人工神经网络（深度学习）

王 瑞 轩

http:// www.isee-ai.cn/~wangruixuan/

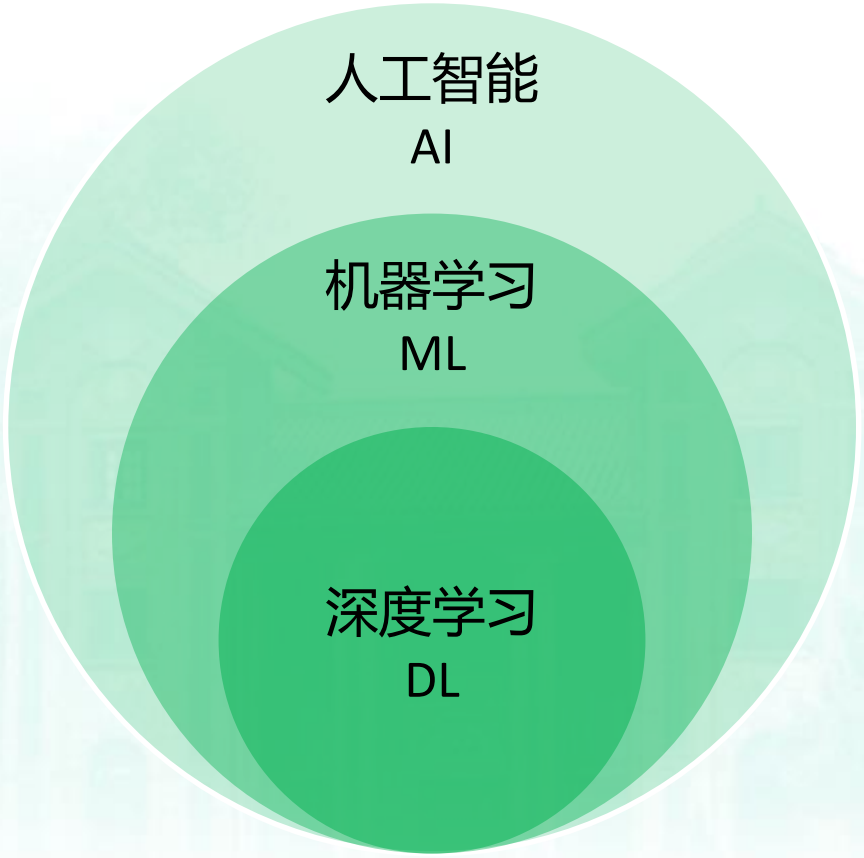
SUN YAT-SEN University

机器智能与先进计算
教育部重点实验室

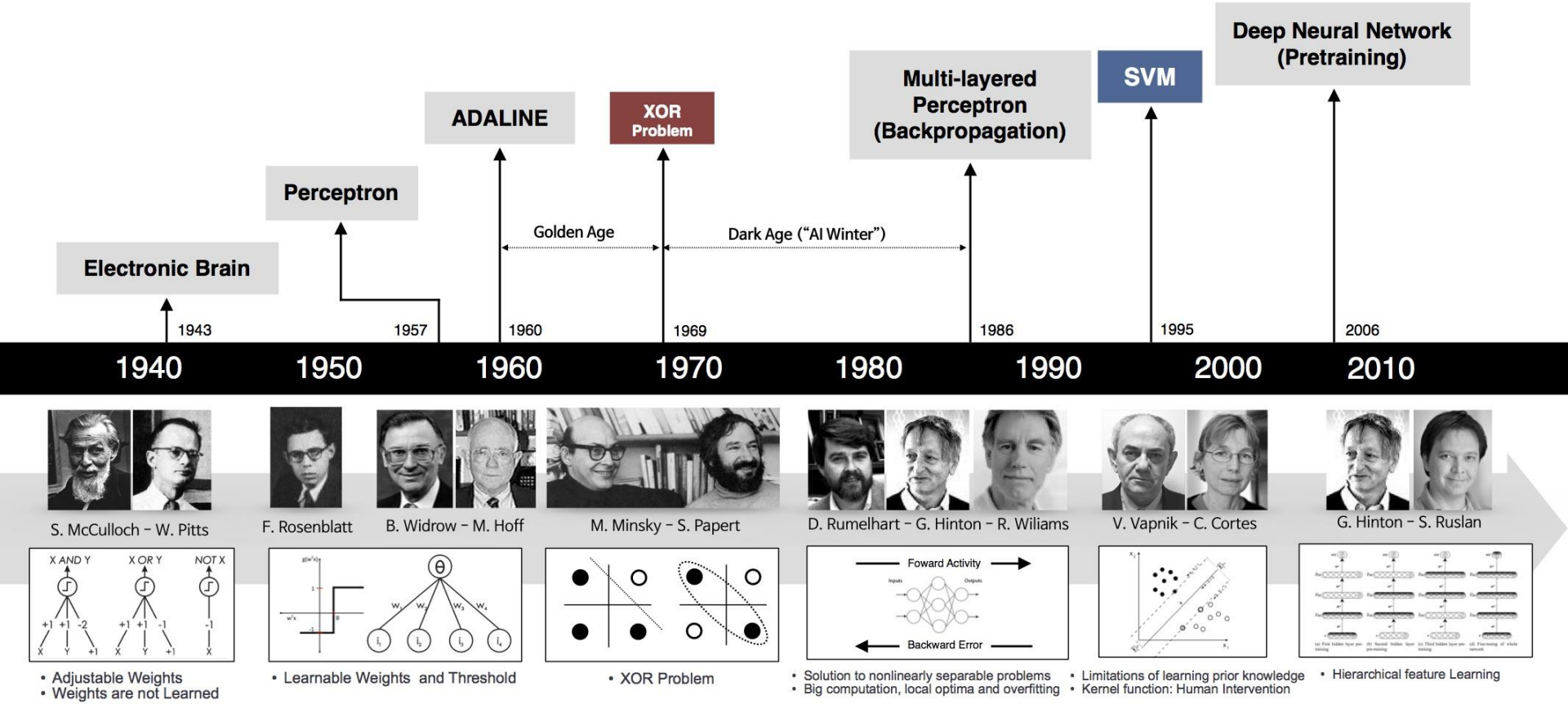
深度学习概览

- What: 揭神秘面纱
- Wow: 赏群模乱舞
- Why: 寻万能之源
- Where: 追研究前沿
- Whoops: 探未解之谜
- While: 评大众观点

揭神秘面纱:深度学习与AI的关系



揭神秘面纱:深度学习发展历史



揭神秘面纱:深度学习研究的坚守者

- 选定人生方向，坚持走下去!
- 要尊重他人的付出与成果!

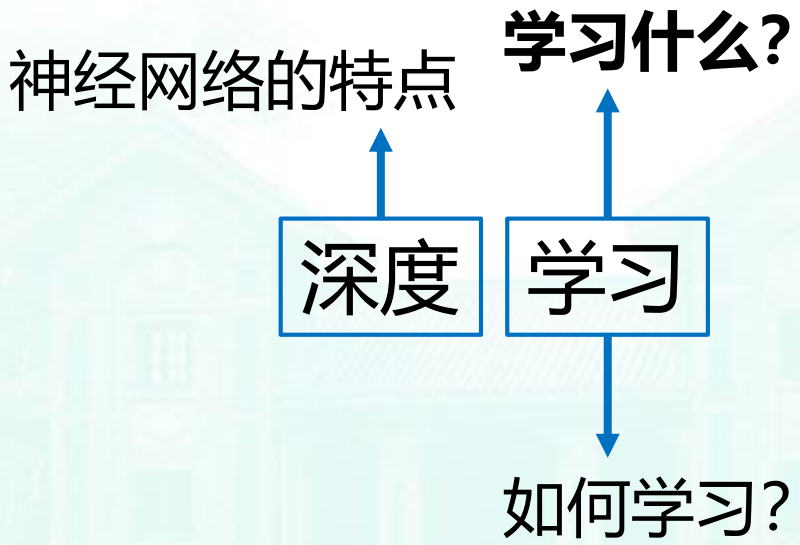
Yoshua Bengio　Geoffrey Hinton　Yann LeCun

Turing Award 2019

Jürgen Schmidhuber

LSTM, 1997

揭神秘面纱



揭神秘面纱:要解决之任务（示例）

- 回归(Regression)任务：“通过身高预测体重”

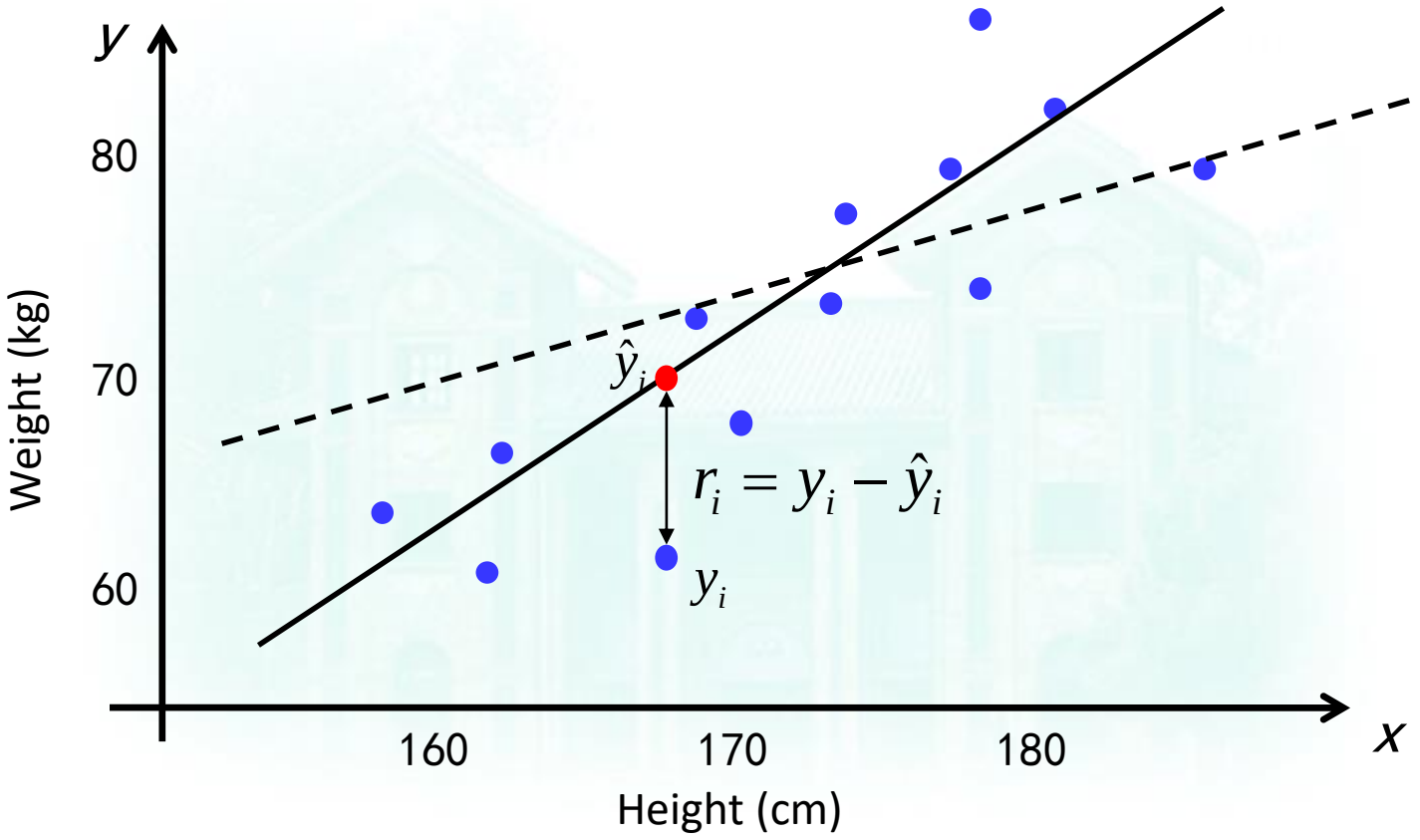
第一步：建模 $y = f(x) = wx + b$

第二步：收集数据 $D = \{(x_i, y_i)\}$

第三步：利用D寻找模型最佳参数 $\boldsymbol{\theta}^* = \{w^*, b^*\}$

揭神秘面纱:要解决之任务（示例）

- 回归(Regression)任务：“通过身高预测体重”



原则：希望预测的体重与观测的体重越接近越好！

揭神秘面纱:要解决之任务（示例）

- 换句话说：希望模型（函数）的预测误差尽量的小

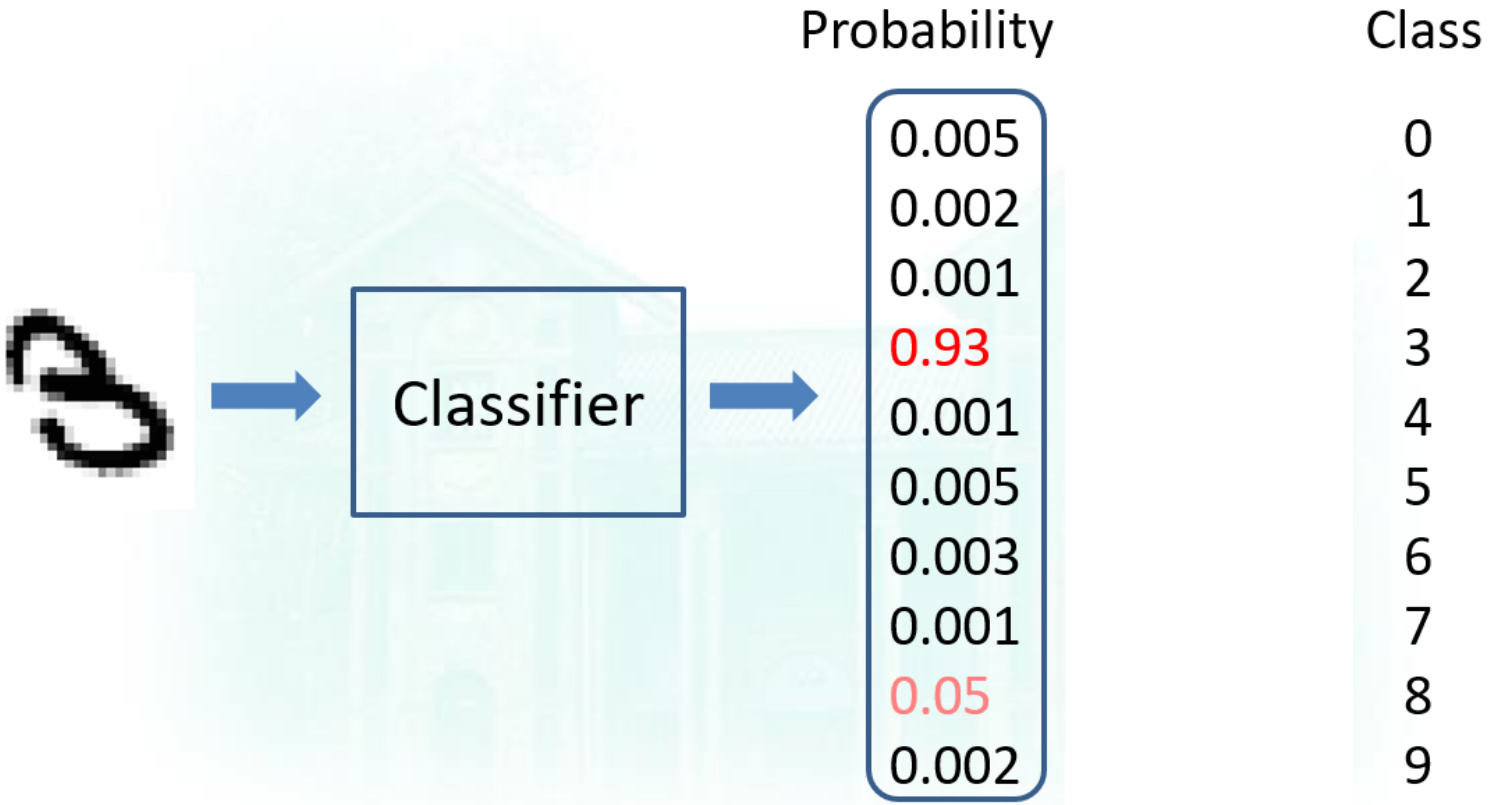
$$\begin{aligned}\min_{\boldsymbol{\theta}} \quad L(\boldsymbol{\theta}) &= \frac{1}{N}\sum_{i=1}^{N} r_i^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}(y_i - wx_i - b)^2\end{aligned}$$

