深蓝学院
shenlanxueyuan.com

深度学习理论与实践

第6课：图像分类

主讲人 洪振

中国科学院自动化研究所博士
模式识别国家重点实验室

课程内容

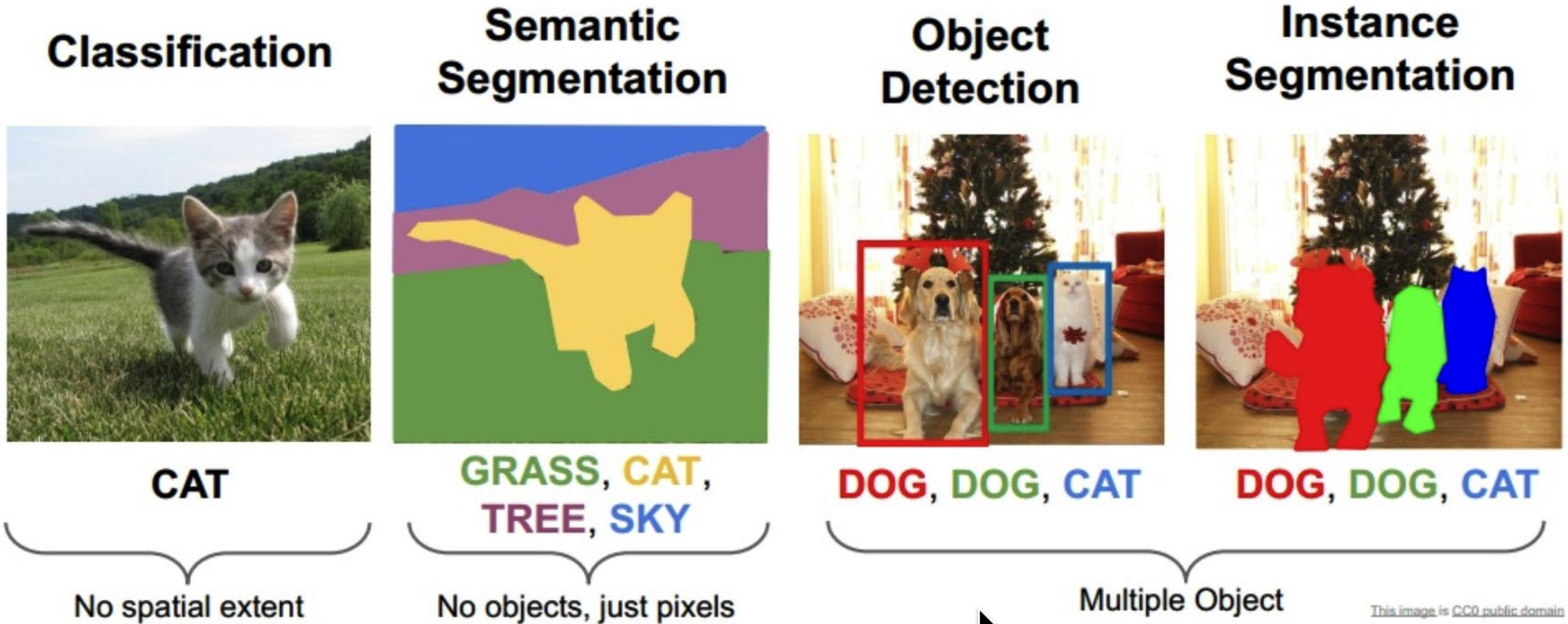
✓图像分类介绍

✓常见分类网络介绍

✓从图像分类到语义分割

图像分类简介

□任务定义



Classification

Semantic Segmentation

Object Detection

Instance Segmentation

CAT

GRASS, CAT, TREE, SKY

