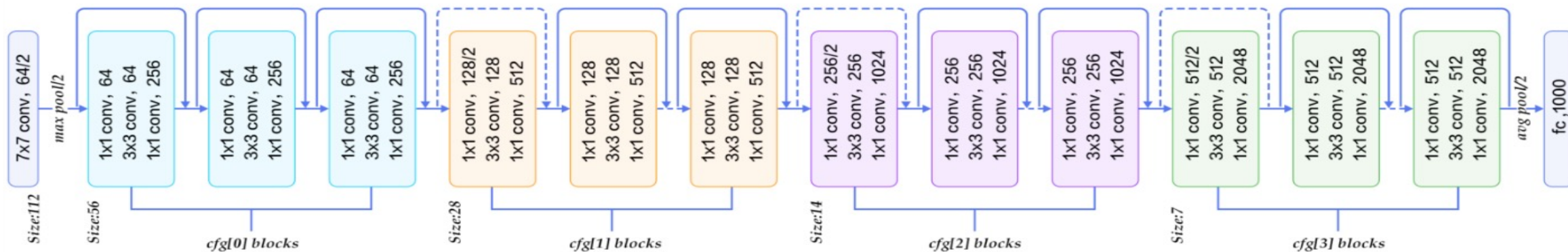
ResNet: 2015年冠军(3.57%, 首次超越人类, 152层, 6千万参数)



Kaiming He



50 layers	$cfg=[3,4,6,3]$
101 layers	$cfg=[3,4,23,8]$
152 layers	$cfg=[3,8,36,3]$



He, Kaiming, et al. "Deep residual learning for image recognition." CVPR, 2016.



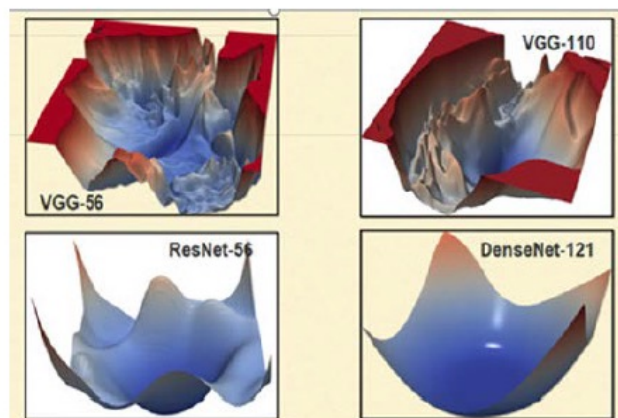
➤ 更激进的密集连接机制

每个层都会接受其前面所有层作为其额外的输入

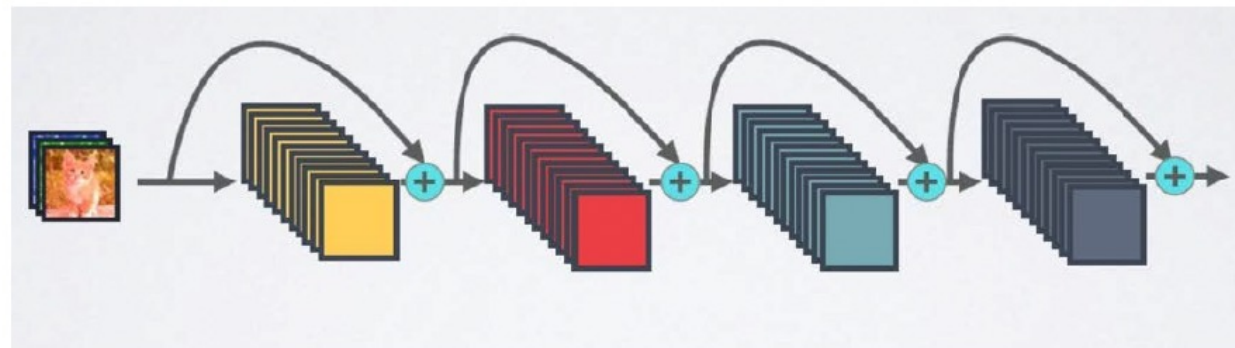
ResNet: 元素级相加

DenseNet: channel级连接

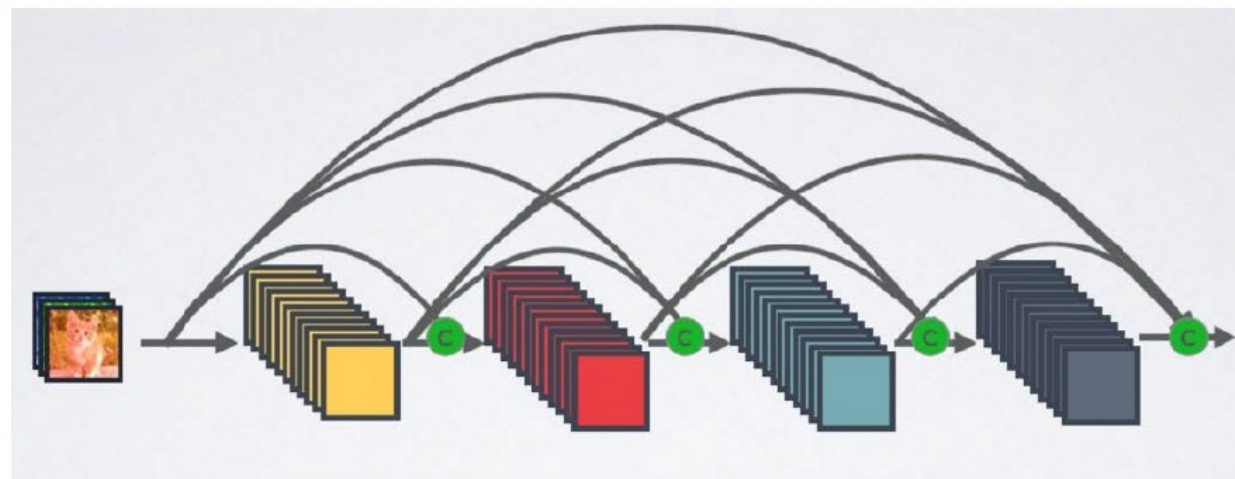
神经网络损失函数可视化曲面: 从VGG到DenseNet越来越光滑, 算法更容易找到好的优化点