寻找模型（函数）最佳参数 = 最小化损失函数$L(\boldsymbol{\theta})$

学习什么：模型的最佳参数！而最佳模型表示的是...
身高（**输入**）与体重（**输出**）之间的准确**关系**！

揭神秘面纱:要解决之任务（示例）

- 分类（Classification）任务：“手写体数字识别”



给一个输入数据（的表示），模型预测该数据属于每一类的概率

希望模型的输出与理想输出越接近越好！

揭神秘面纱:要解决之任务（示例）

- 模型的输出与理想输出各是一个离散概率分布

模型输出：$\hat{\mathbf{y}}_i = \mathbf{f}(\mathbf{x}_i; \boldsymbol{\theta}) = (\hat{y}_{i1}, \hat{y}_{i2}, \ldots, \hat{y}_{iK})$

理想输出：$\mathbf{y}_i = (y_{i1}, \ldots, y_{iK}) = (0, \ldots, 1, \ldots, 0)$

- 如何测量两个概率分布的差别？Cross-entropy loss!

$$
\begin{aligned}
l(\mathbf{y}_i, \mathbf{f}(\mathbf{x}_i; \boldsymbol{\theta})) &= \sum_{k=1}^{K} y_{ik} \log \frac{1}{\hat{y}_{ik}} \\
&= -\sum_{k=1}^{K} y_{ik} \log \hat{y}_{ik}
\end{aligned}
$$

揭神秘面纱:要解决之任务（示例）

- 与回归任务的目标类似：

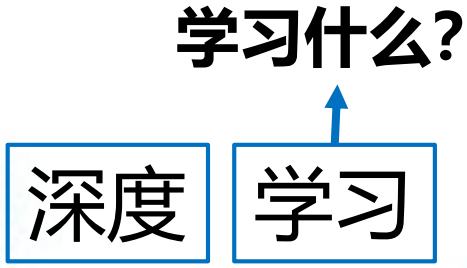
$$\min_{\boldsymbol{\theta}} \quad L(\boldsymbol{\theta}) \quad = \quad \frac{1}{N}\sum_{i=1}^{N} l(\mathbf{y}_i, \mathbf{f}(\mathbf{x}_i; \boldsymbol{\theta}))$$

$$= \quad -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik} \log \hat{y}_{ik}$$

寻找模型（函数）最佳参数 = 最小化损失函数$L(\boldsymbol{\theta})$

学习什么：模型的最佳参数！而最佳模型表示的是...
图象（**输入**）与数字类别（**输出**）之间的准确**关系**！

揭神秘面纱：学习什么



任一任务：

学习的是系统输入与输出的关系！

如果系统由一个数学函数表示，
学习的是函数（数学模型）的最佳参数！

揭神秘面纱：模型/函数

如何设计模型（函数）？

$$\min_{\boldsymbol{\theta}} \quad L(\boldsymbol{\theta}) = \frac{1}{N}\sum_{i=1}^{N} l(\mathbf{y}_i, \mathbf{f}(\mathbf{x}_i;\boldsymbol{\theta}))$$

人工神经网络！

$$= -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik}\log \hat{y}_{ik}$$

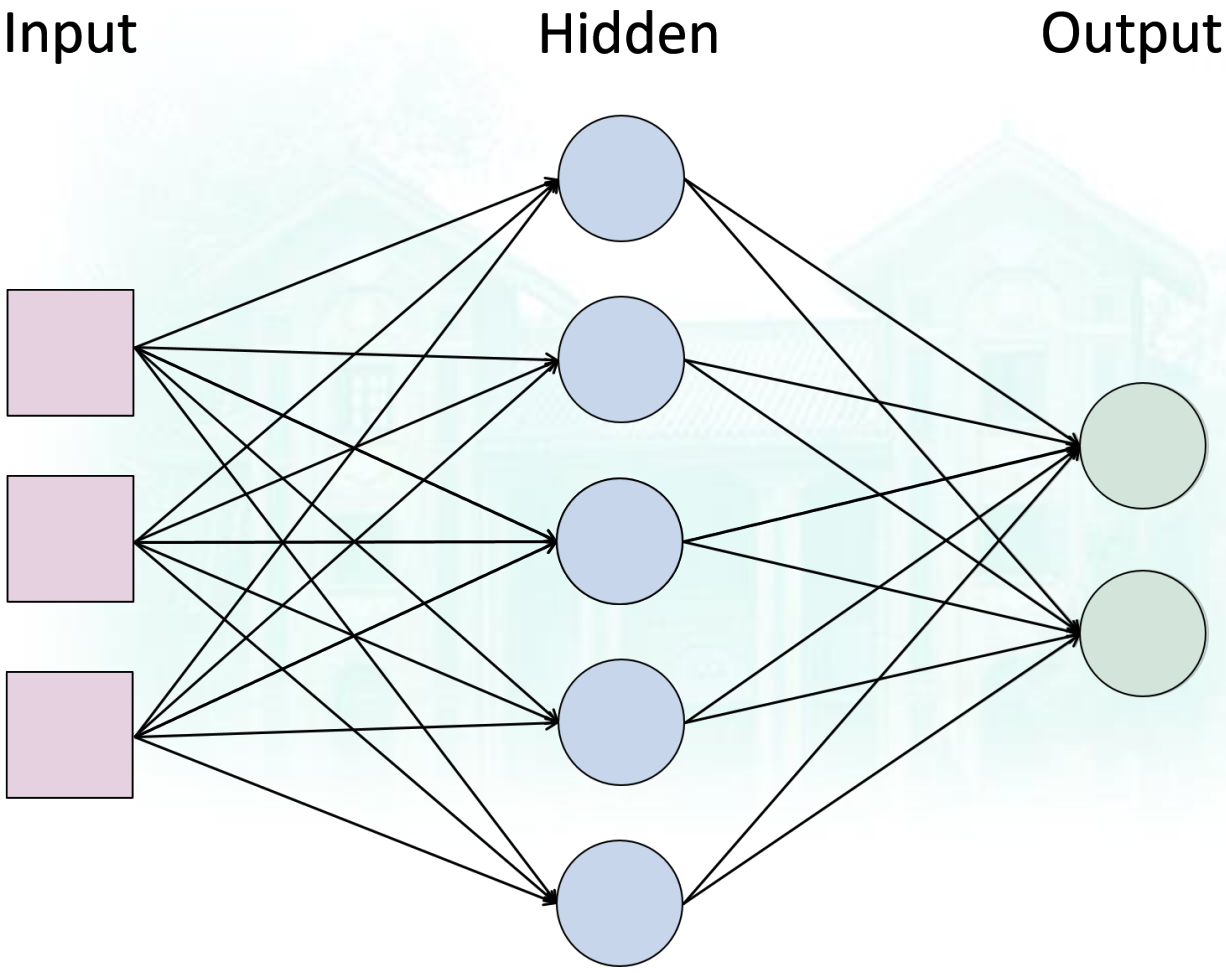
揭神秘面纱：为什么用神经网络模型？

万能近似定理（universal approximation theorem）(Hornik et al., 1989;Cybenko, 1989)：一个两层人工神经网络如果具有足够多的隐藏单元，它可以以任意的精度来近似任何一个函数（即可以表示任意的复杂输入-输出关系）。

更多层的神经网络比两层神经网络能够以更少的神经元表示具有同样复杂度的函数，并且性能更好！

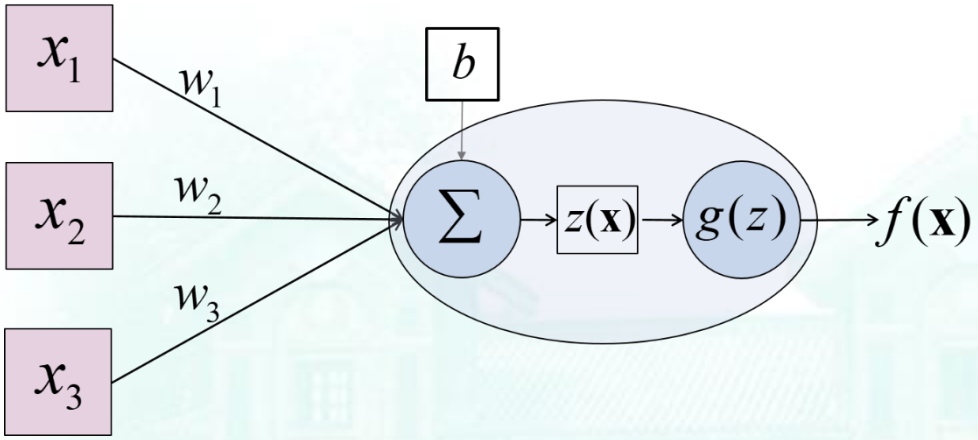
揭神秘面纱:传统全连接神经网络

- 两层全连接神经网络结构



揭神秘面纱:传统全连接神经网络

- 全连接神经网络中单个神经元:



$$z(\mathbf{x}) = \sum_i w_i x_i + b = \mathbf{w}^T\mathbf{x} + b$$

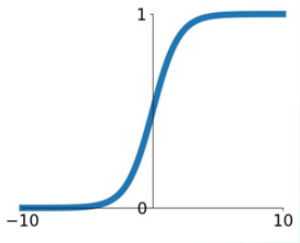
$$f(\mathbf{x}) = g(z) = g(\mathbf{w}^T\mathbf{x} + b)$$

$g(z)$: 激活函数

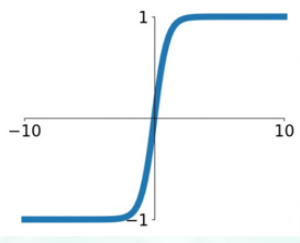
$\boldsymbol{\theta} = \{\mathbf{w}, b\}$: 模型参数

揭神秘面纱:传统全连接神经网络

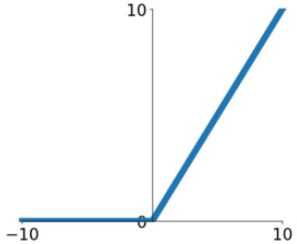
- 常见激活函数



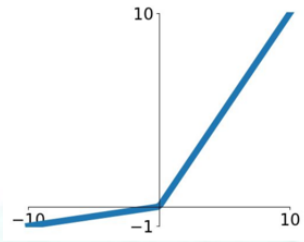
Sigmoid



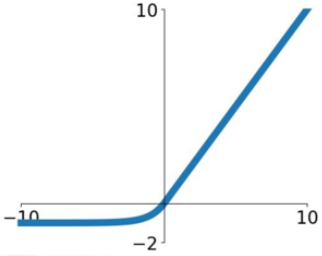
tanh



ReLU



Leaky ReLU



ELU

$$g(z)=\frac{1}{1+e^{-z}}$$

$$\tanh(z)$$

$$\max(0,z)$$

$$\max(\alpha z,z)$$

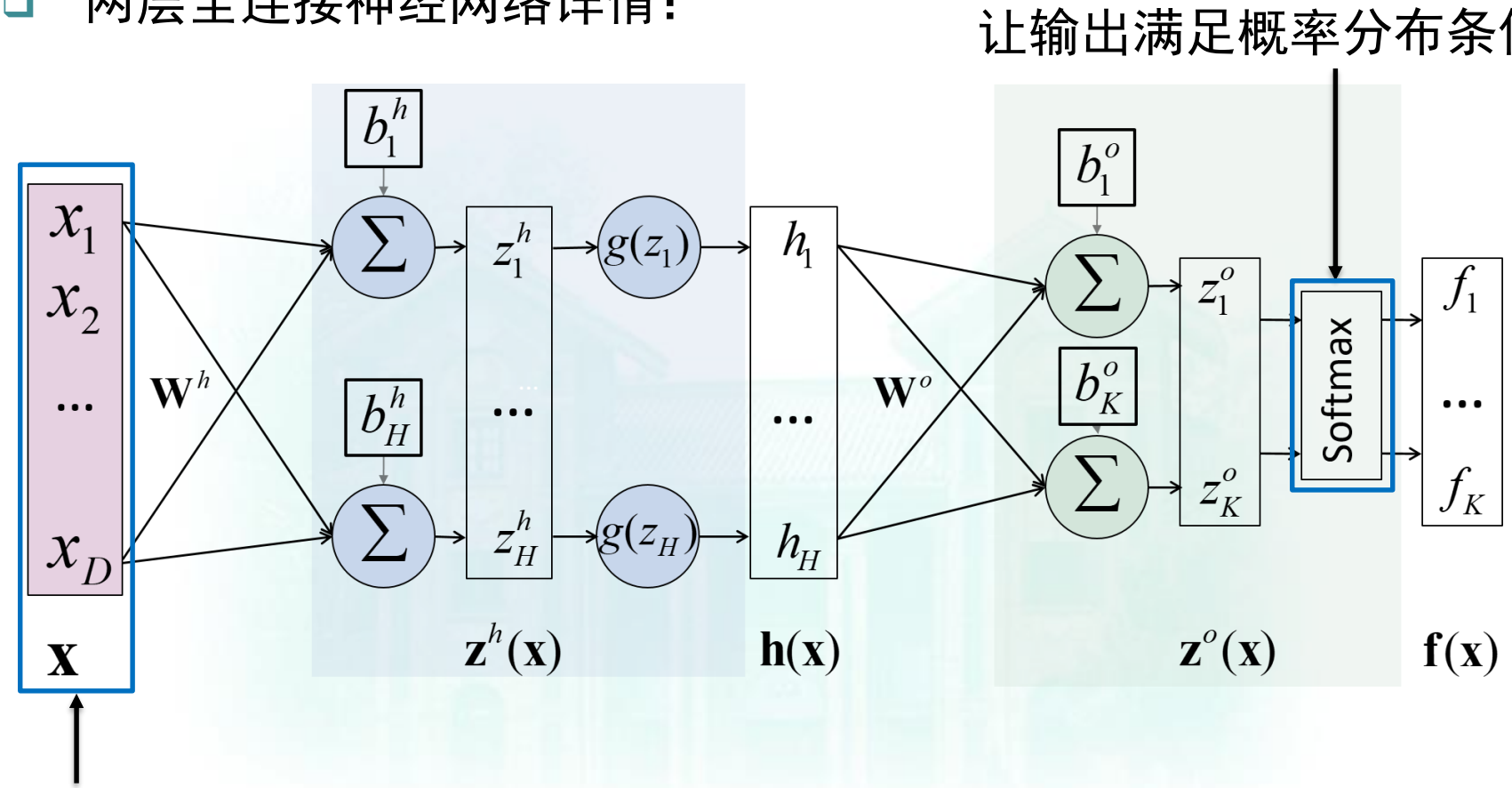
$$\begin{cases} z & z\geq 0 \\ \alpha(e^{z}-1) & z<0 \end{cases}$$