DOG, DOG, CAT

DOG, DOG, CAT

No spatial extent

No objects, just pixels

Multiple Object

This image is CC0 public domain

图像分类简介

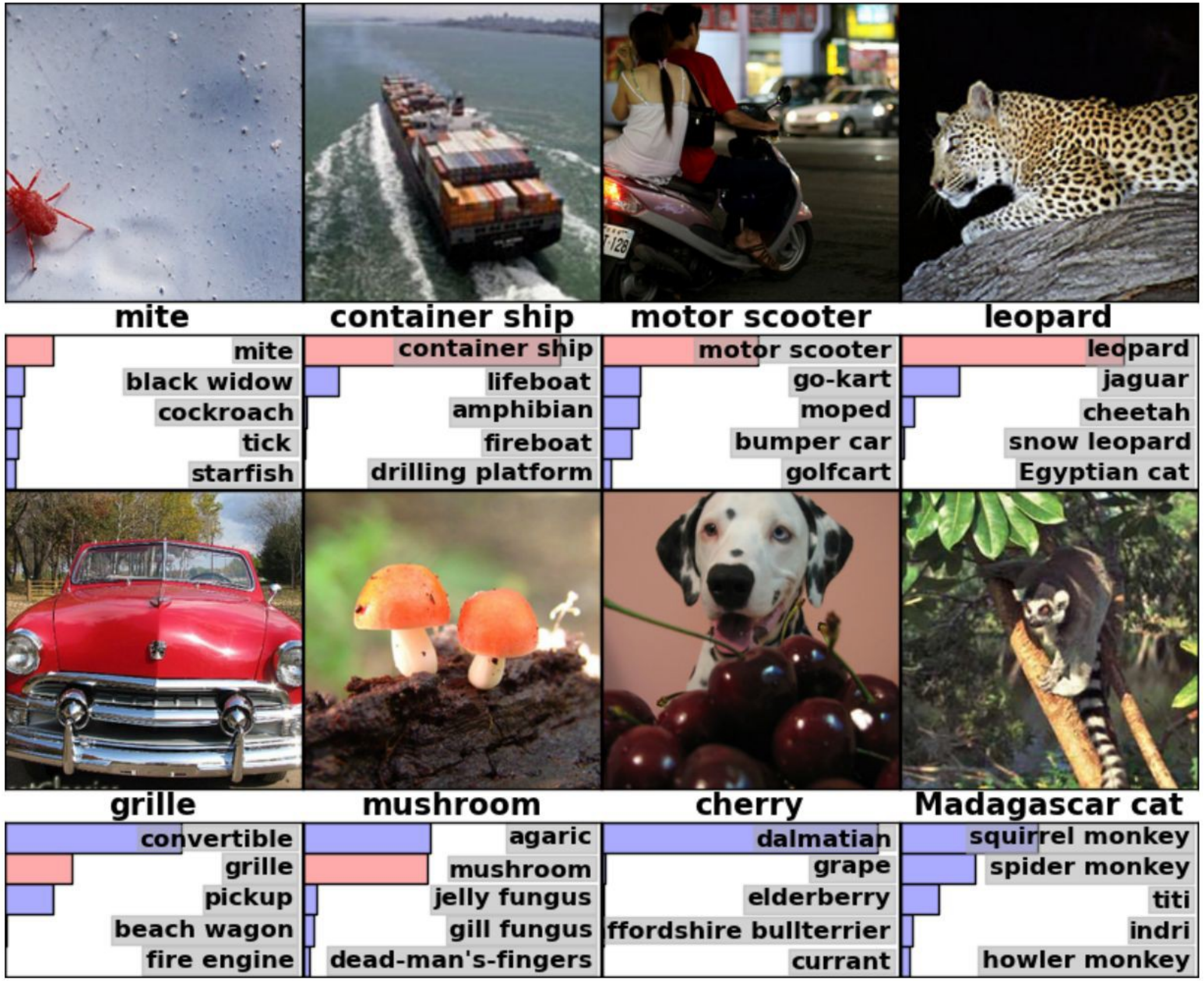
□数据集

●1000 类

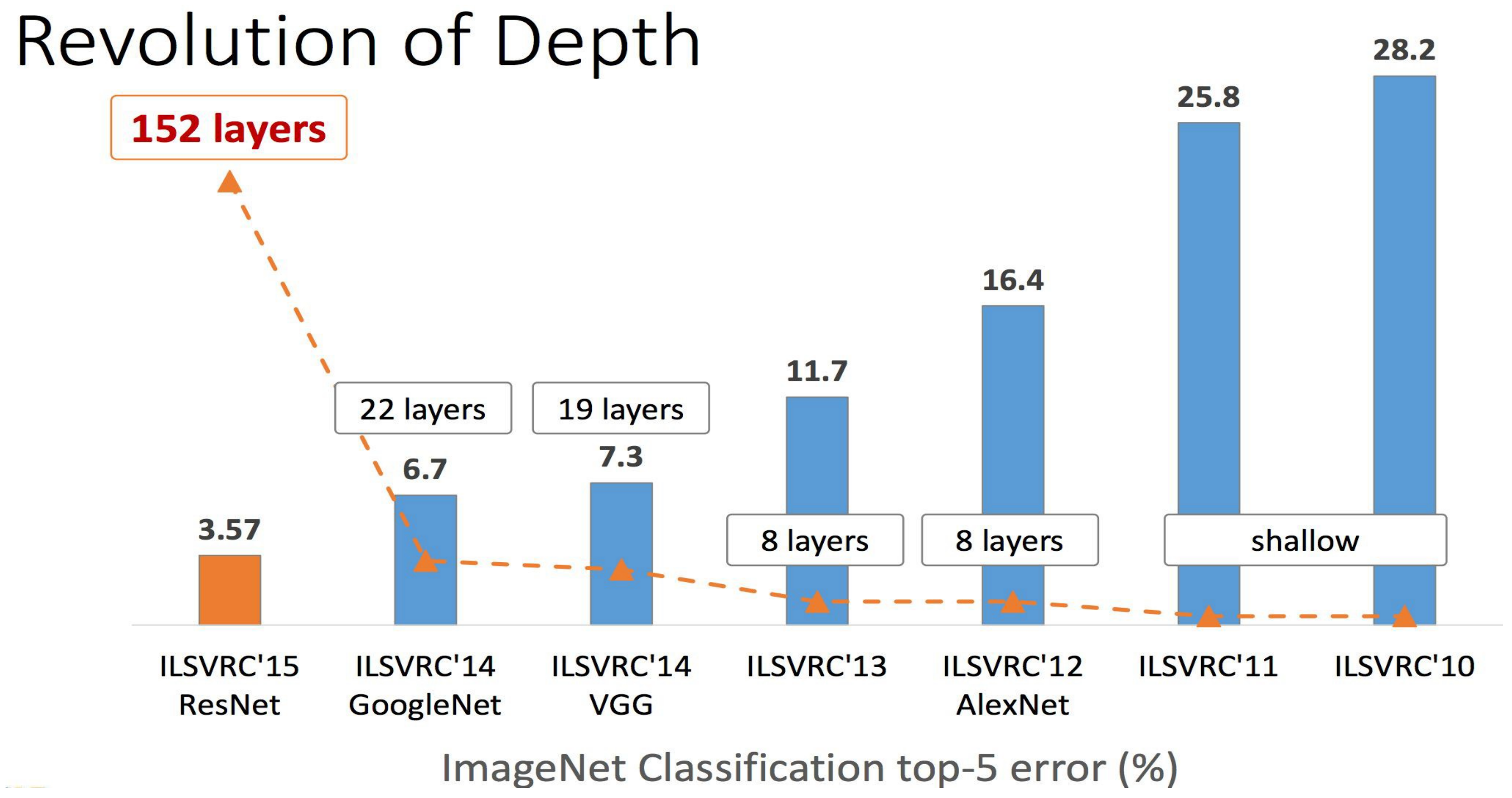
●Images:

- •1.2 M train
- •100k test.