DenseNet



ResNet



DenseNet

ViT: 把图像划分成16x16的方块，并将每个方块当成一个单词

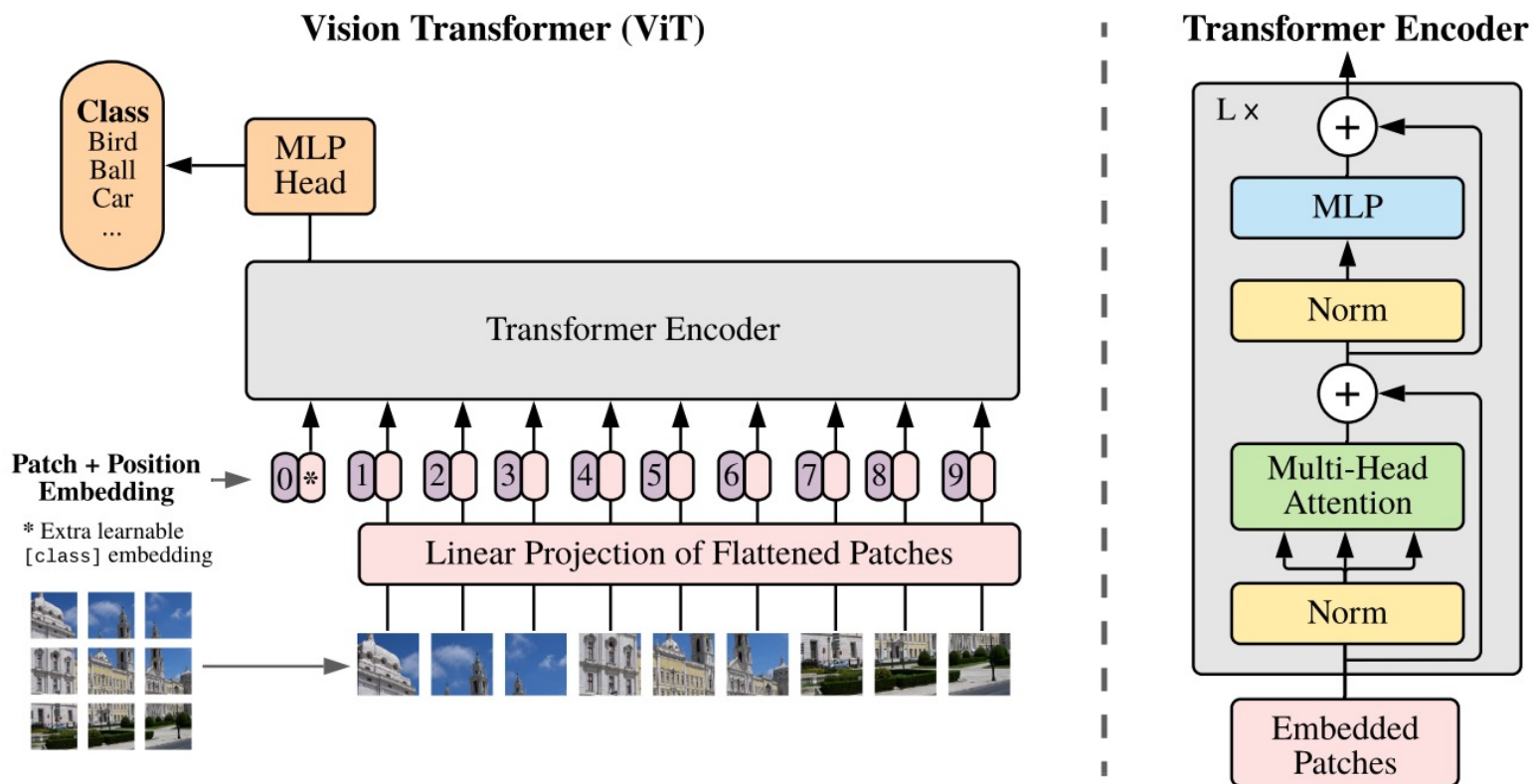
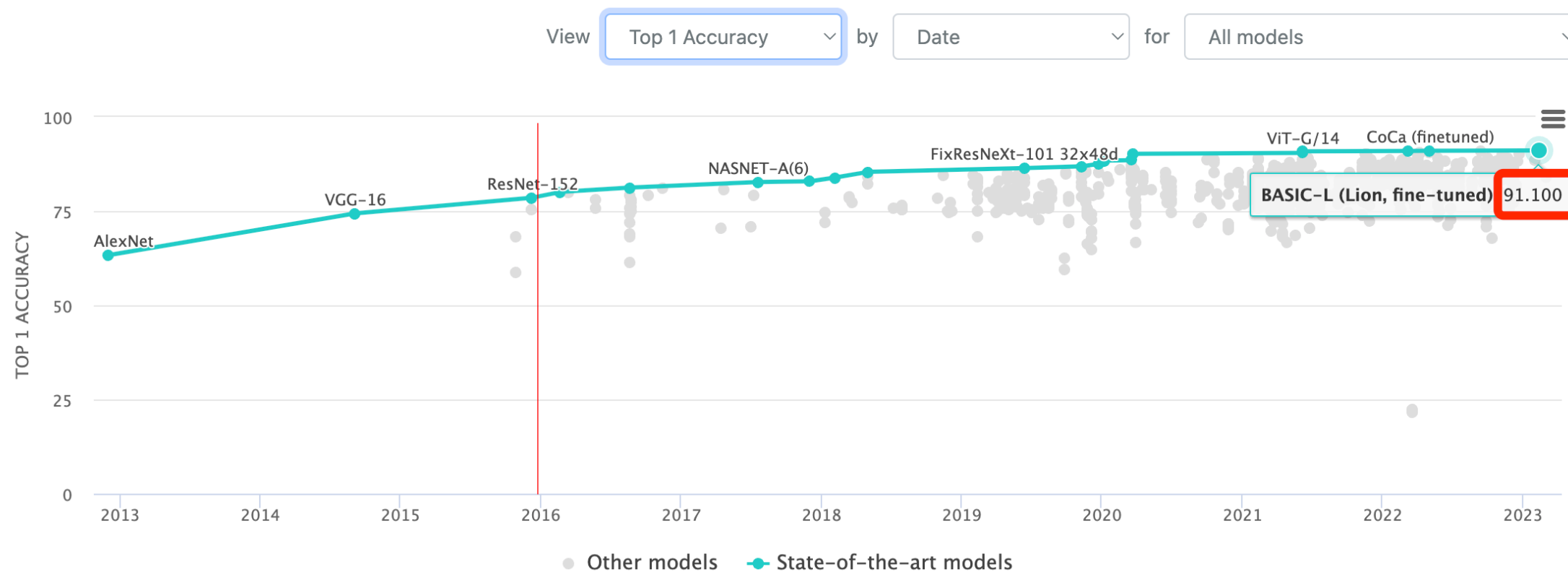