为什么需要激活函数？

让神经网络成为非线性函数，进而可以
表示输入-输出间的潜在复杂非线性关系！

揭神秘面纱：传统全连接神经网络

- 两层全连接神经网络详情：



数据表示为向量

上图中模型的参数包括？

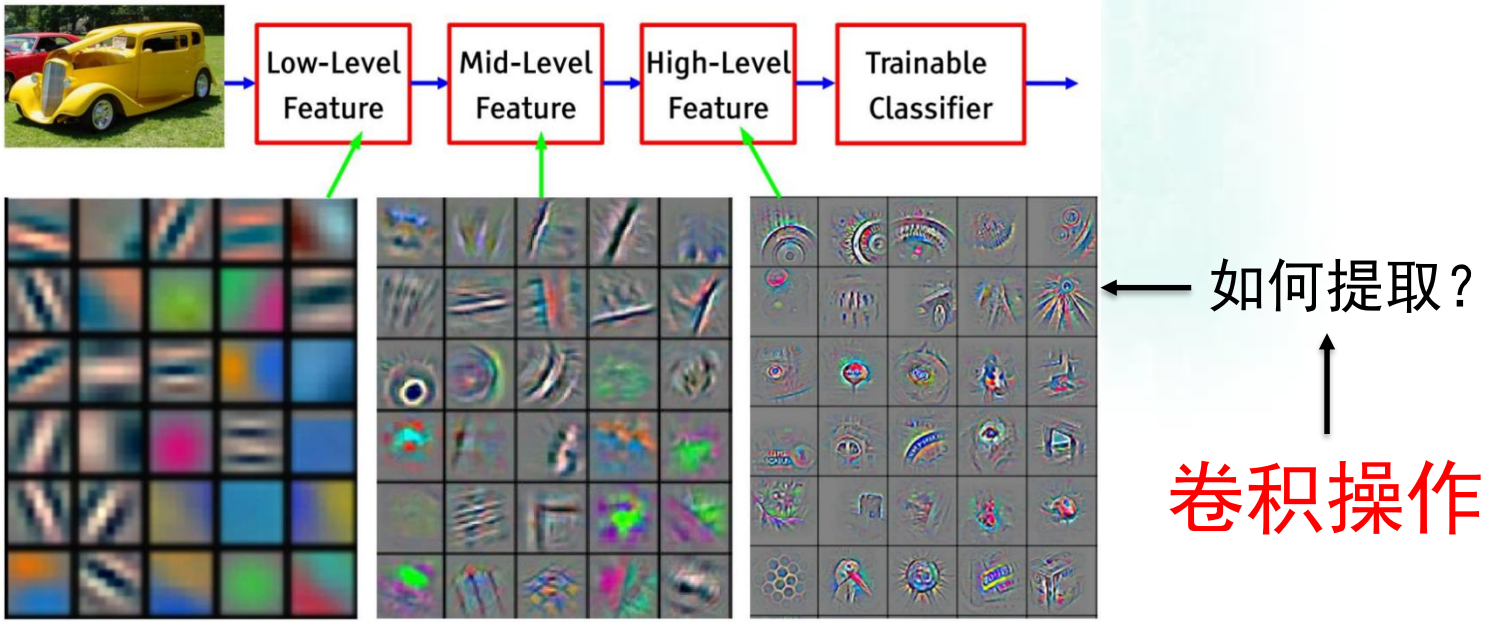
揭神秘面纱:数据的表示

- 如何将数据表示为向量，尤其是图像/视频/文本等数据？

- 以前：设计算法从数据中提取特征，比如SIFT+Bag of Words

  缺点–人工设计、与具体任务无关、可能漏掉有用特征

- 是否可以自动学习抽取与特定任务相关的特征？



揭神秘面纱:卷积

卷积是函数与函数之间的一种操作，结果为另一个函数

$$(f * g)(t) \equiv \int_{-\infty}^{\infty} f(\tau) g(t-\tau) d\tau$$

离散函数之间的卷积：

$$(f * g)[i] \equiv \sum_{m=-\infty}^{\infty} f[m]\, g[i-m] = \sum_{m=-\infty}^{\infty} f[i-m]\, g[m]$$

当函数$g(m)$只在$m \in [-M, M]$时非零：

$$(f * g)[i] = \sum_{m=-M}^{M} f[i-m]\, g[m]$$

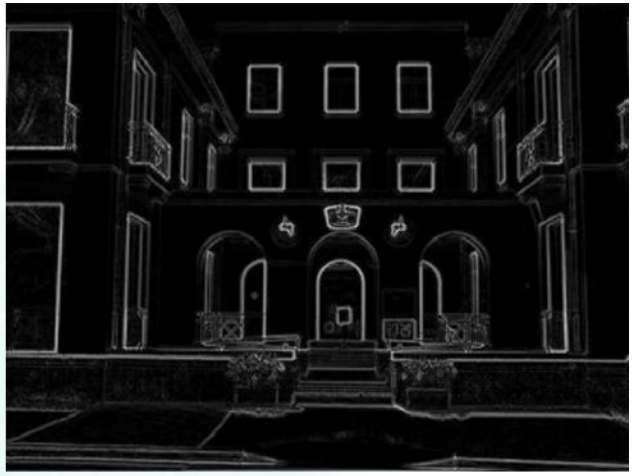
卷积可以扩展到二维函数：

$$(f * g)[i,j] = \sum_{m=-M}^{M} \sum_{n=-N}^{N} f[i-m, j-n]\, g[m,n]$$

揭神秘面纱:卷积

- 传统的边缘提取就是一个具体的卷积操作

$$* \begin{bmatrix} 1 & 0 & -1 \\ 0 & 0 & 0 \\ -1 & 0 & 1 \end{bmatrix} =$$



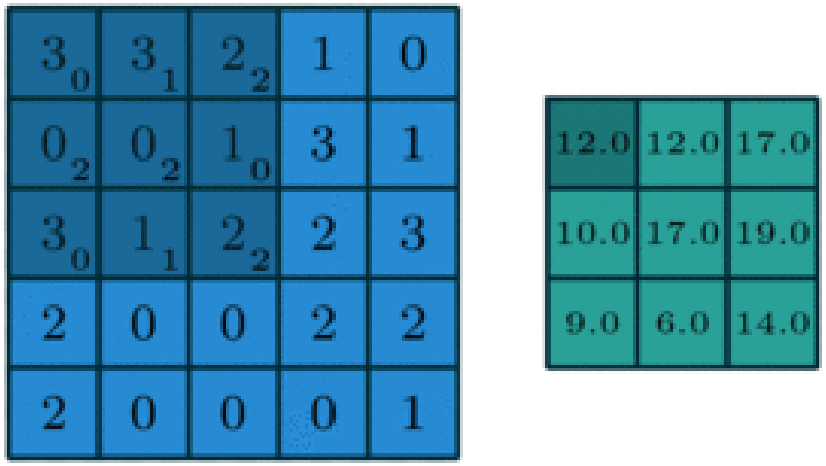
$f[i,j]$ ↑ 输入

$g[m,n]$ ↑ 卷积核(Kernel)

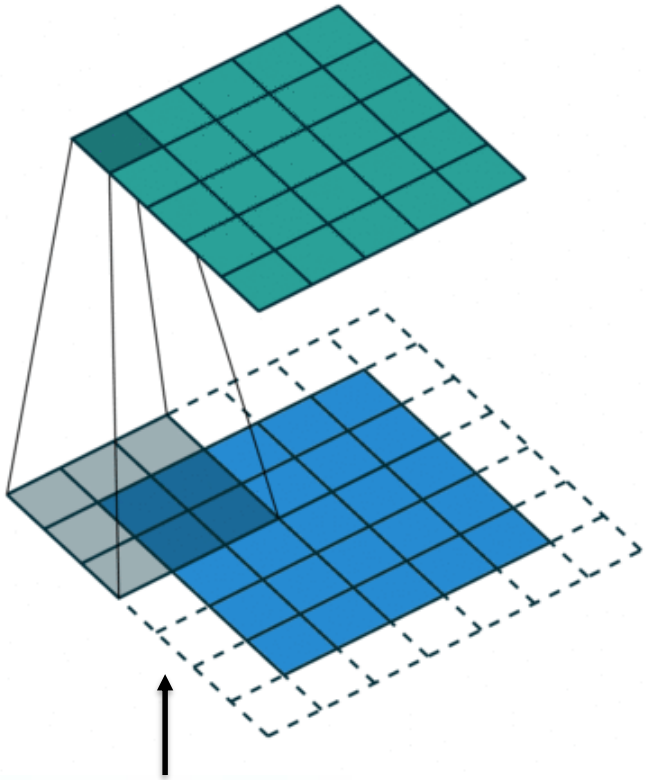
$(f*g)[i,j]$ ↑ 特征图(Feature map)

揭神秘面纱:卷积

- 卷积具体操作（例子）：大小为5x5的输入、3x3的卷积核



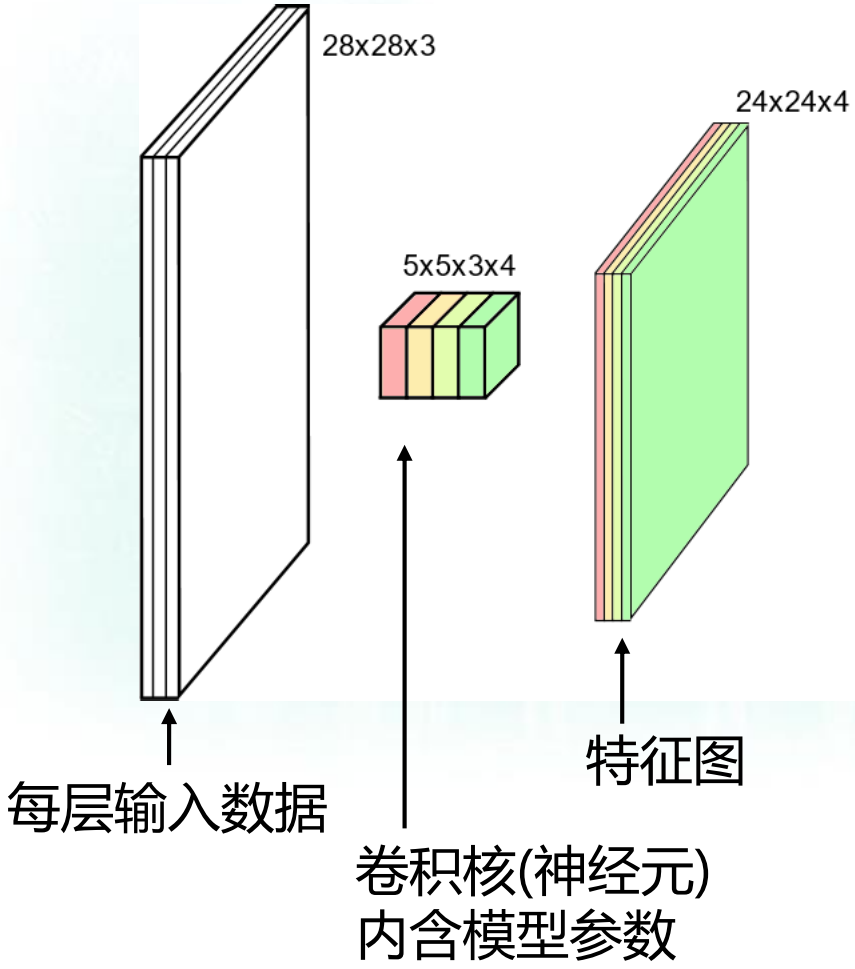
输出（特征图）小于输入的空间尺寸



Padding:对输入边缘补零，使卷积结果与输入尺寸一样

揭神秘面纱：卷积层

- 一个卷积层包含多个卷积操作，输出多个特征图
- 卷积层输出（经过激活函数等）作为下一卷积层的输入



Questions:

每个卷积核的维度？

卷积核通道数与输入通道数（Channels）的关系？

揭神秘面纱:卷积提取高层特征

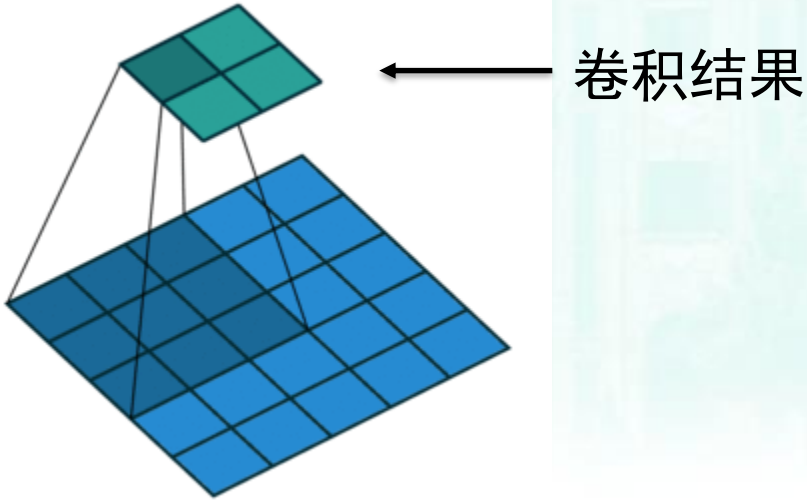
- 如何有效提取高层语义特征？
  - 语义特征对应更大图像区域（大尺寸卷积核？）
  - 语义特征需要忽略细节（大尺寸卷积核一般对细节敏感）

- 如何在不增加卷积核尺寸情况下提取语义特征？

揭神秘面纱：卷积中Stride

- 为了提取语义特征，需要减小（低层）特征图尺寸，然后与（高层）卷积核做卷积操作

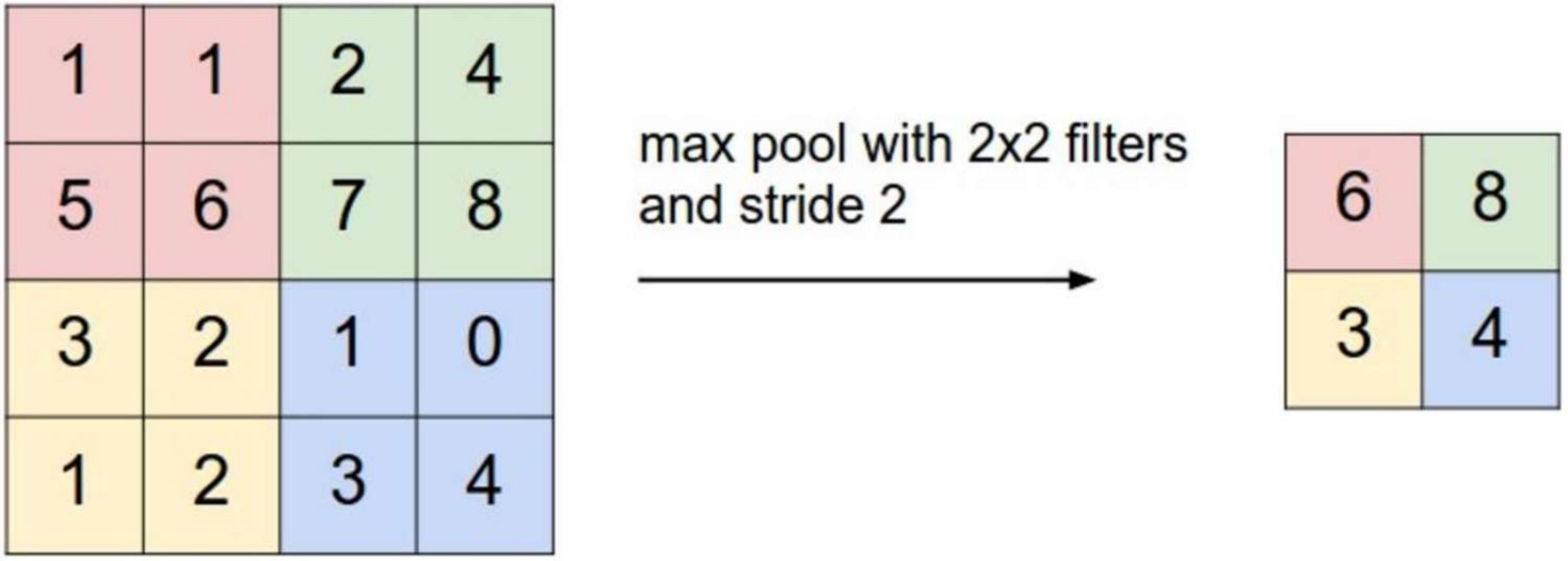
- Stride：卷积操作中卷积核相对于特征图每次移动距离
- 可有效减小特征图的空间尺寸