IMAGENET



基于CNN图像分类



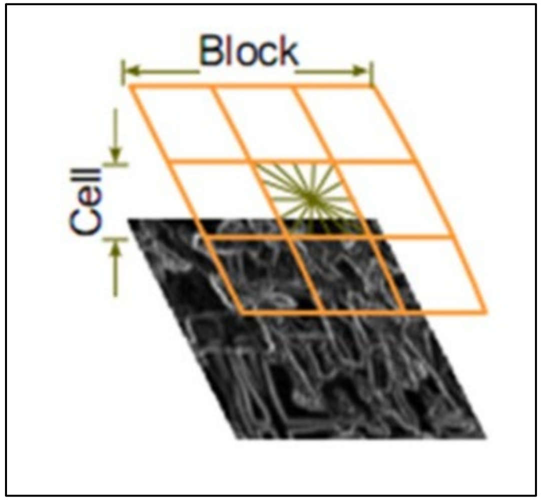
He, Kaiming, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. "Deep residual learning for image recognition." CVPR 2016. [slides]

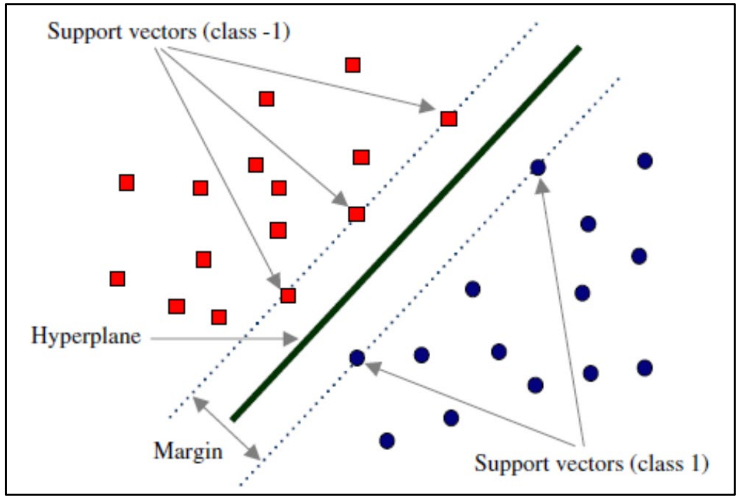
传统图像分类

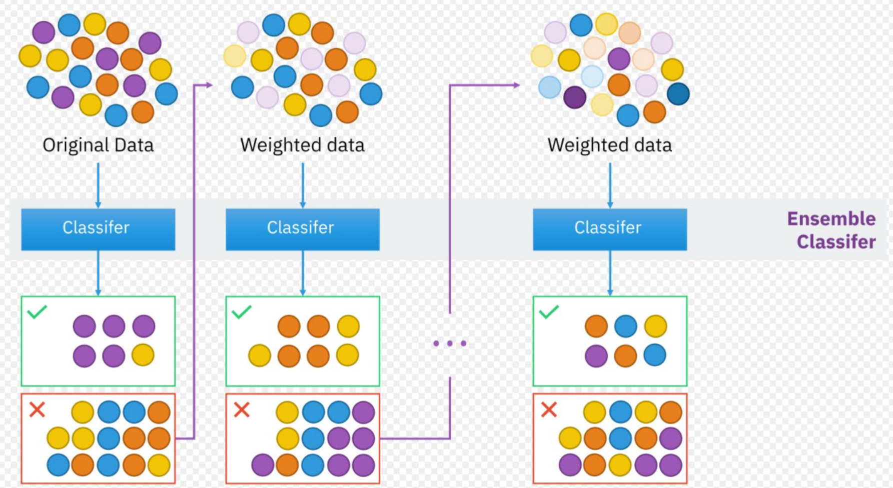
□特征提取 + 分类器

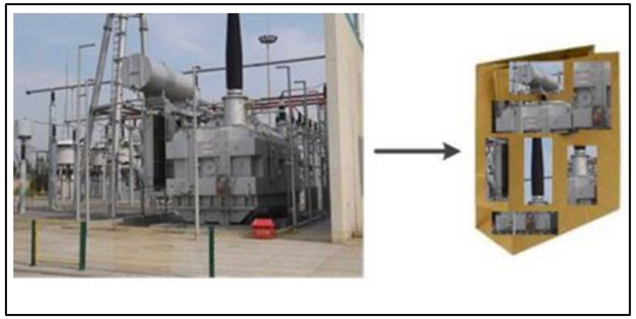
SIFT、HOG、词袋

SVM、boost









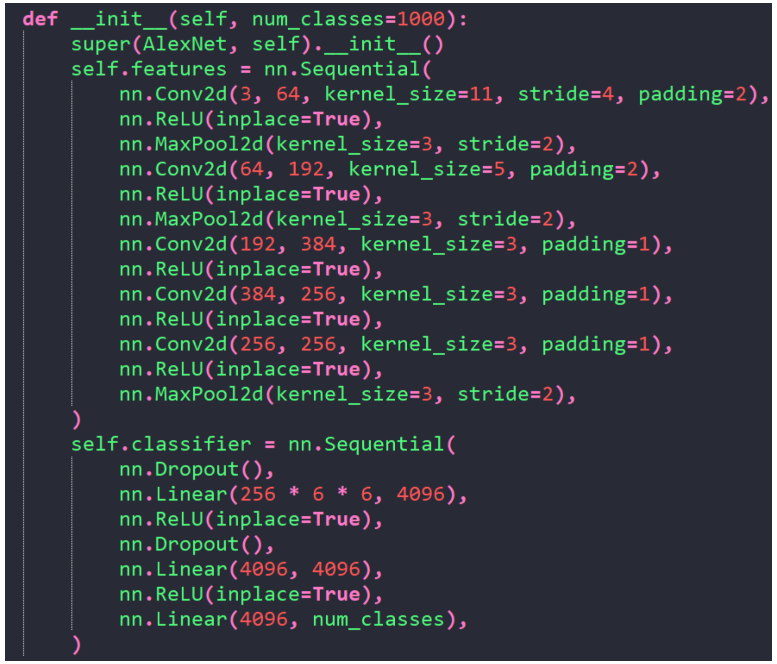
课程内容

- ✓图像分类介绍
- ✓**常见分类网络介绍**
- ✓从图像分类到语义分割

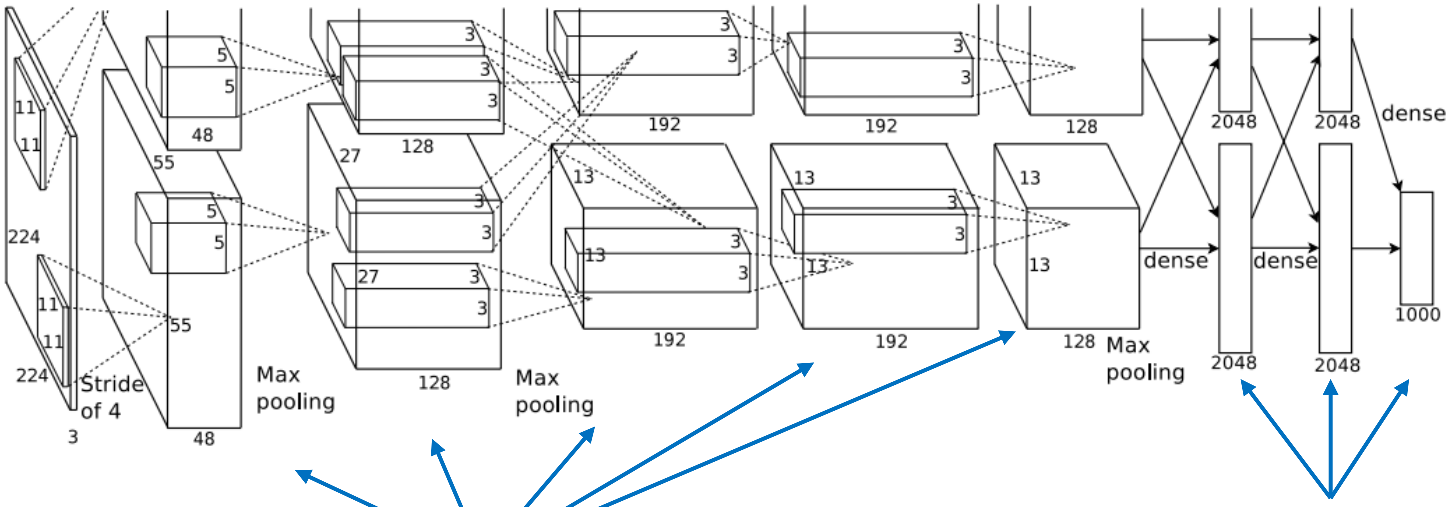
AlexNet

□贡献点

- sigmoid到Relu
- 数据增广
- dropout
- LRN归一化
- GPU训练



```
def __init__(self, num_classes=1000):
    super(AlexNet, self).__init__()
    self.features = nn.Sequential(
        nn.Conv2d(3, 64, kernel_size=11, stride=4, padding=2),
        nn.ReLU(inplace=True),
        nn.MaxPool2d(kernel_size=3, stride=2),
        nn.Conv2d(64, 192, kernel_size=5, padding=2),
        nn.ReLU(inplace=True),
        nn.MaxPool2d(kernel_size=3, stride=2),
        nn.Conv2d(192, 384, kernel_size=3, padding=1),
        nn.ReLU(inplace=True),
        nn.Conv2d(384, 256, kernel_size=3, padding=1),
        nn.ReLU(inplace=True),
        nn.Conv2d(256, 256, kernel_size=3, padding=1),
        nn.ReLU(inplace=True),
        nn.MaxPool2d(kernel_size=3, stride=2),
    )
    self.classifier = nn.Sequential(
        nn.Dropout(),
        nn.Linear(256 * 6 * 6, 4096),
        nn.ReLU(inplace=True),
        nn.Dropout(),
        nn.Linear(4096, 4096),
        nn.ReLU(inplace=True),
        nn.Linear(4096, num_classes),
    )
```