Image Classification on ImageNet

Leaderboard

Dataset



面向不同任务

目标检测

Faster RCNN
Cascade RCNN
SSD
YOLO
Mask RCNN
...

图像分割

FCN
DeepLab系列
SegNet
U-Net
Mask RCNN
...

前述卷积神经网络是其中的一部分

轻量级神经网络

相关技术

参数修剪和共享
低秩因子分解
紧凑卷积滤波器
知识蒸馏
神经网络搜索

典型结构

MobileNet系列
ShuffleNet系列
MixNet
Xception
...

模型小、资源消耗低，计算速度快、性能好

易于部署在移动端和嵌入式端

目录

1

计算机视觉简介

2

卷积神经网络

3

视觉模型演化历程

4

计算机视觉应用场景



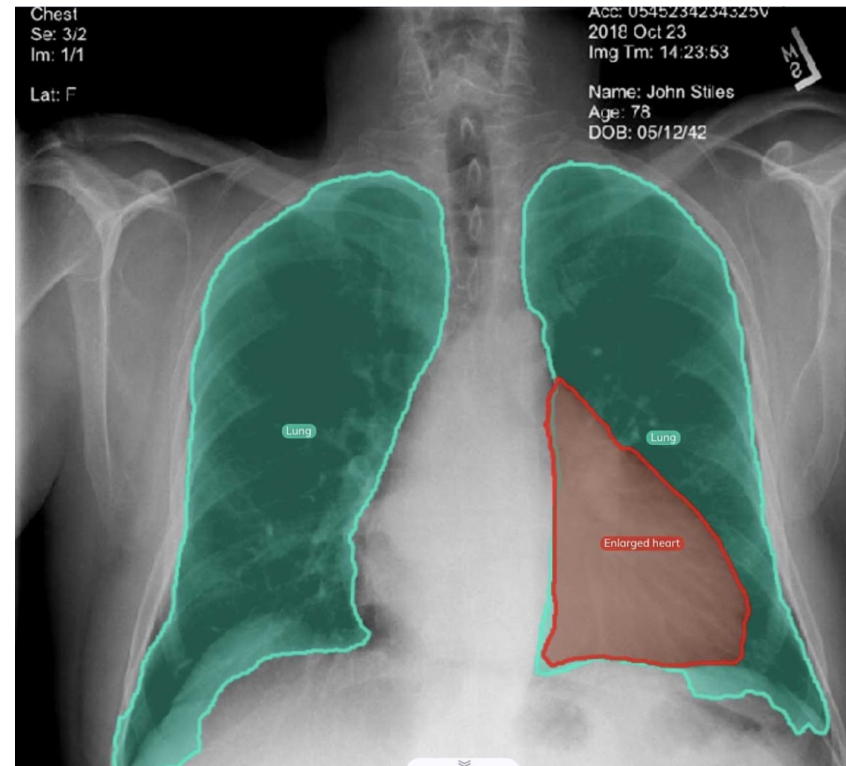
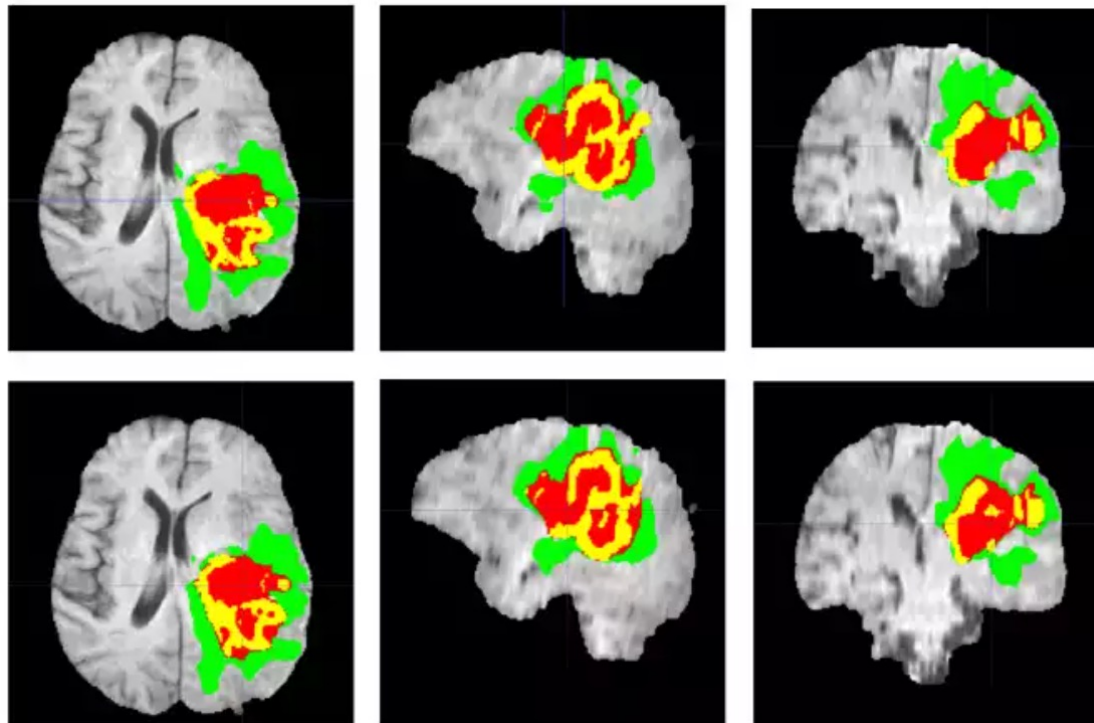
自动驾驶