Stride=2 时的卷积操作

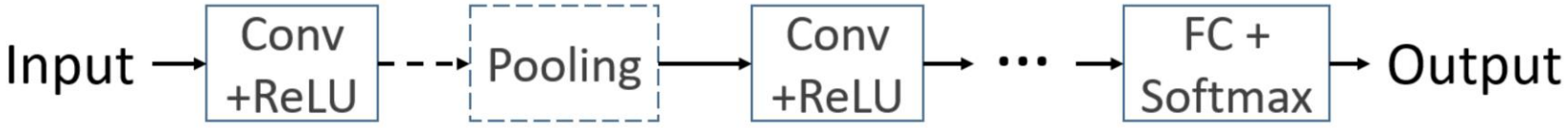
揭神秘面纱:卷积后池化

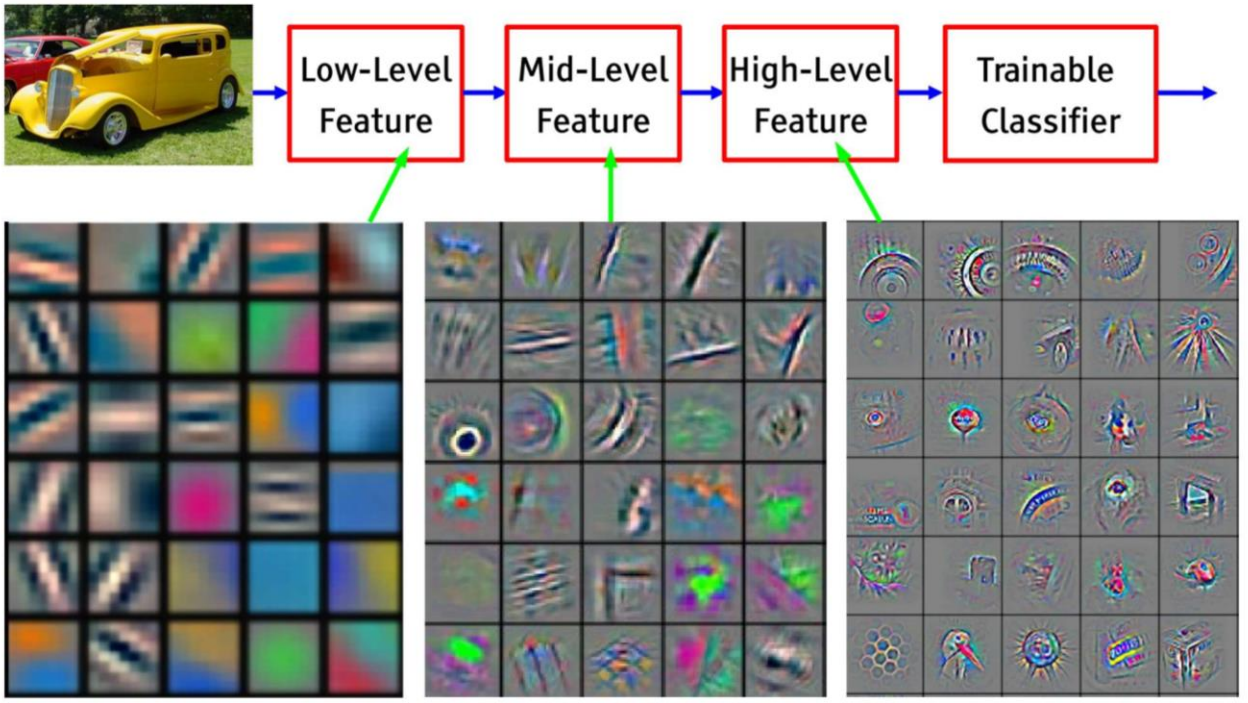
- 池化(pooling)：将每个特征图划分为多个（可重叠）局部区域，每个局部区域求均值或极大值输出
- 它是另外一种减少特征图尺寸的方法



揭神秘面纱:卷积神经网络CNN

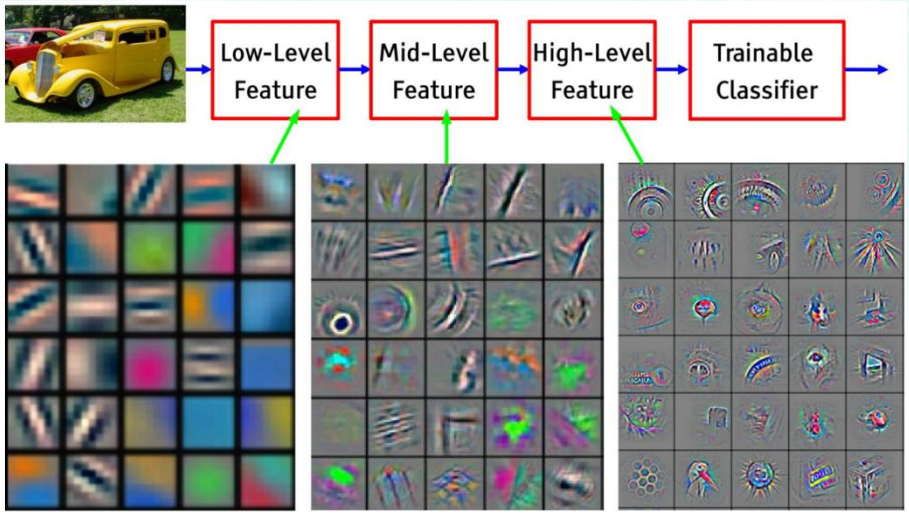
- CNN：多个（卷积层+激活函数）+多次池化+全连接层



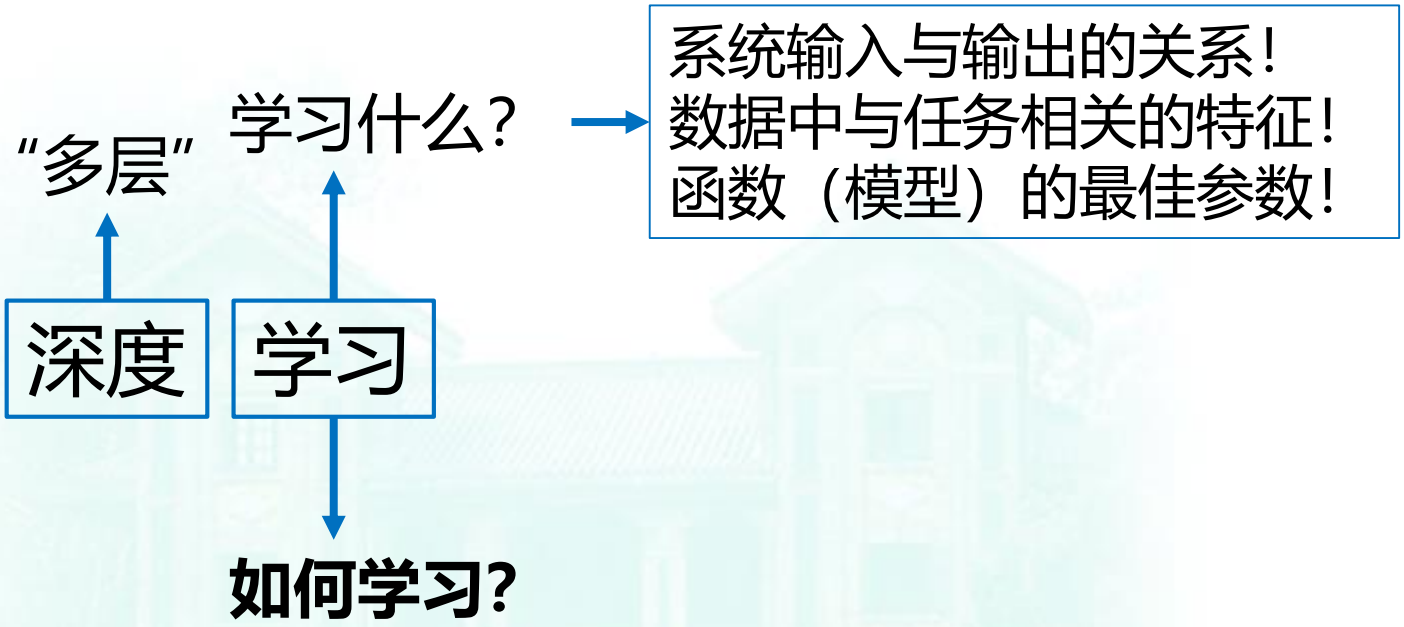


揭神秘面纱:深度学习(Deep Learning)

- “深度”：网络层数多
- “学习”：利用数据训练网络，找到每个卷积核的最佳参数，实现自动提取与任务相关的特征，即“特征学习”
- 注：每个卷积核（参数）不是人为设计的，而是自动学习的！
- 输入端是原始数据，输出端是预测结果，中间过程全部自动化（不用人为设计特征提取算法），所以叫“端到端学习”！
- 最终：学习到输入-输出关系



揭神秘面纱



揭神秘面纱：如何学习

- 比如针对分类任务

训练数据　模型参数

$$\min_{\boldsymbol{\theta}} \quad L(\boldsymbol{\theta}) \quad = \quad \frac{1}{N}\sum_{i=1}^{N} l(\mathbf{y}_i, \mathbf{f}(\mathbf{x}_i; \boldsymbol{\theta}))$$

$$= \quad -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_{ik} \log \hat{y}_{ik}$$

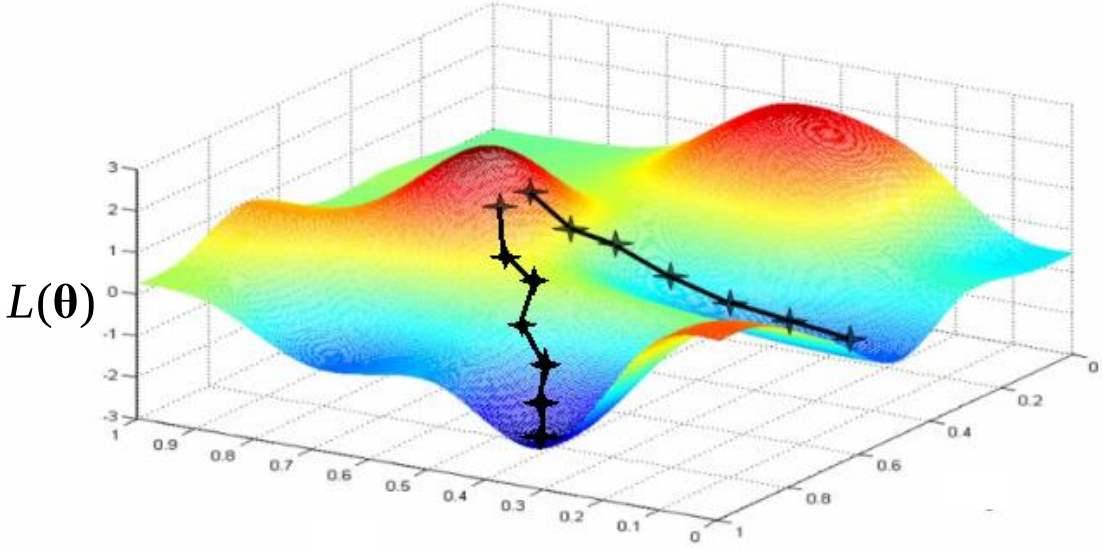
寻找模型（函数）最佳参数 = 最小化损失函数$L(\boldsymbol{\theta})$

梯度下降法！

基于训练数据，通过梯度下降法最小化损失函数，以此找到最佳的模型参数！

揭神秘面纱:如何学习好

- 梯度下降法不是只能找到局部最优解么?



局部最优解和全局最优解对应的模型性能常常类似!

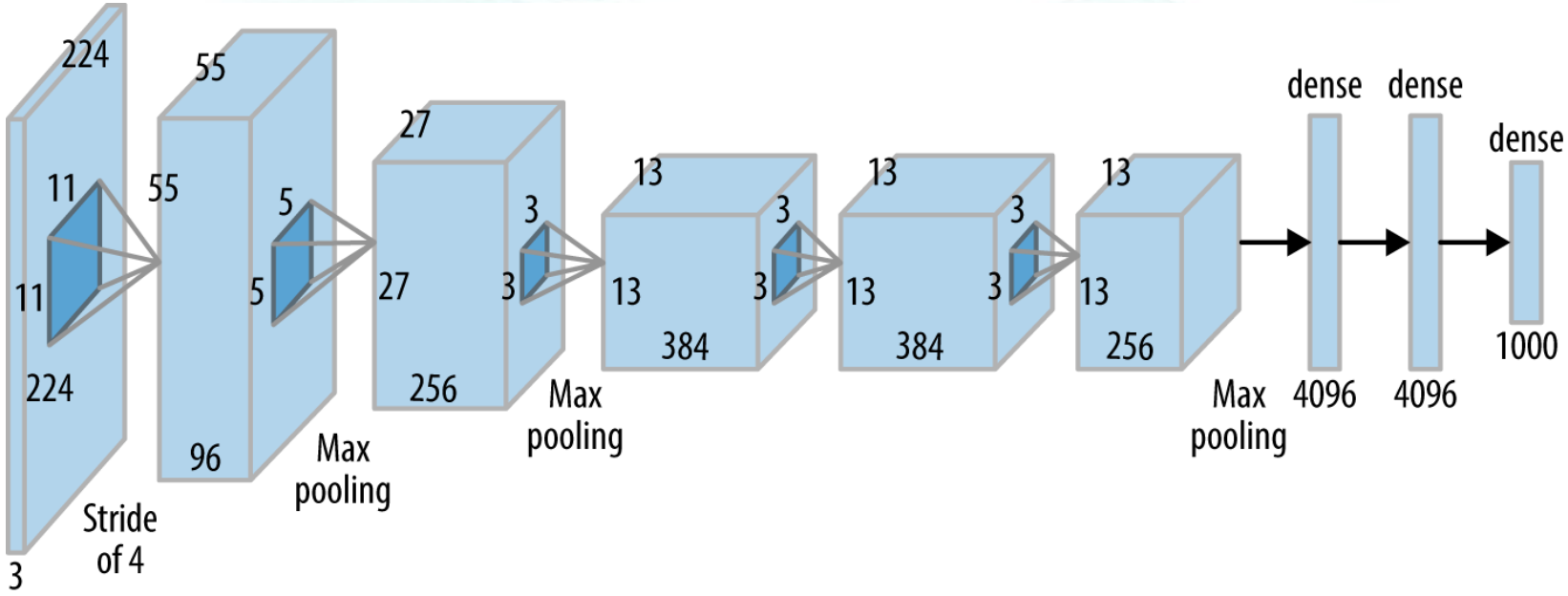
各种训练策略(正则化,Dropout等)提升模型的预测性能!