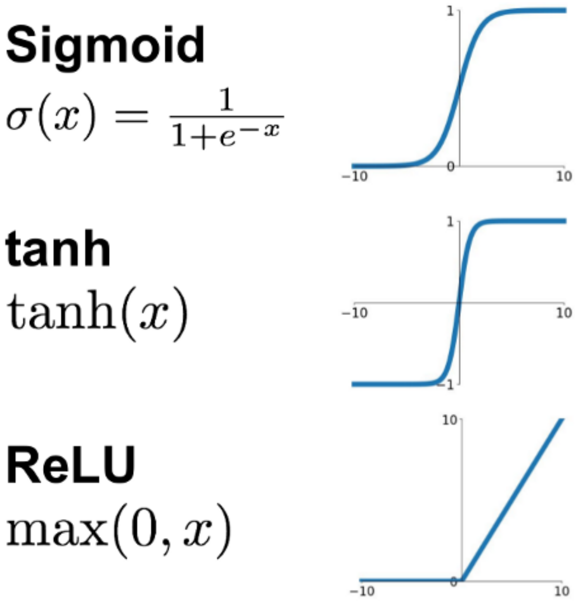
5个卷积层

3个全连接层

AlexNet

□Relu优点

- 无梯度消失
- 计算简单
- 网络稀疏性



Sigmoid

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

tanh

$\tanh(x)$

ReLU

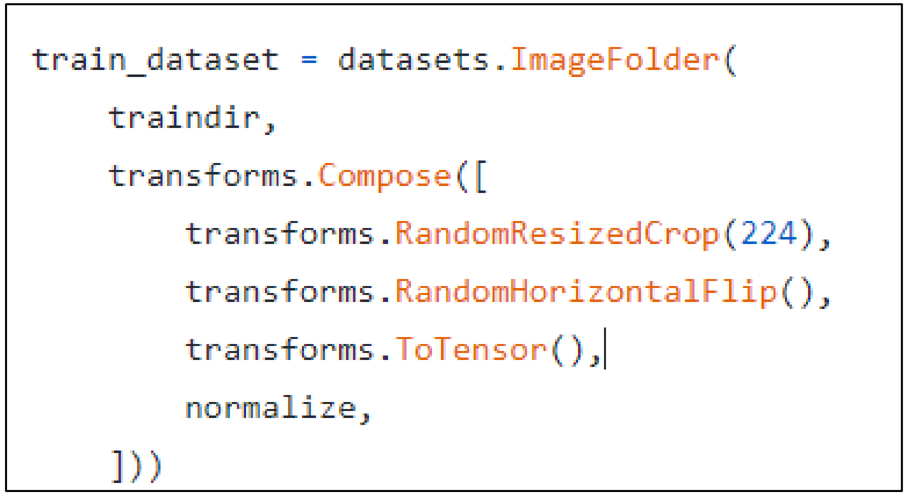
$\max(0, x)$

AlexNet 2012

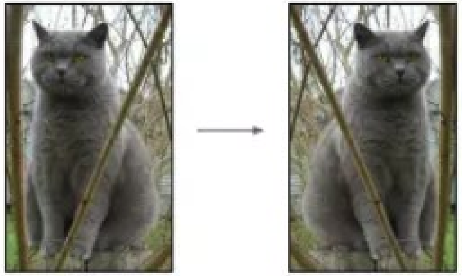
□ 数据增广

- 随机crop
- 镜像
- 光照色彩变换

□代码

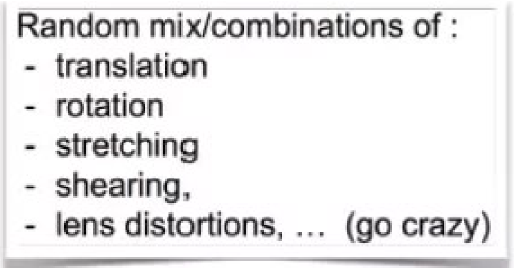


```
train_dataset = datasets.ImageFolder(
    traindir,
    transforms.Compose([
        transforms.RandomResizedCrop(224),
        transforms.RandomHorizontalFlip(),
        transforms.ToTensor(),
        normalize,
    ]))
```



Flip horizontally

Random crops/scales

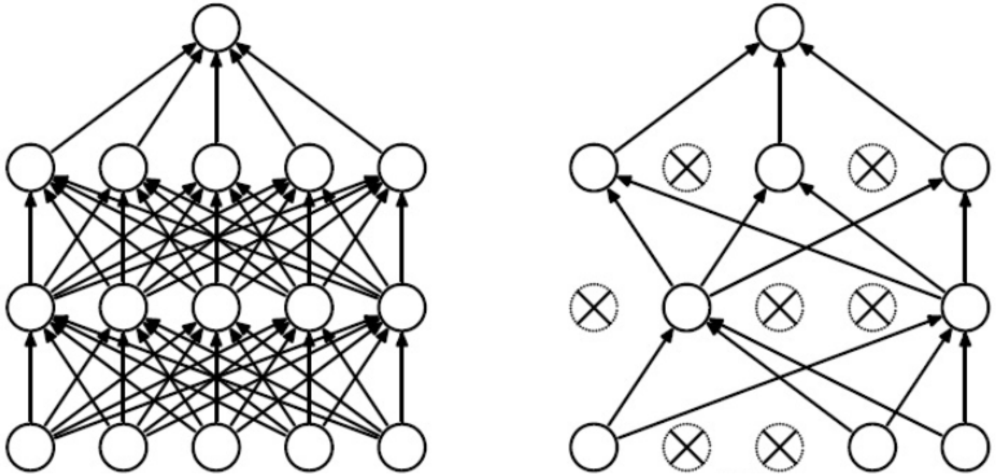


Random mix/combinations of :
- translation
- rotation
- stretching
- shearing,
- lens distortions, … (go crazy)