交通灯检测



医疗诊断：脑肿瘤诊断



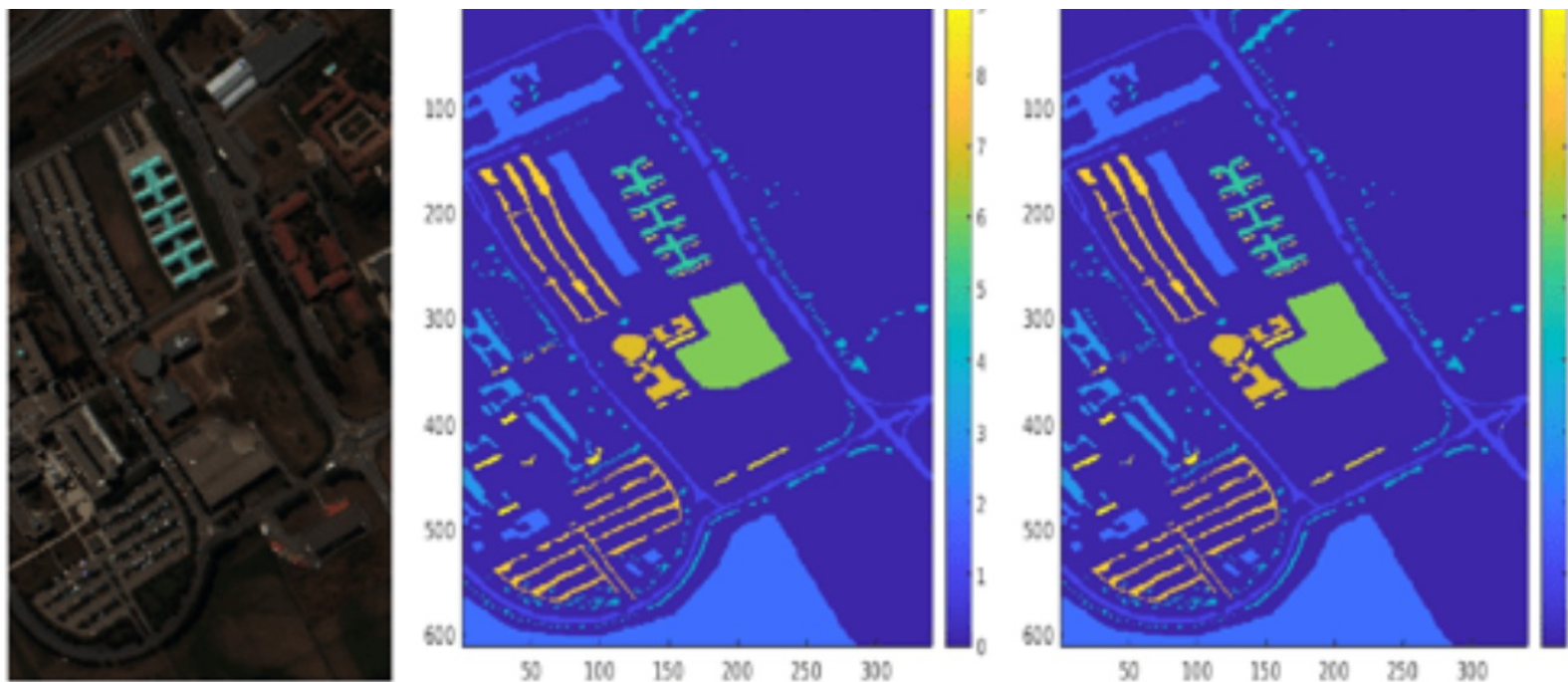
人脸识别



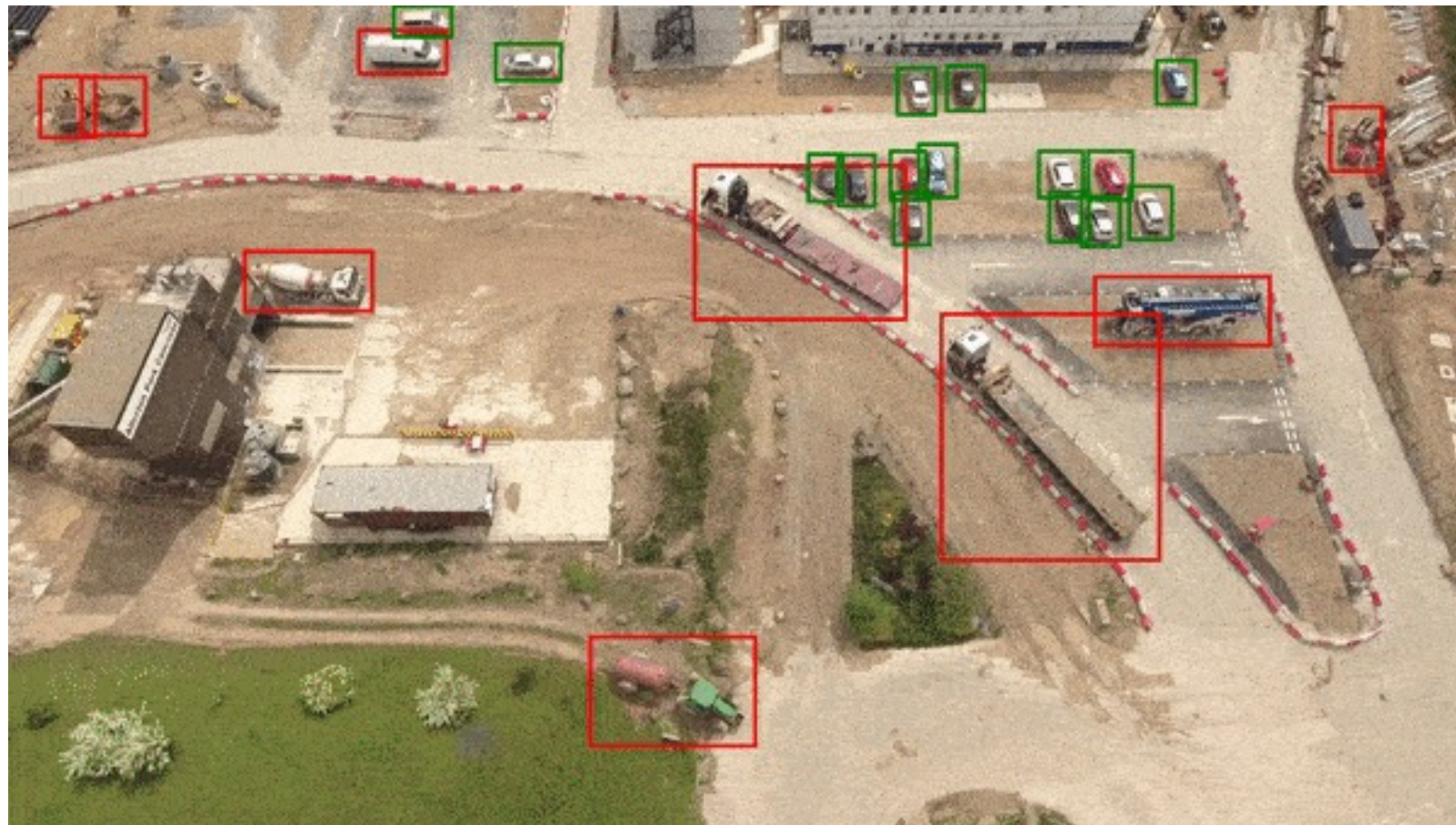
路况分割



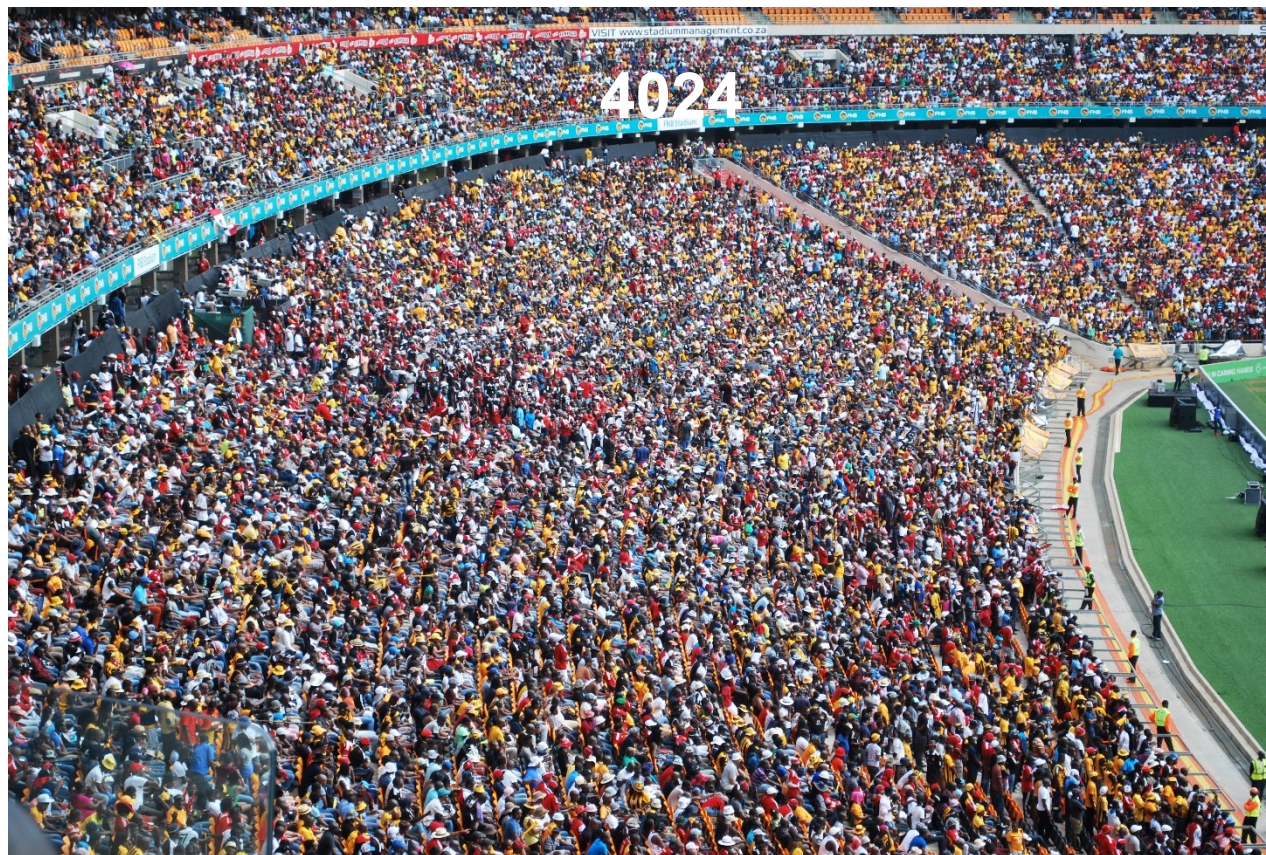