赏群模乱舞

深度学习模型结构和应用场景多种多样！

群模乱舞之图像分类:AlexNet （2012）

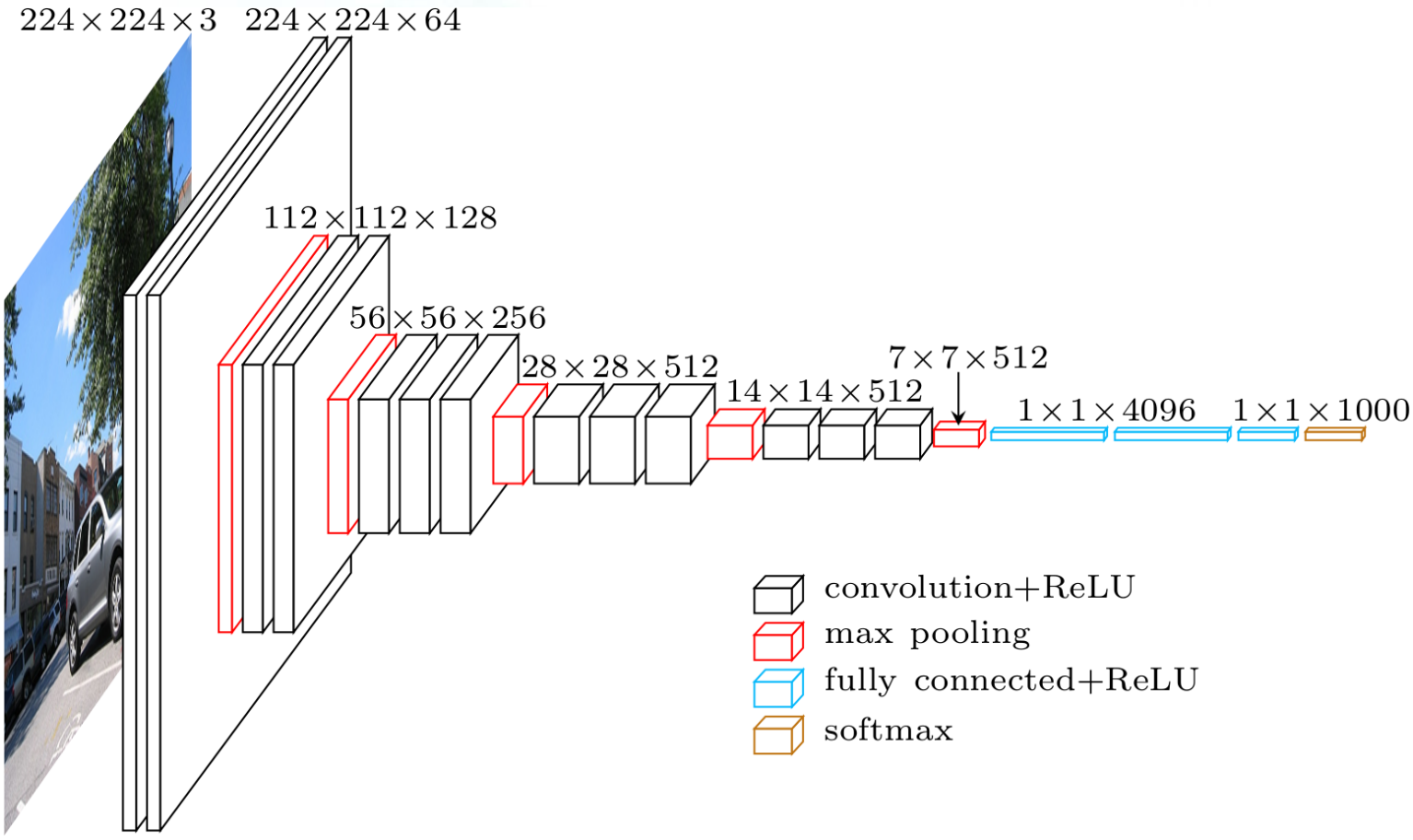
- AlexNet: CNN，5卷积层+3全连接层，1000类图像分类器
- 分类性能远超人工设计的特征提取器+训练的分类器
- 从此深度学习开始进入科研人员视野！



注：每个卷积层输出之后都接有ReLU激活函数

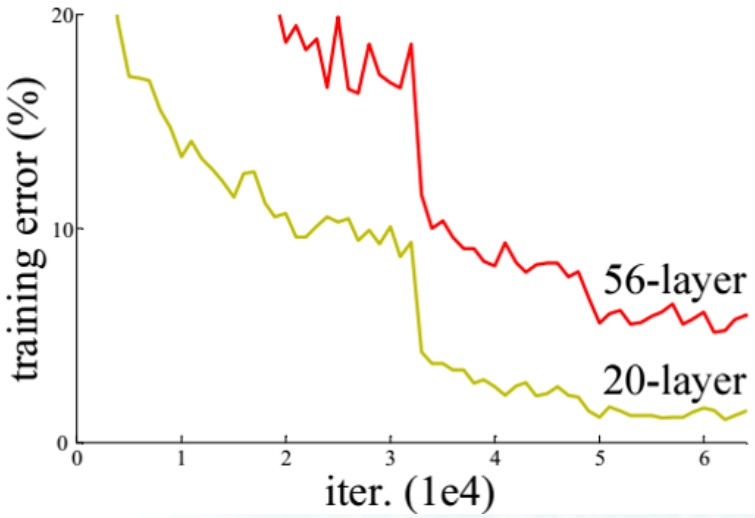
群模乱舞之图像分类:VggNet （2014）

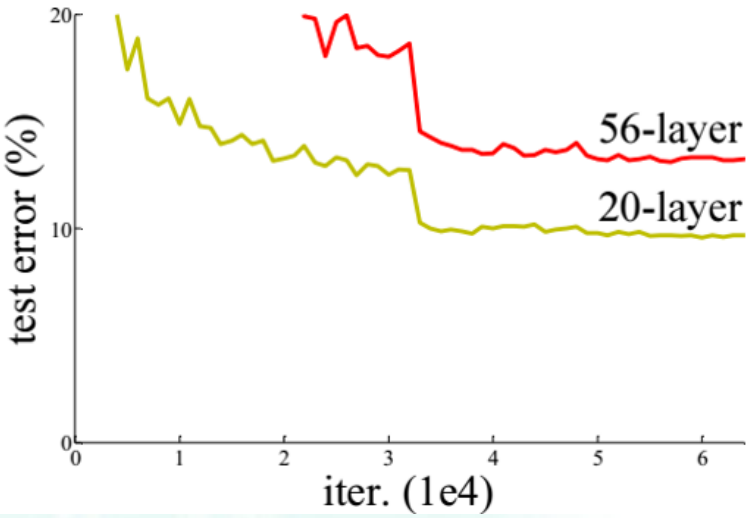
- 更多卷积层，每个卷积核大小为3x3
- 层数越多，越能表示更复杂的输入-输出关系



群模乱舞之图像分类:ResNet (2015)

- 奇怪：56层CNN分类器在训练集上分类误差大于20层CNN分类器！
- 说明：网络太深（层数太多），难于训练（被优化）！





群模乱舞之图像分类:ResNet (2015)

- 残差神经网络(Residual Network)
- 创新点：Skip connection

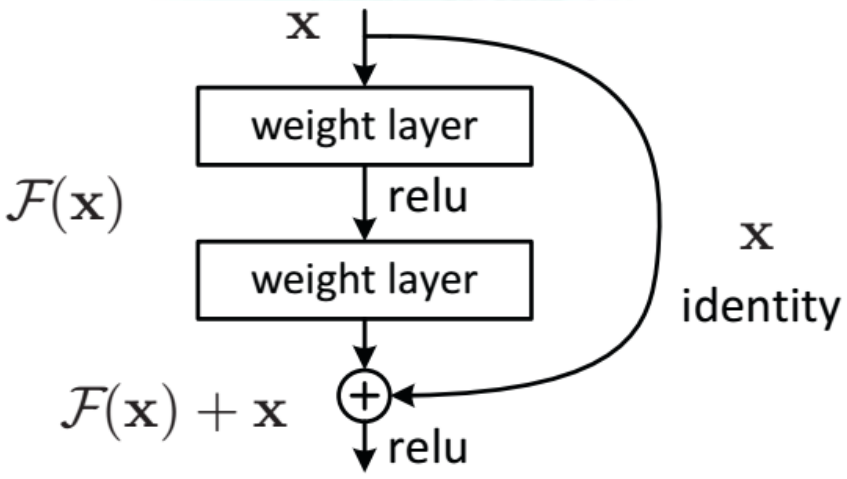
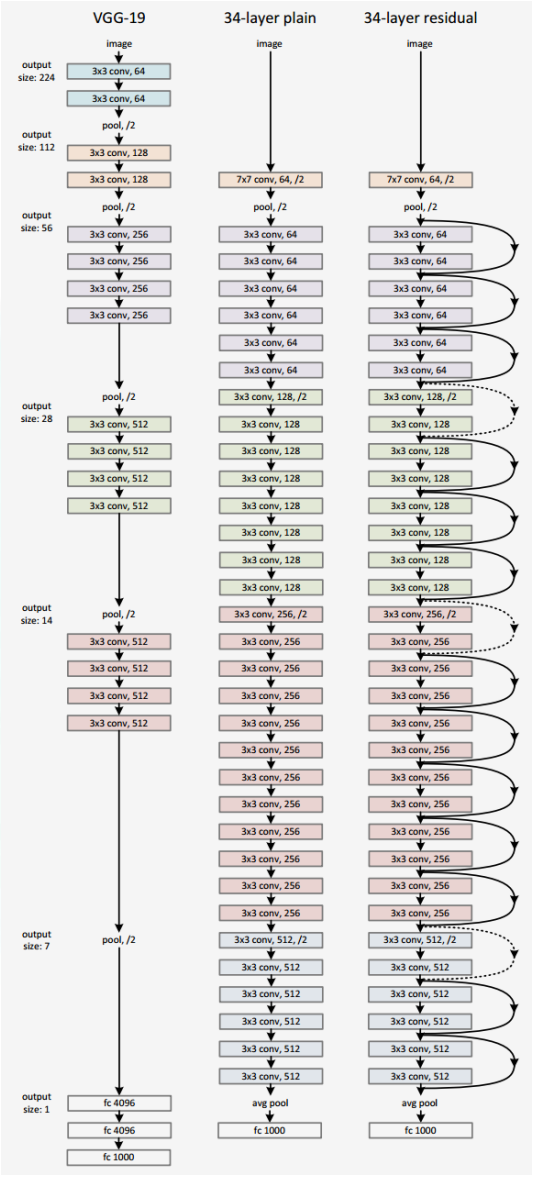


Figure 2. Residual learning: a building block.

$$\mathcal{H}(\mathbf{x}) = \mathcal{F}(\mathbf{x}) + \mathbf{x}$$

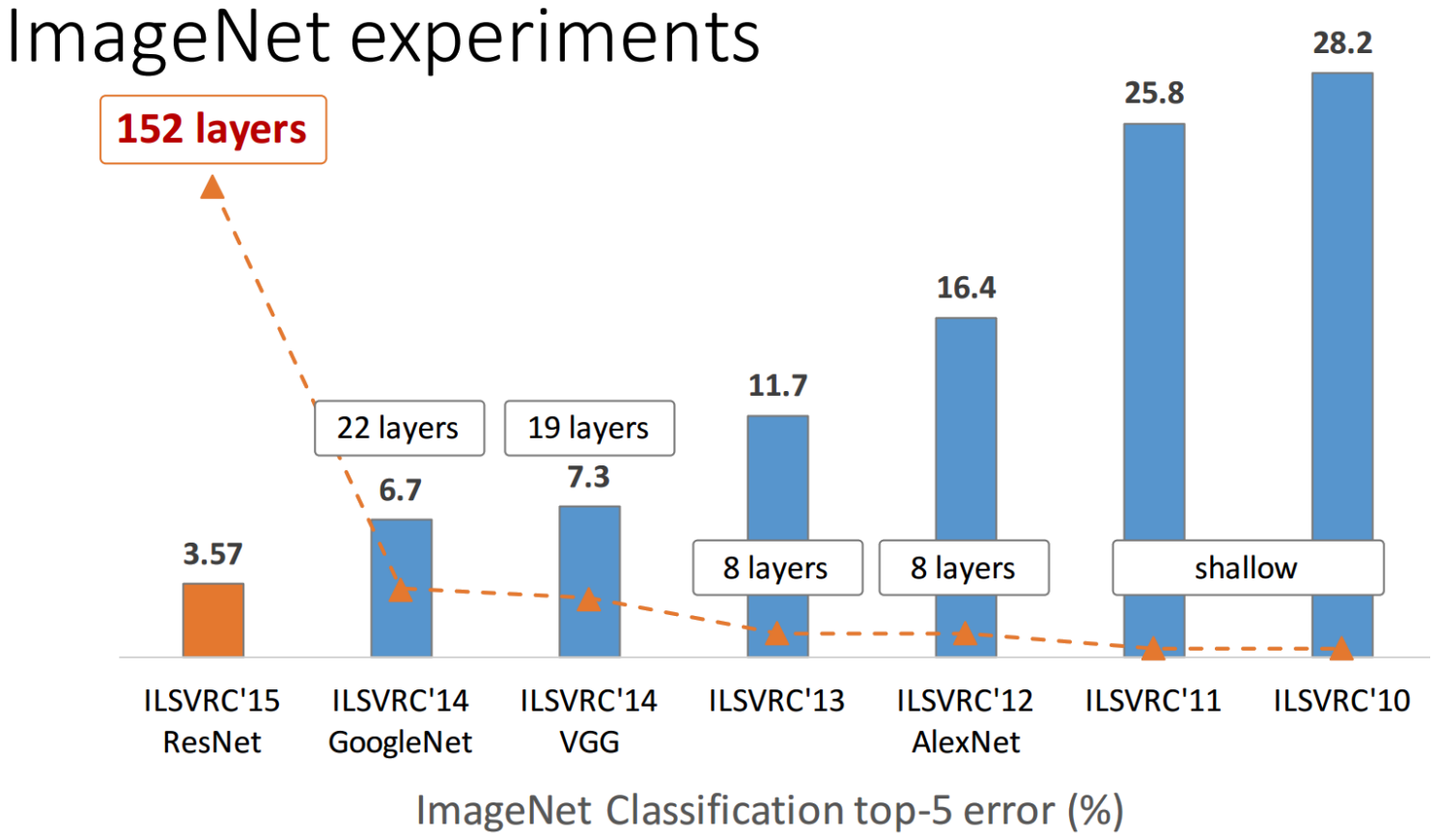
$$\mathcal{F}(\mathbf{x}) = \boxed{\mathcal{H}(\mathbf{x}) - \mathbf{x}}$$

Residual



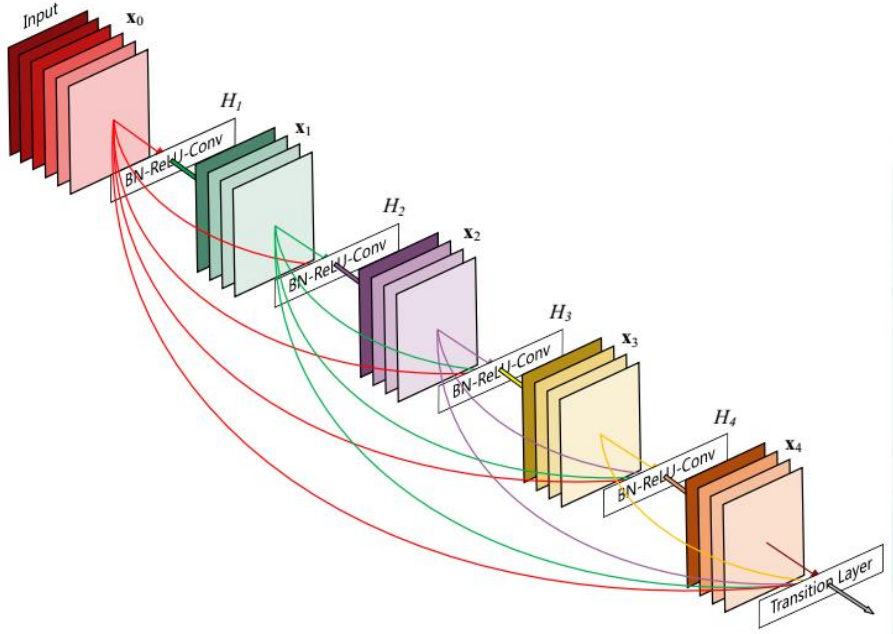
群模乱舞之图像分类:ResNet (2015-2016)

- 在1000类图像分类上首次超过人类表现!