AlexNet

□ Dropout

- 增加网络泛化性能
- 训练过程中随机抑制节点输出
- 测试中采用所有



□ LRN

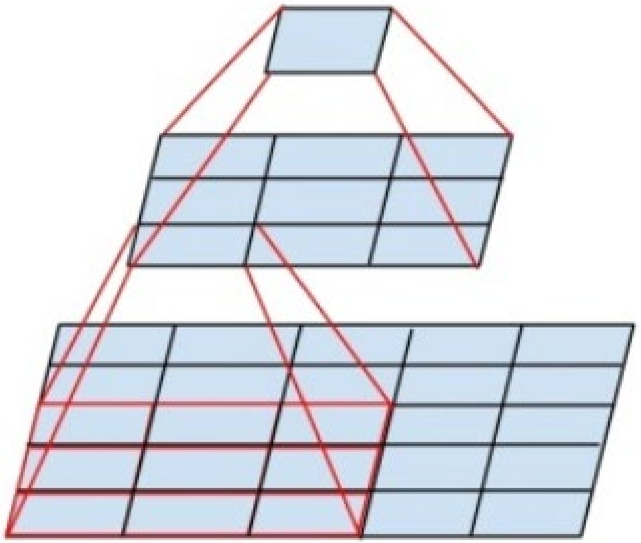
- 提高响应较大的神经元，反之抑制

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

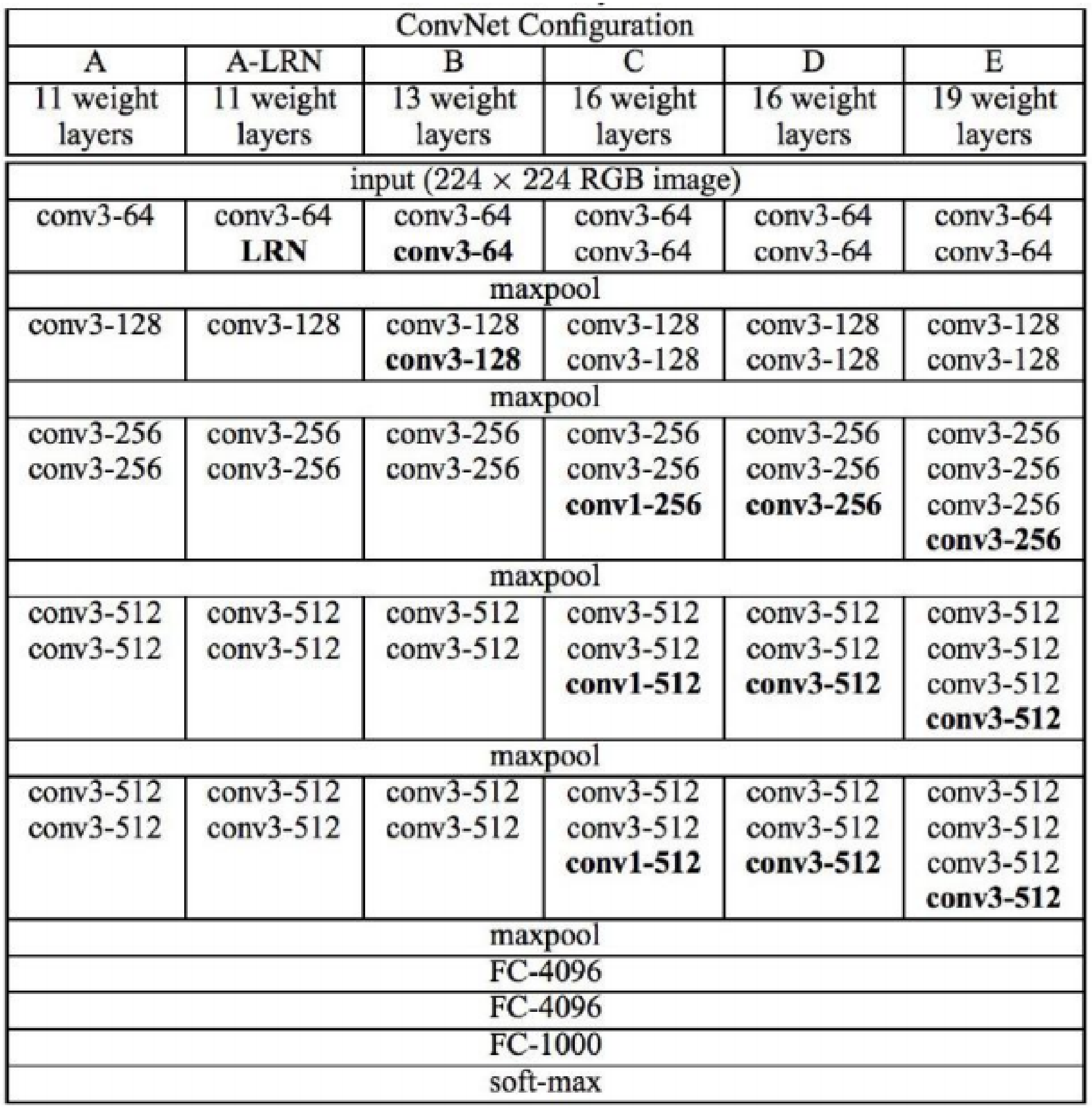
VGGNet 2014

□ 特点

- 级联3*3卷积 替代 大卷积核
- 结构整齐简单



2个3x3卷积核可代替5*5卷积核

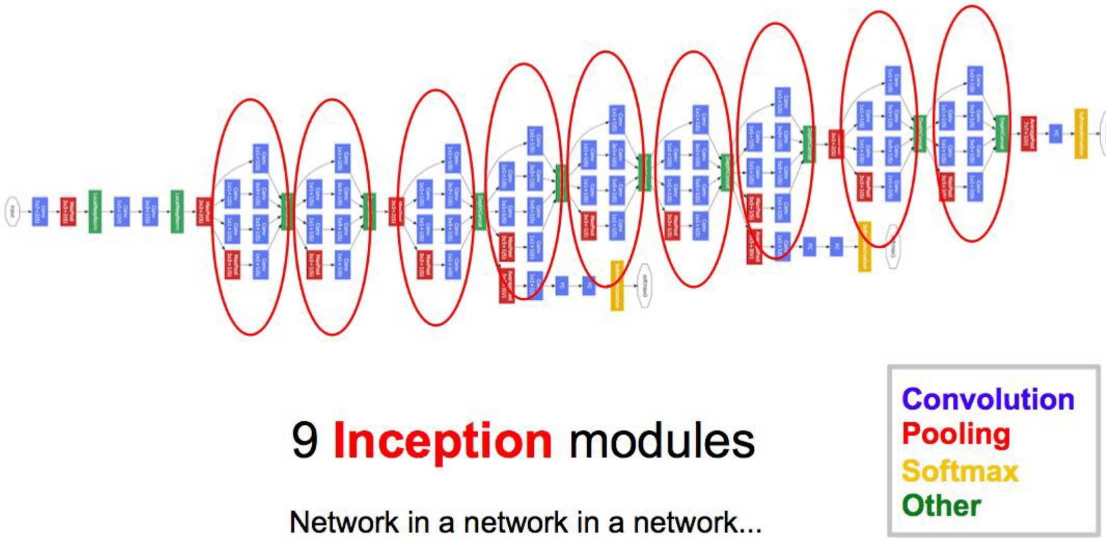


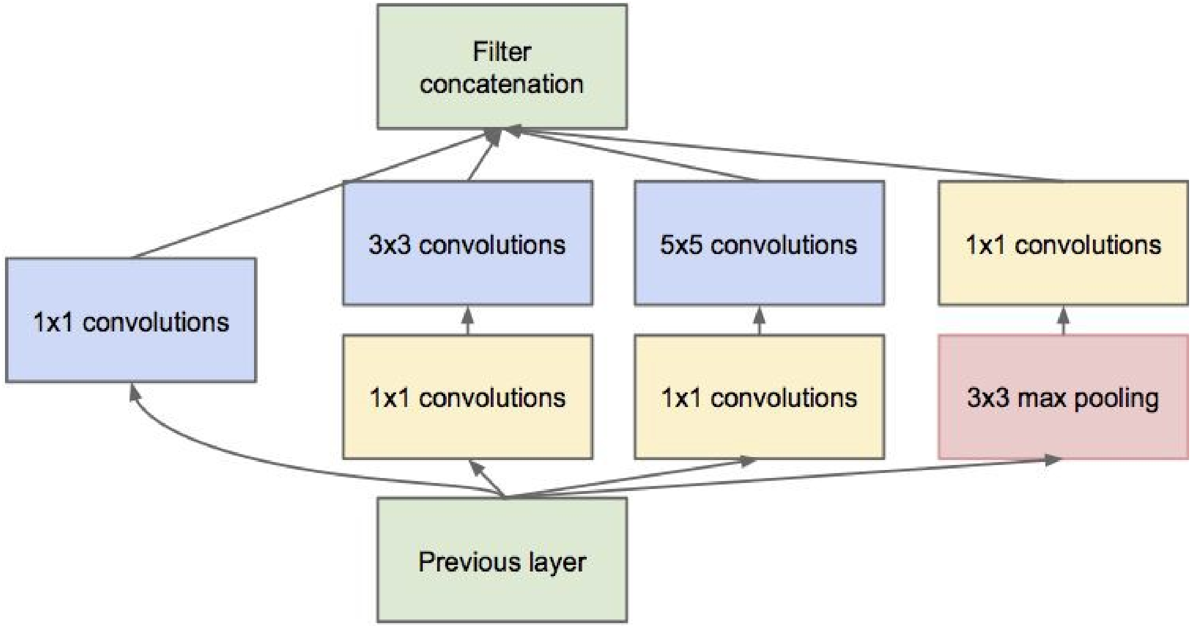
ConvNet Configuration					
A	A-LRN	B	C	D	E
11 weight layers	11 weight layers	13 weight layers	16 weight layers	16 weight layers	19 weight layers
input (224 × 224 RGB image)					
conv3-64	conv3-64 **LRN**	conv3-64 **conv3-64**	conv3-64 conv3-64	conv3-64 conv3-64	conv3-64 conv3-64
maxpool					
conv3-128	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					