遥感图像分割



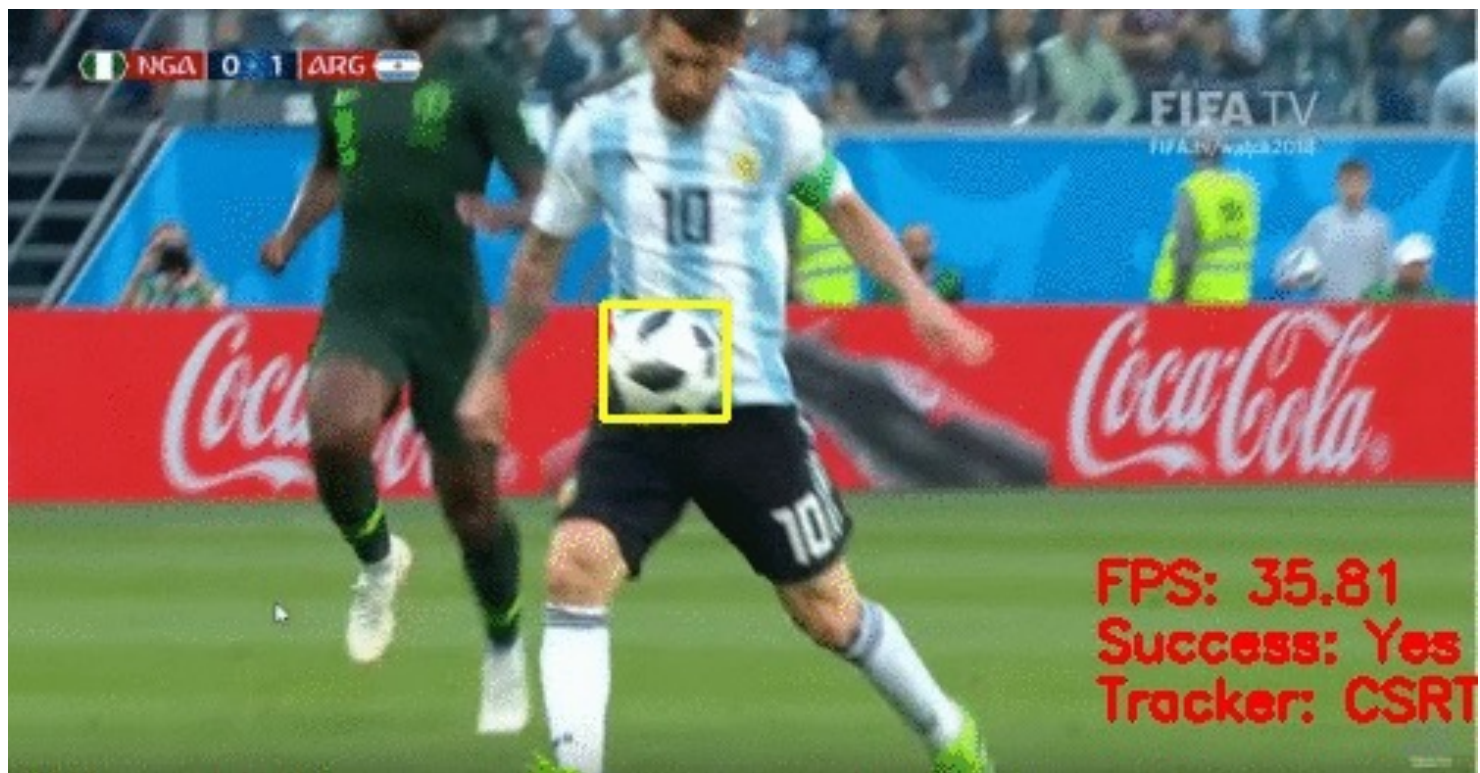
无人机物体识别



人群计数



物体追踪



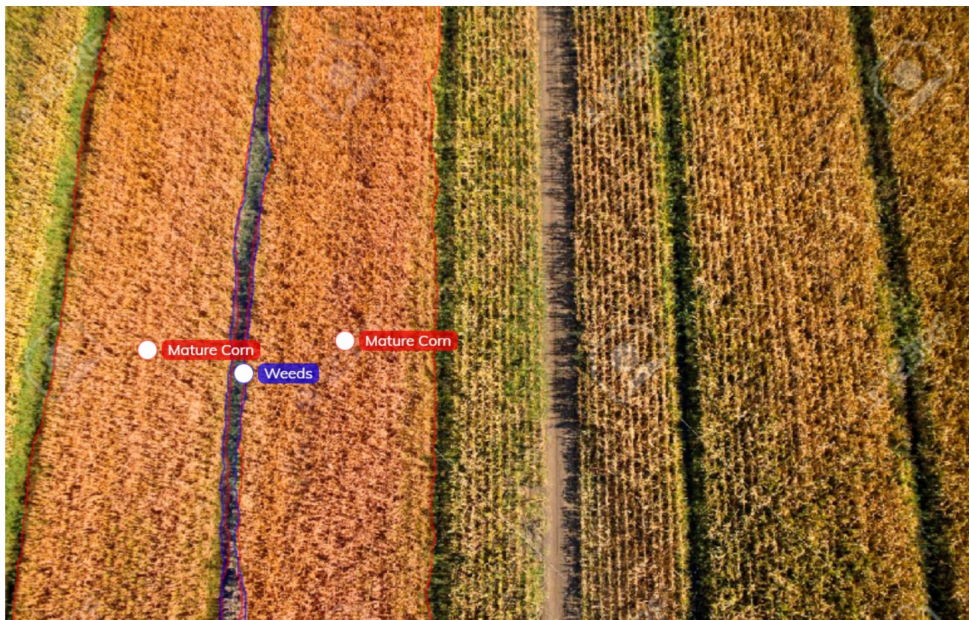
工业生产: 缺陷检测