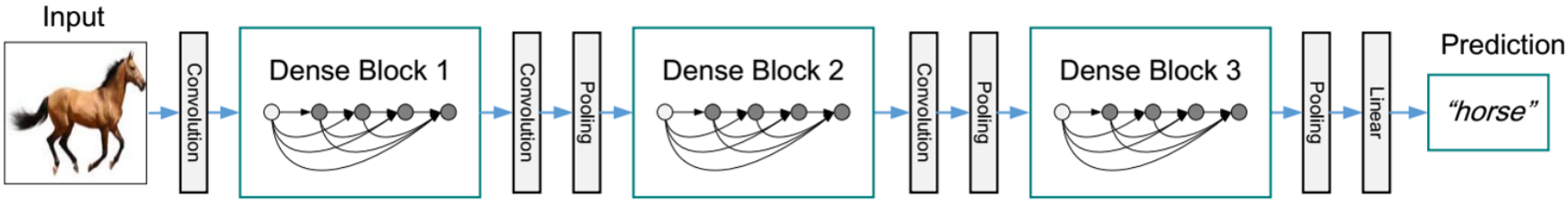


群模乱舞之图像分类:DenseNet (2017)

- 更多卷积层之间加入skip connections



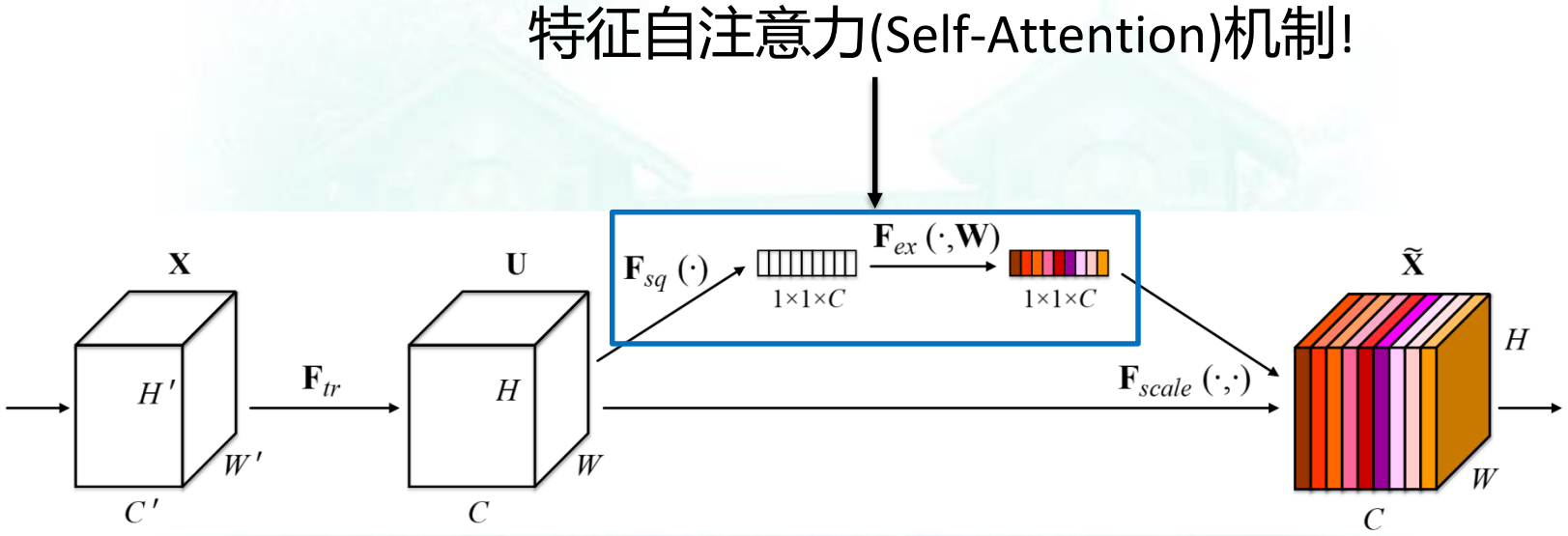
Dense block



DenseNet一般不需要太多卷积层

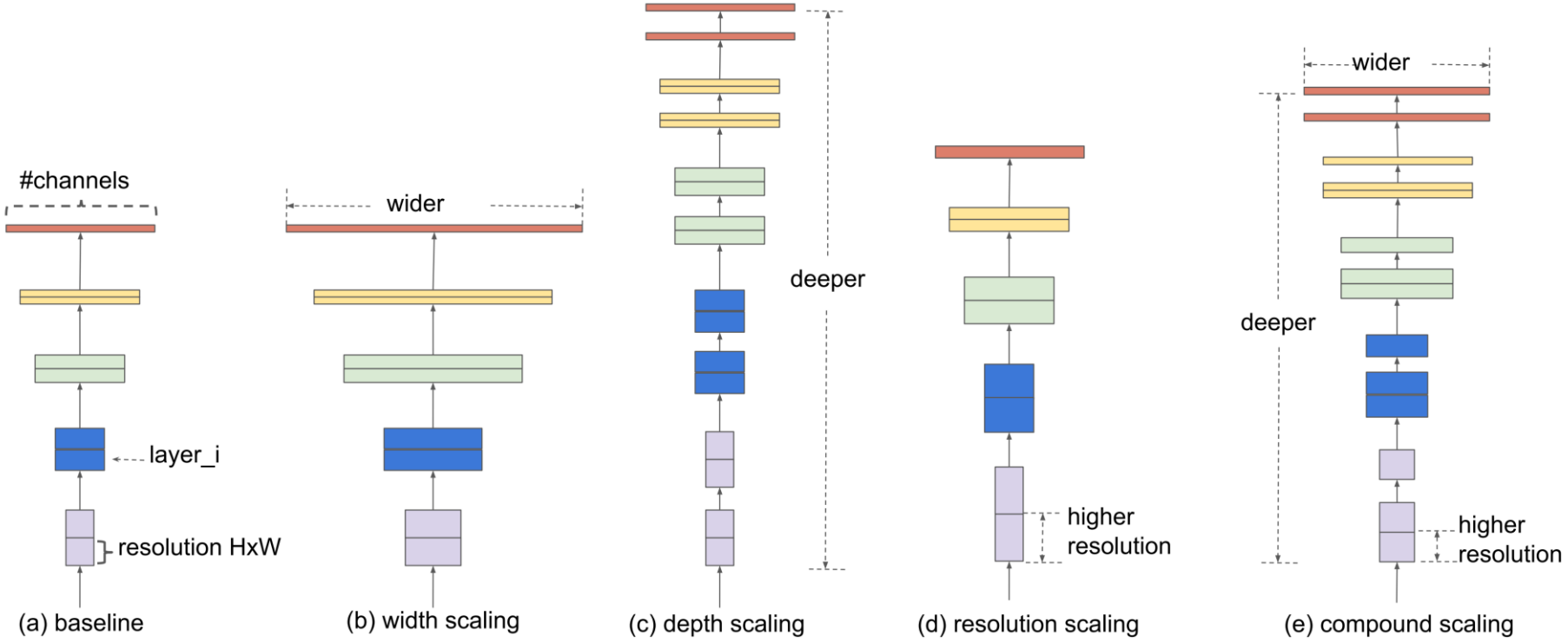
群模乱舞之图像分类:SENet (2018)

- Squeeze-and-Excitation Network:每个卷积层输出端加入SE模块
- 自动估计每个特征图的重要性，得到加权的特征图



群模乱舞之图像分类:EfficientNet(2019)

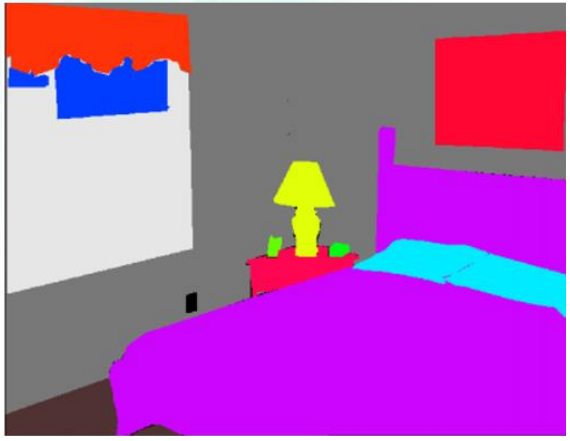
- 协同增加模型深度、宽度、输入尺寸
- EfficientNet B0 – B7: 模型尺寸逐渐增加;
- 与ResNet比：分类更准确，模型尺寸更小/运算更快



EfficientNet

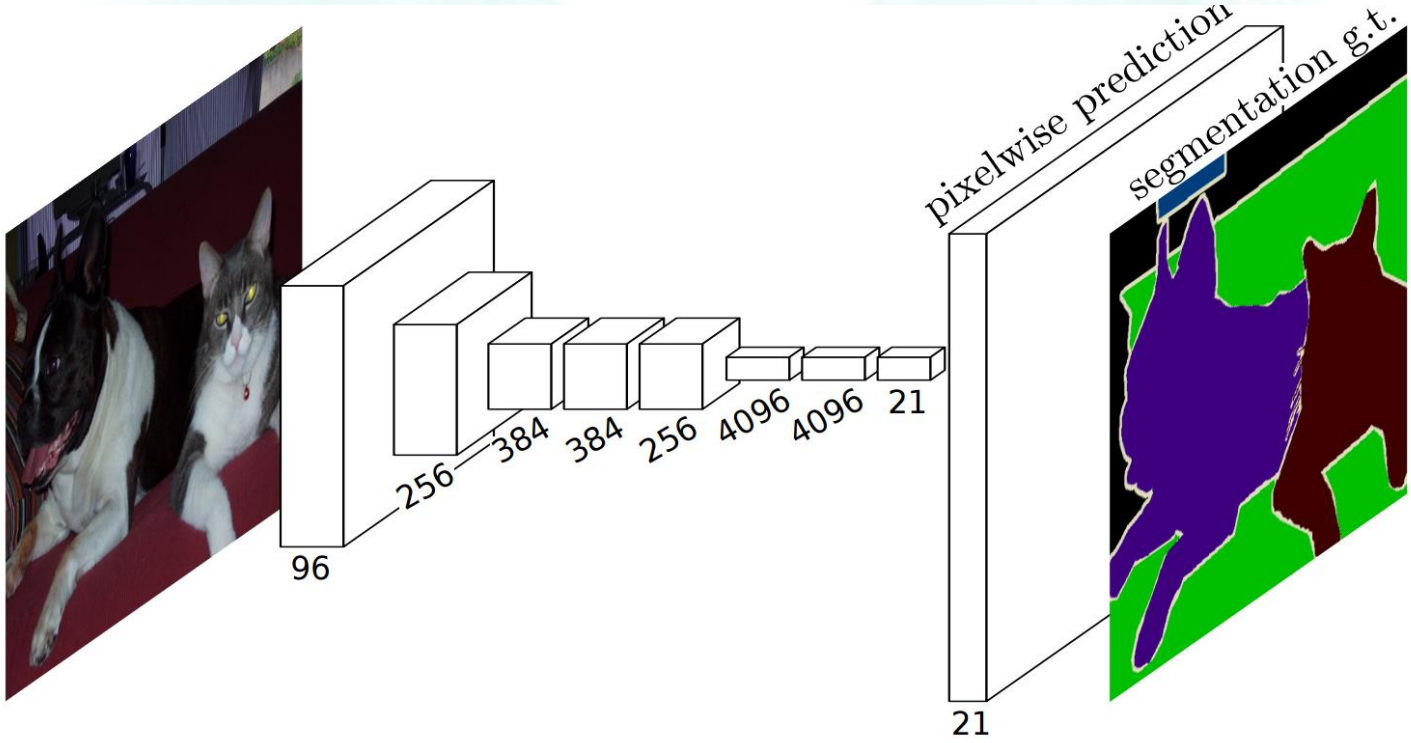
群模乱舞之图像语义分割

- 图像语义分割：将图像分割为不同种类的区域
- 目标：得到每个像素的种类



群模乱舞之图像语义分割

- 分割模型输出：大小与输入一样，层数等于区域类型的个数
- 输出包含了每个像素属于每一类别的概率
- 挑战：多层卷积后的输出尺寸远小于输入数据
- 思路：需要某种上采样操作将卷积输出尺寸变为与输入一样

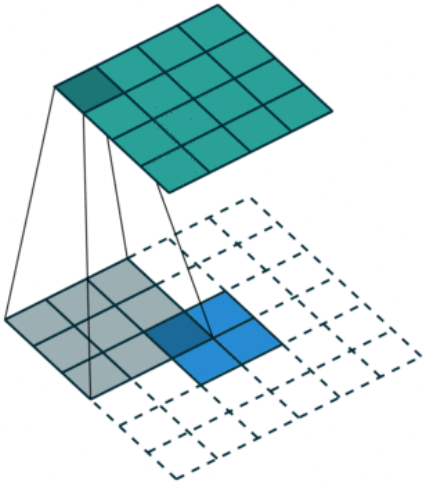


群模乱舞之图像语义分割：反卷积

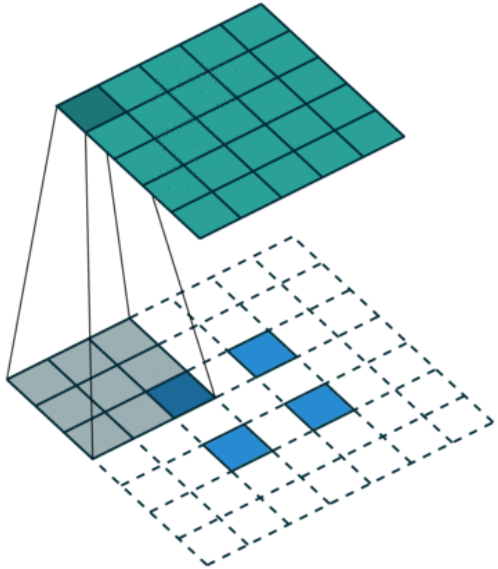
- Deconvolution（反卷积）：仍然是卷积操作
- Stride和Padding的作用与传统卷积里的作用几乎相反

输出
(绿色)

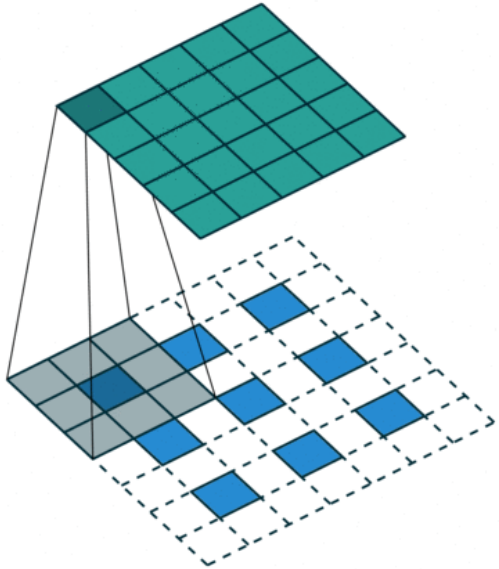
输入
(蓝色)