GoogleNet 2014

□ 特点

- 级联inception模块
- 增加网络深度、宽度
- 采用1*1卷积降低参数量





ResNet 2015

□ 特点

- 级联残差卷积模块
- 网络深度极大提升
- 网络更容易收敛

$$H(x) = F(x) + x$$

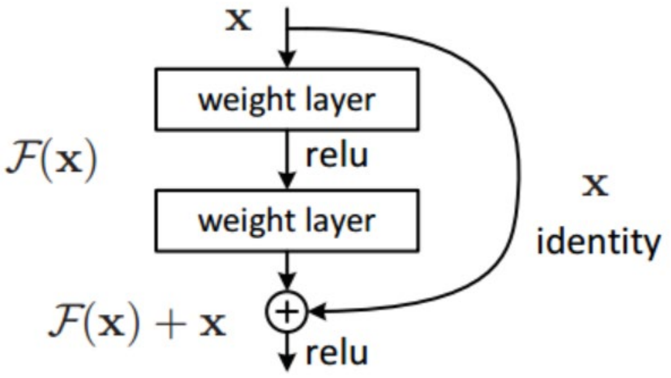
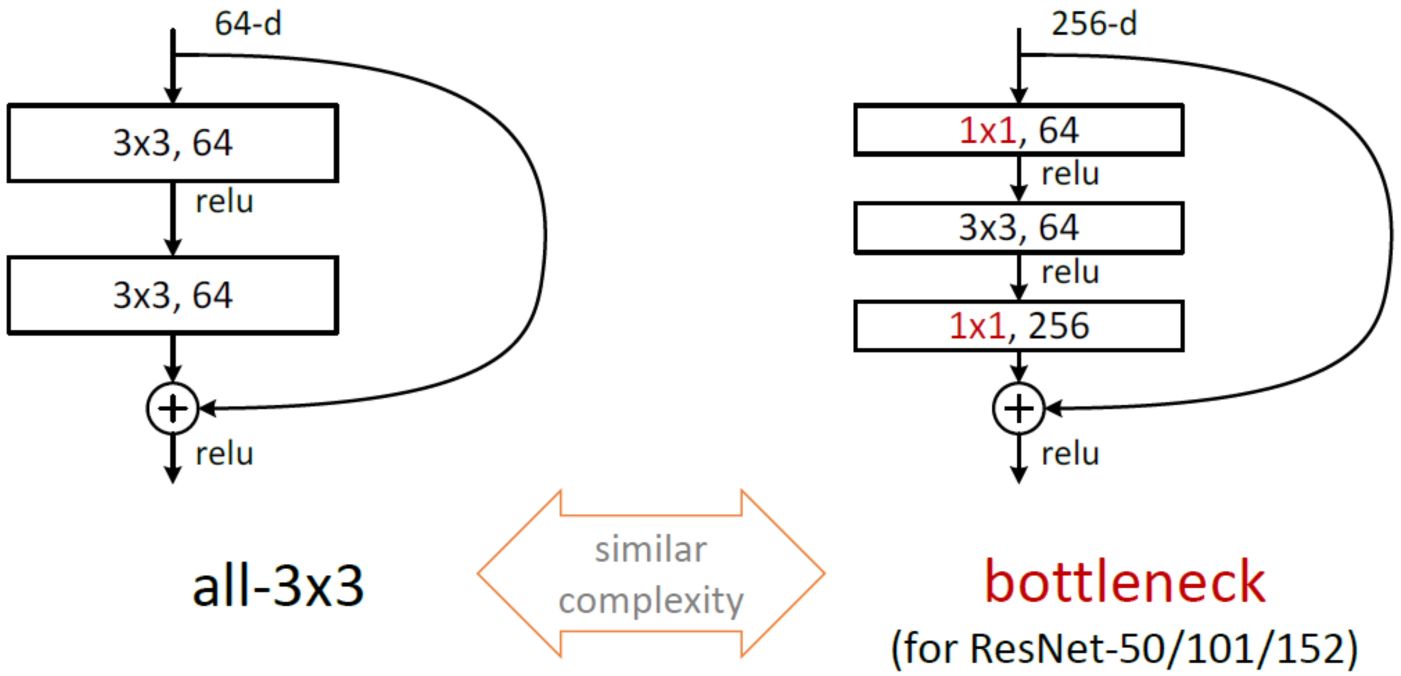

x
weight layer
relu
weight layer
$\mathcal{F}(\mathbf{x})$
x
identity
$\mathcal{F}(\mathbf{x}) + \mathbf{x}$
relu

Figure 2. Residual learning: a building block.