建筑业: 安全风险检测



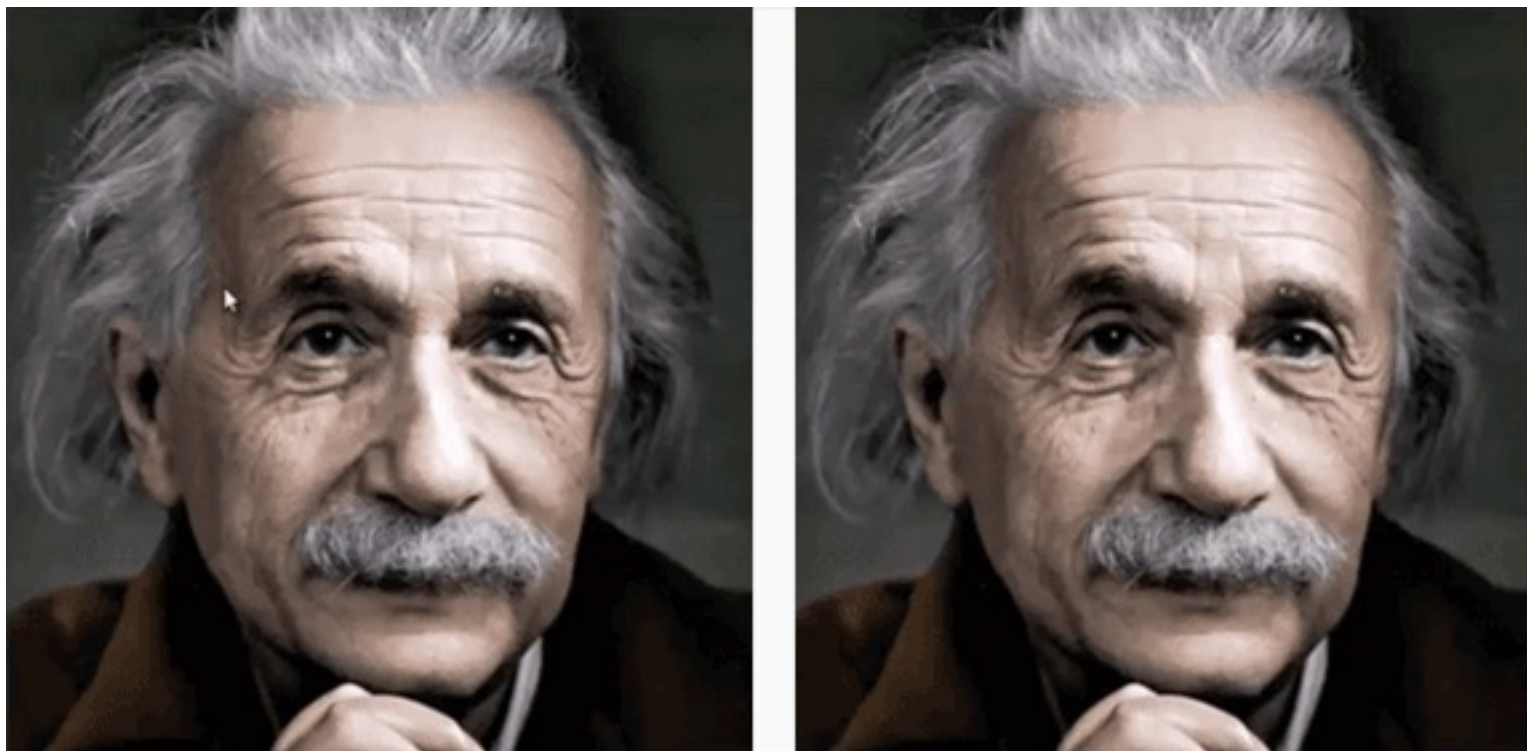
农业: 农作物健康监测



超分辨率



图像修复 (Image Inpainting)



视频修复 (Video Inpainting)