Stride=1
No padding



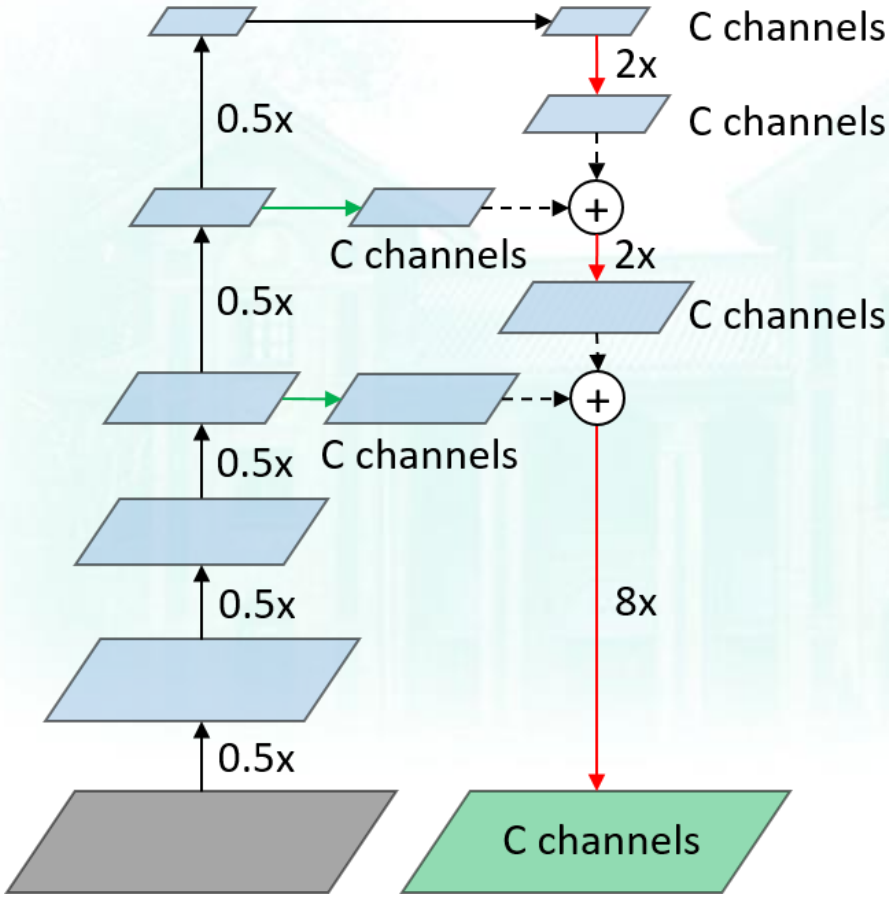
Stride=2
No padding



Stride=2
Padding=1

群模乱舞之图像语义分割：FCN

- 全卷积神经网络（Fully Convolutional Network）
- 结构：编码器（卷积层）+译码器（反卷积层）



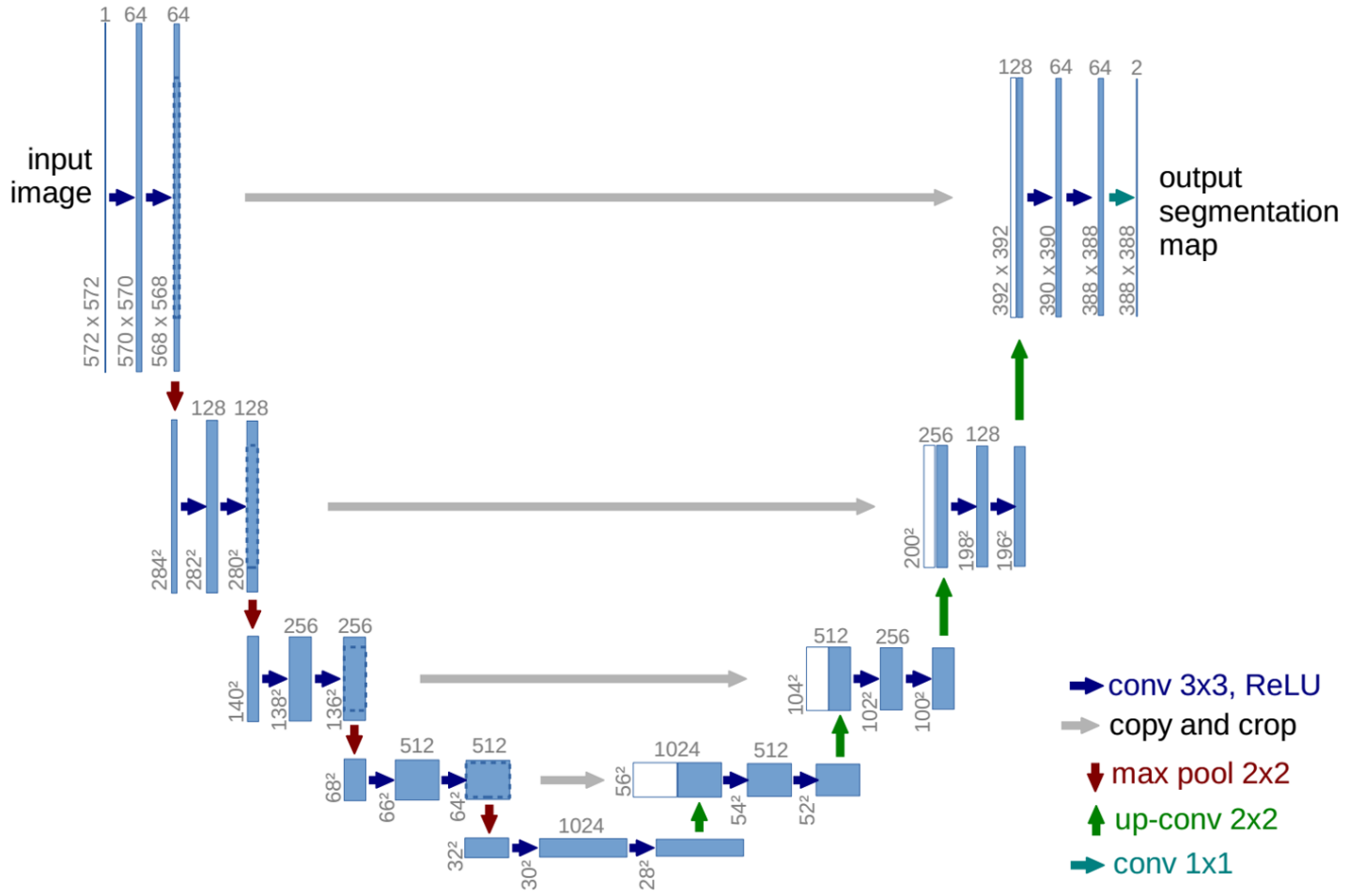
Stride=2
Kernel size 4x4

Stride=2
Kernel size 4x4

Stride=8
Kernel size 16x16

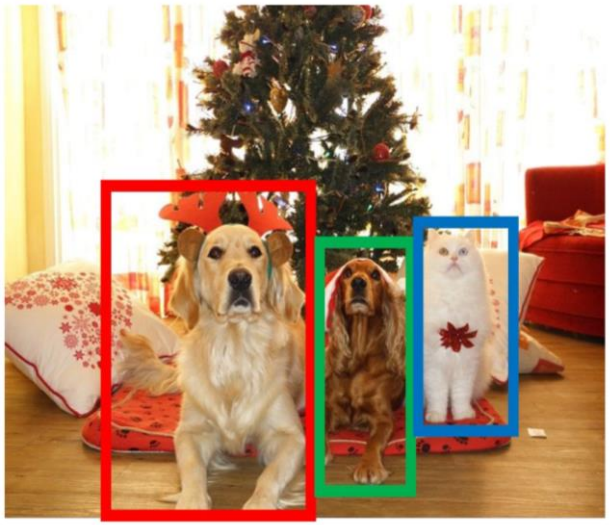
群模乱舞之分割模型：U-Net

- U-Net:FCN改进版，在上采样通道中叠加下采样通道的信息
- 上采样过程中融合了更多的图像细节信息和不同层次的特征



群模乱舞之目标检测

- 任务：从图像中检测出（未知数量）物体的位置、大小、类型



一般分两步来解决任务：

第一步：找到可能的物体区域（边框）

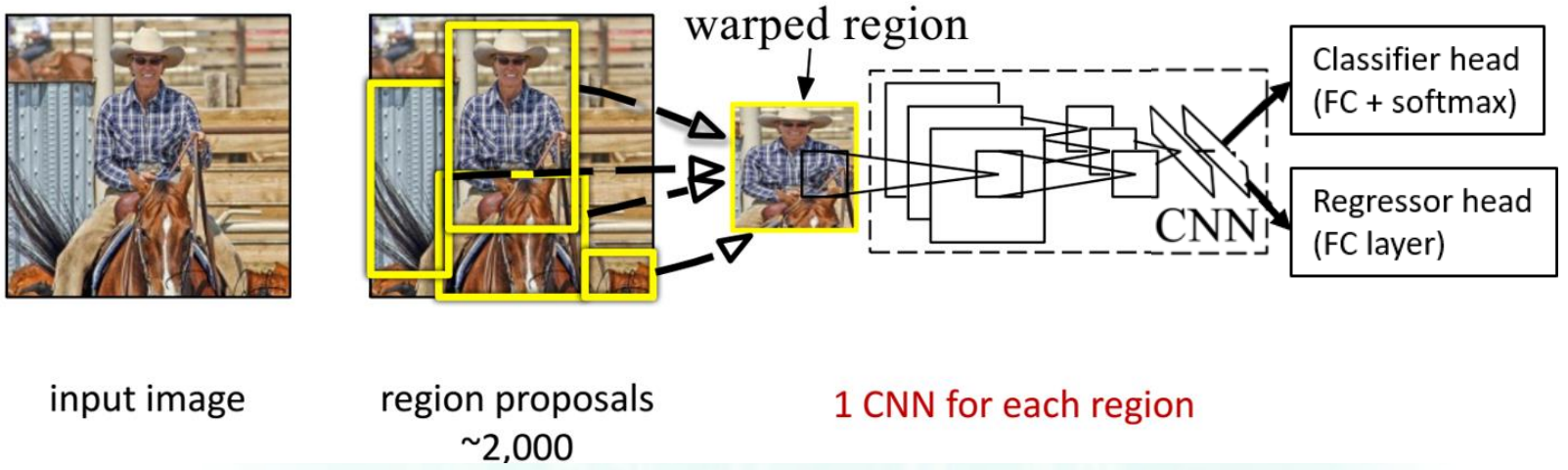
第二步：估计每个物体区域内物体类型，并自动微调边框大小与位置

分类任务

回归任务（边框坐标）

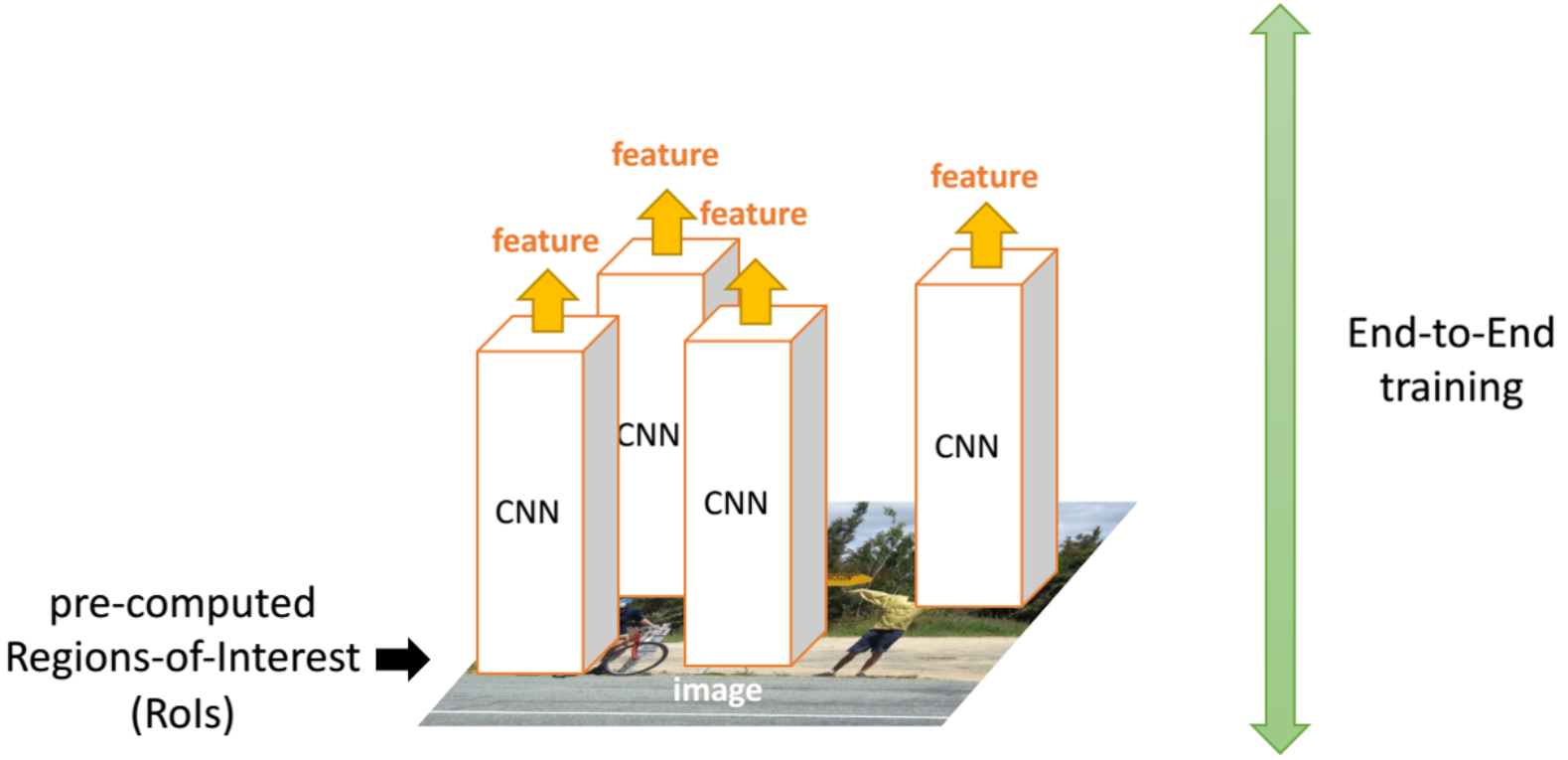
群模乱舞之目标检测：R-CNN

- Region-based CNN
  - 第一步：利用其它已有方法预先找到大量可能的物体区域
  - 第二步：基于这些区域训练共享卷积层的分类器和回归模型



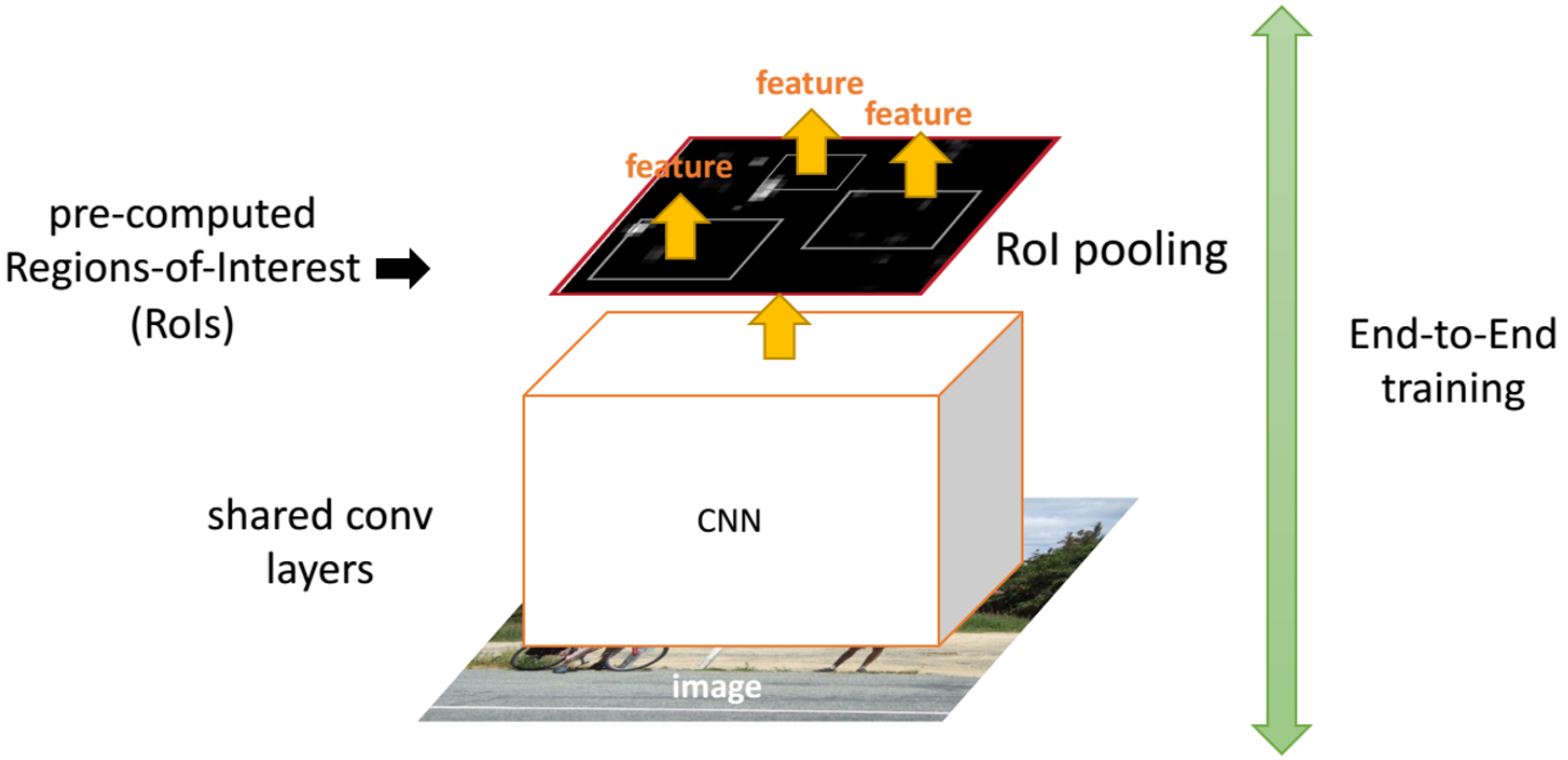
群模乱舞之目标检测：R-CNN

- 可能的物体区域从原始图像中估计得到
- 从图像中进行目标检测时，同一个CNN被执行多次！