ResNet 2015

□ 特点

- 级联残差卷积模块
- 网络深度极大提升
- 网络更容易收敛



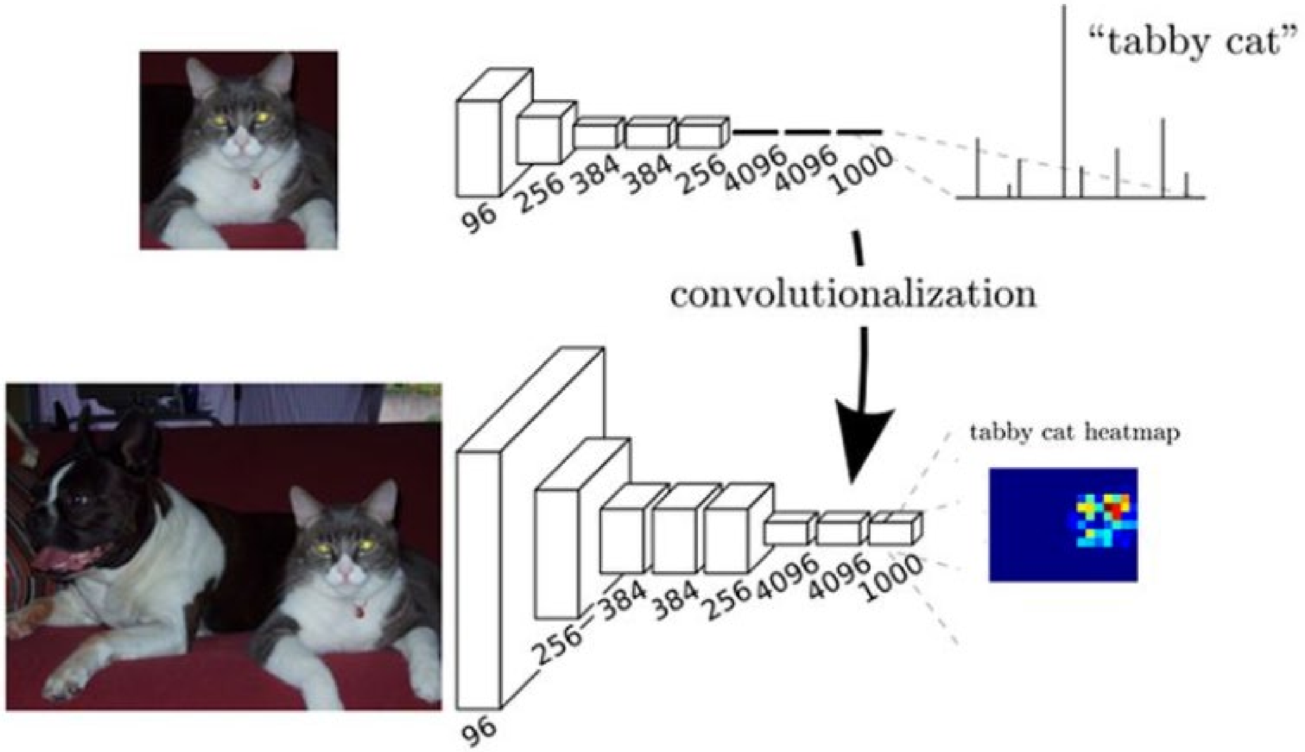
课程内容

- ✓图像分类介绍
- ✓常见分类网络介绍
- ✓**从图像分类到语义分割**

语义分割

□ 从图像分类到像素级分类

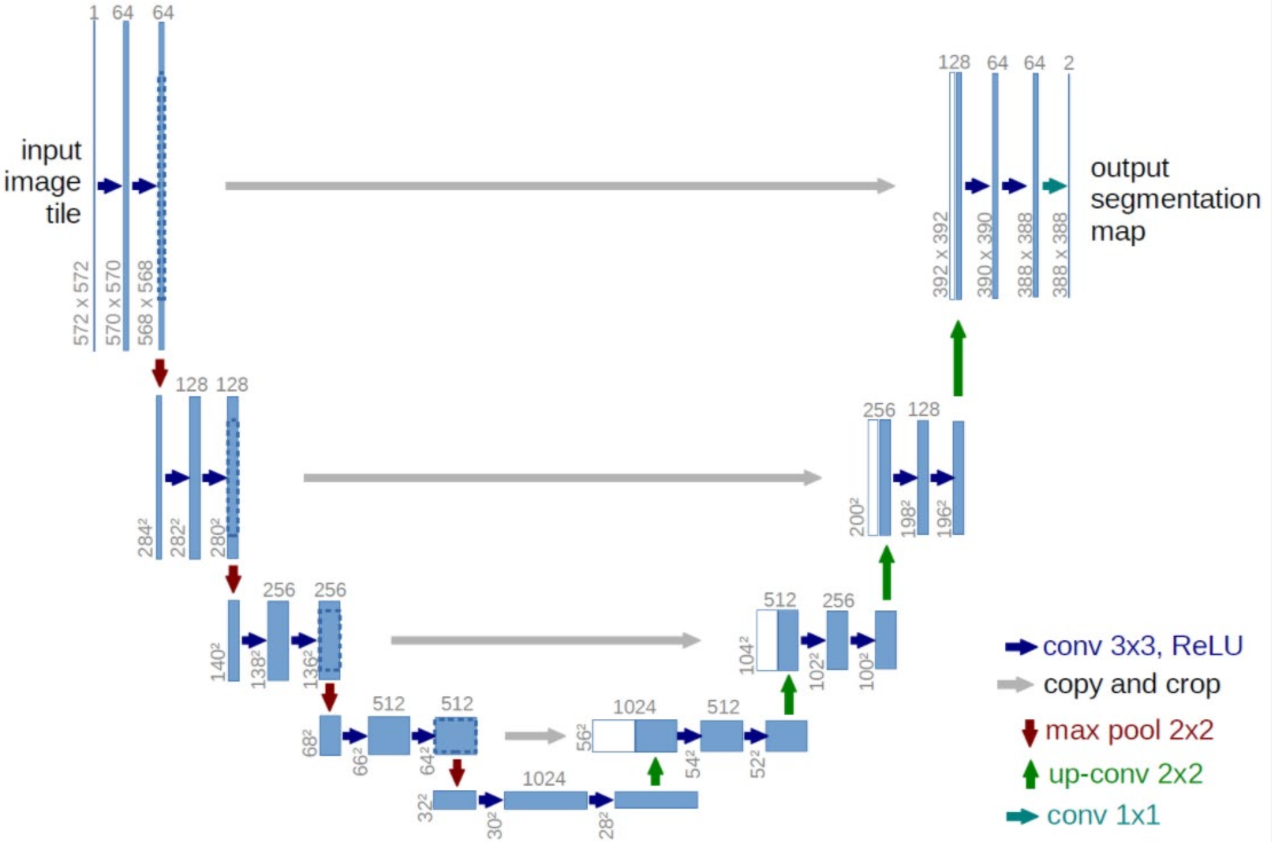
- 全连接层→卷积
- 输入尺寸扩大



语义分割-UNet

□ 从图像分类到像素级分类

- 全连接层→卷积
- 输入尺寸扩大
- 加入上采样→获取更精细化结果



感谢聆听！

Thanks for Listening