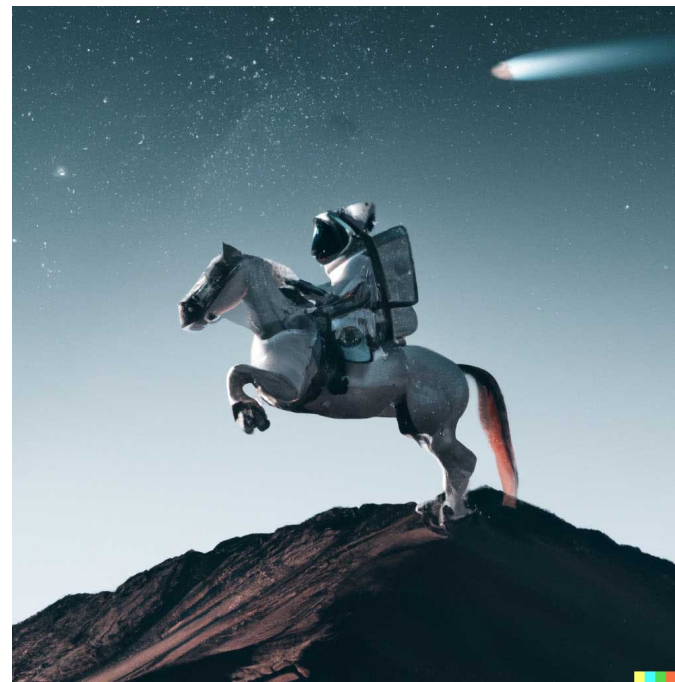
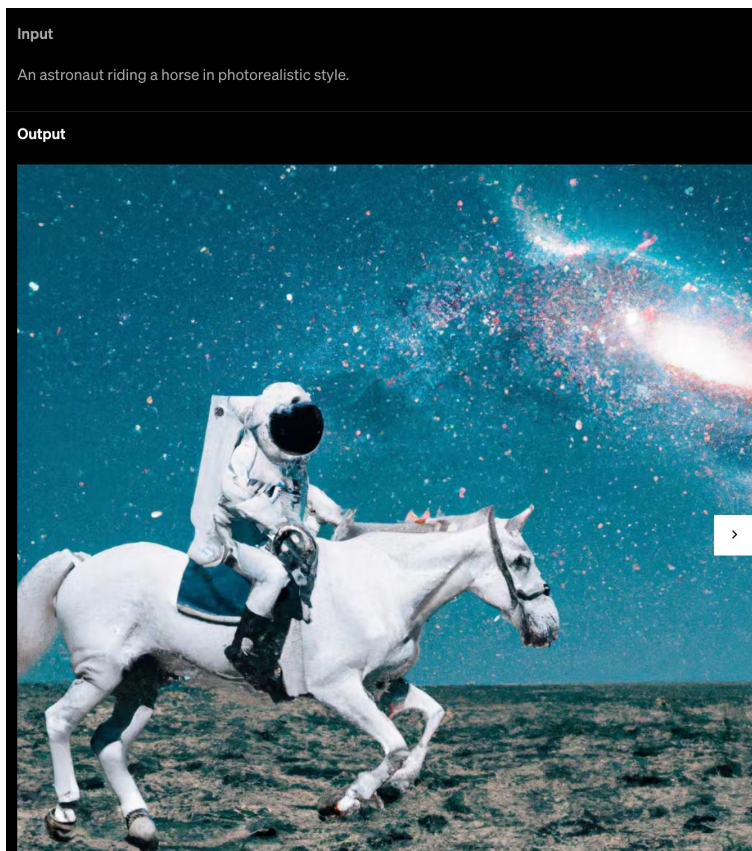


风格迁移 (Style Transfer)



图像生成

知名模型：DALL-E 2、Stable Diffusion 2、Midjourney、Imagen等



视频生成

如Meta公司的Make-A-Video等



A dog wearing a Superhero outfit with red cape flying through the sky



旧视频修复

1920年原片



人工智能上色+帧数修复



谢谢！