- 因此，检测速度很慢！



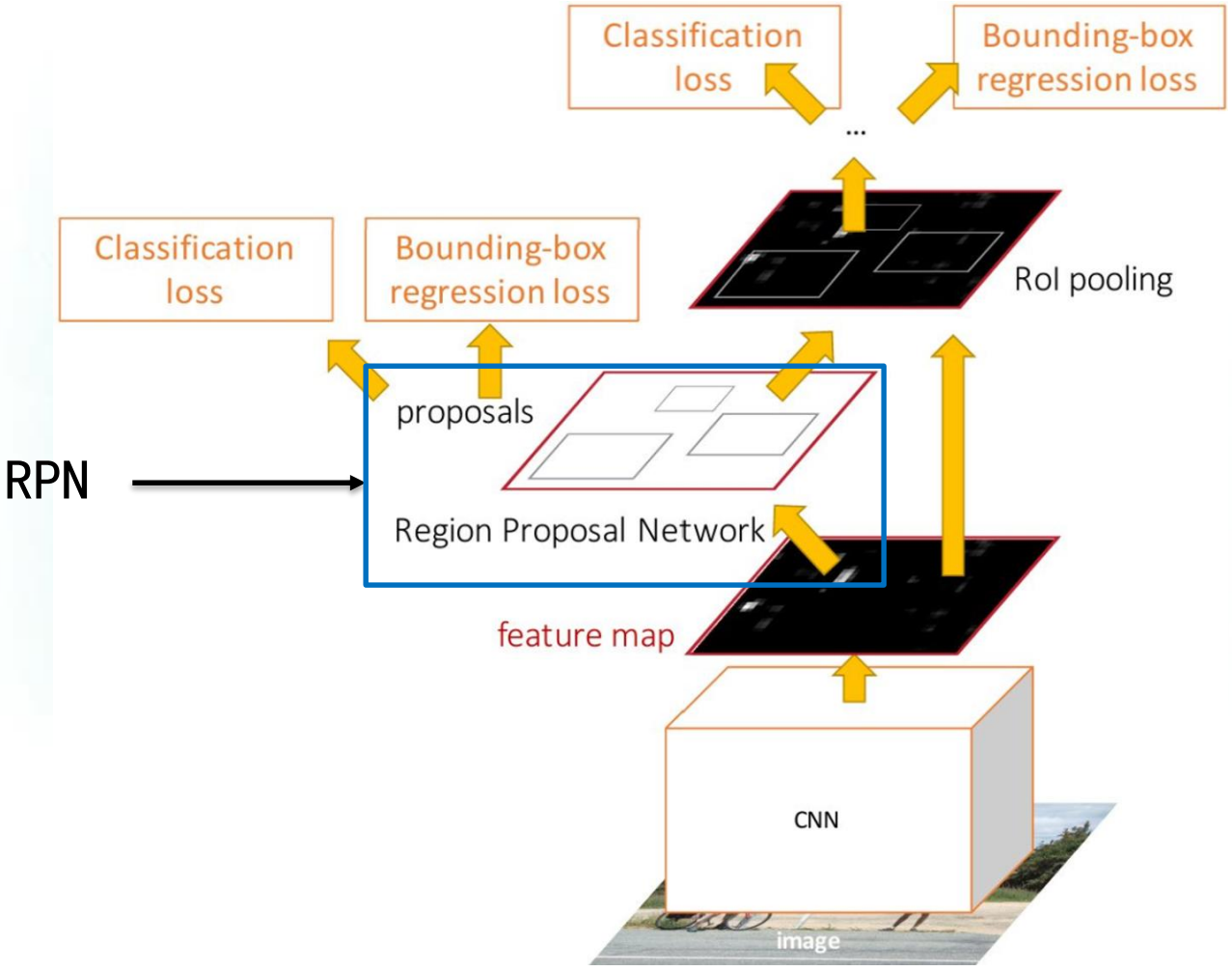
群模乱舞之目标检测：Fast R-CNN

- 加速思路：可能的物体区域从卷积层输出（特征图）得到
- 对一张图像中所有可能的物体检测，卷积操作只被执行一次！
- 大大加快了检测速度！



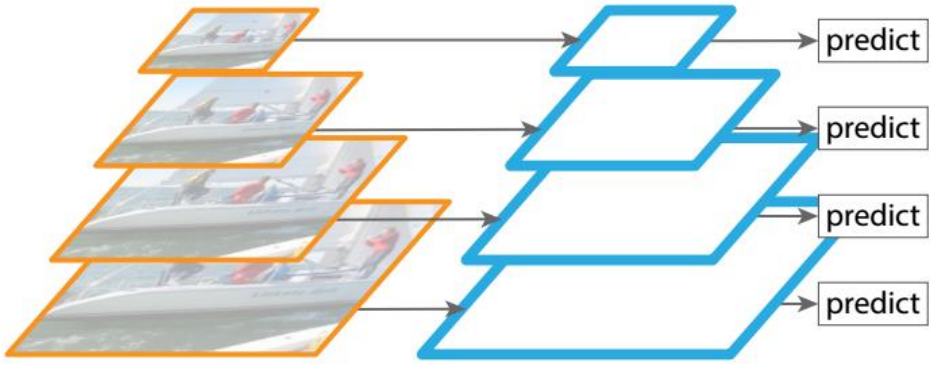
群模乱舞之目标检测：Faster R-CNN

- 进一步加速：同时训练一个子网络RPN寻找可能的物体区域

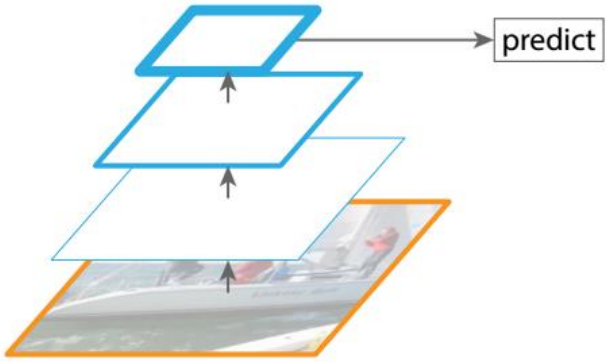


群模乱舞之目标检测：提升检测精度

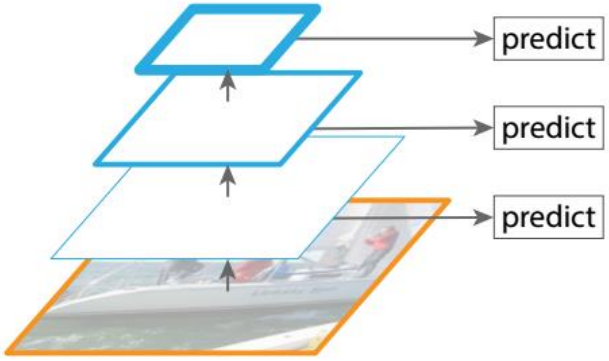
- 特征金字塔网络(FPN)在不同尺度的特征图里检测目标



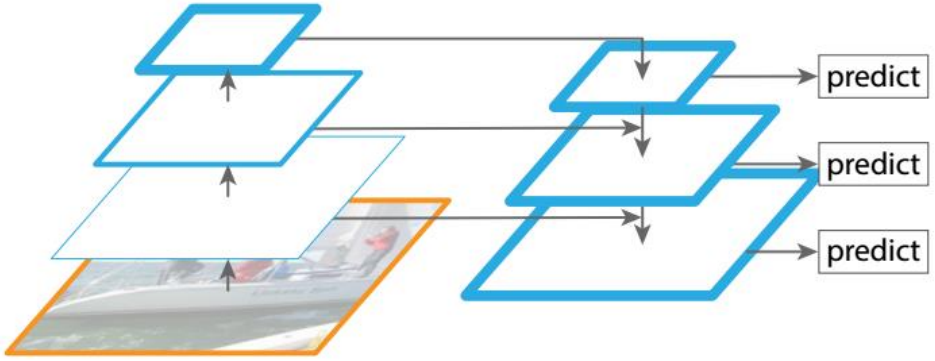
(a) Featurized image pyramid



(b) Single feature map



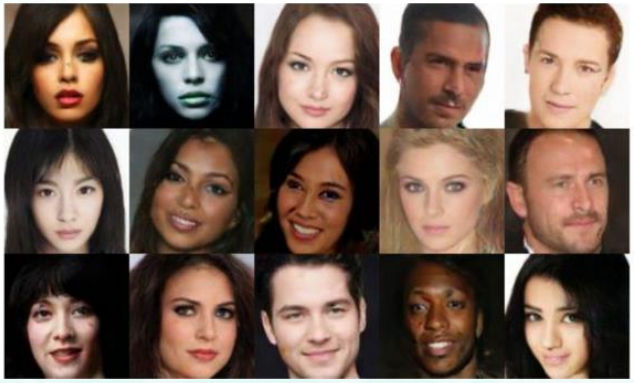
(c) Pyramidal feature hierarchy



(d) Feature Pyramid Network

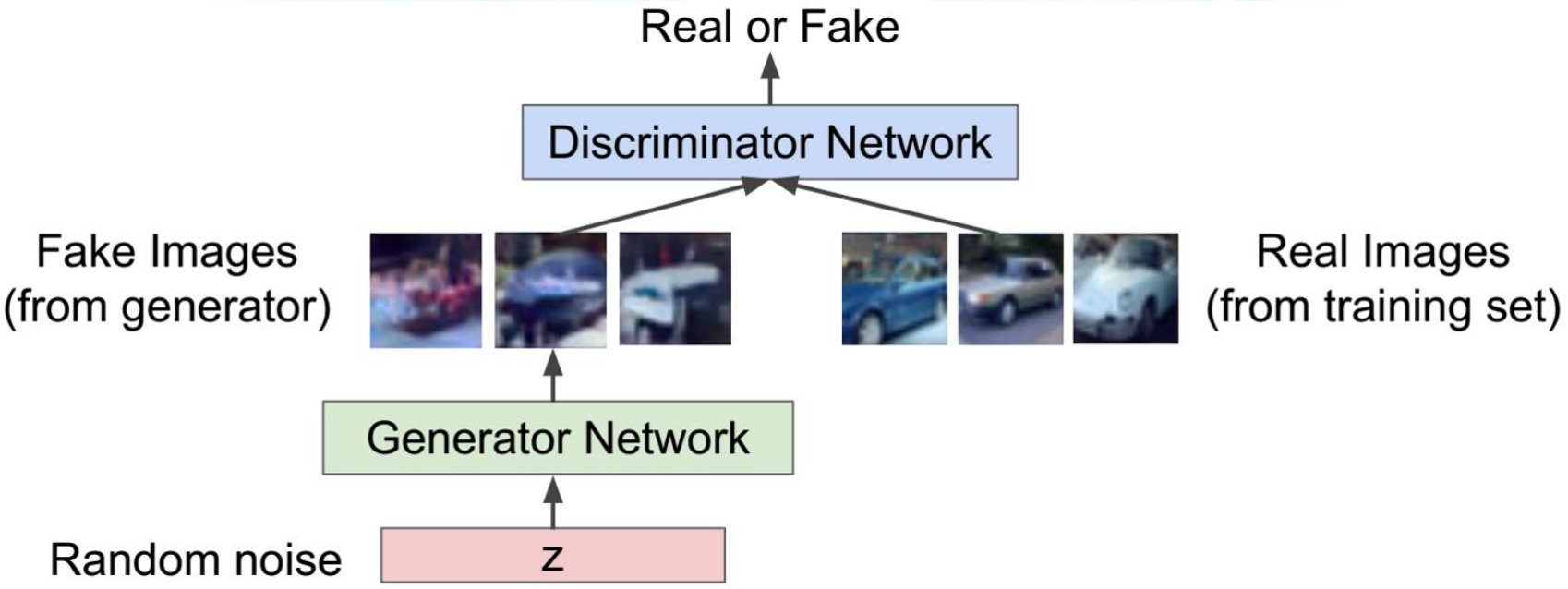
数据生成任务

- 基于已有数据，生成更多的相似数据
- 应用广泛：数据增广，高清成像，艺术设计，风格转移等


群模乱舞之GAN：初始模型

- 生成式对抗网络GAN：生成器网络G + 判别器网络D
- 核心思想：利用判别器判断生成器生成的数据是否足够的真实
  - 训练生成器，使其生成的数据尽量让判别器判断不出真假
  - 训练判别器，使其尽量能区分真实数据与生成的数据

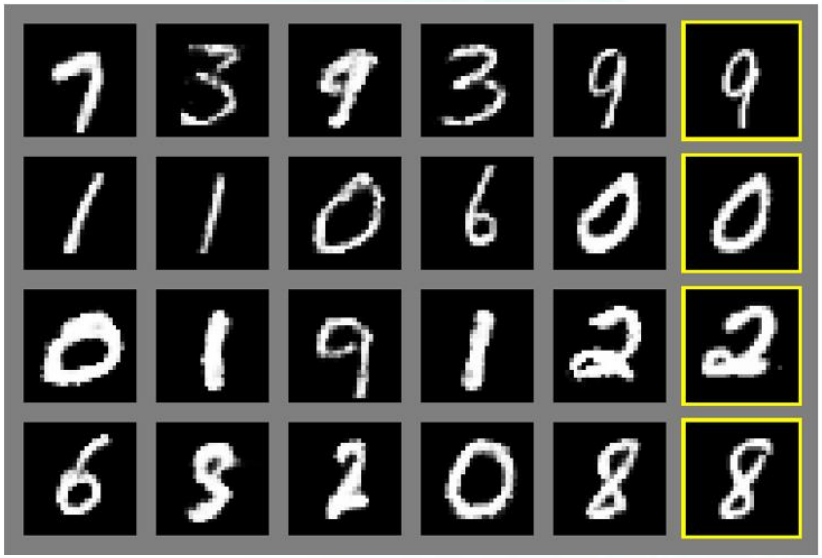


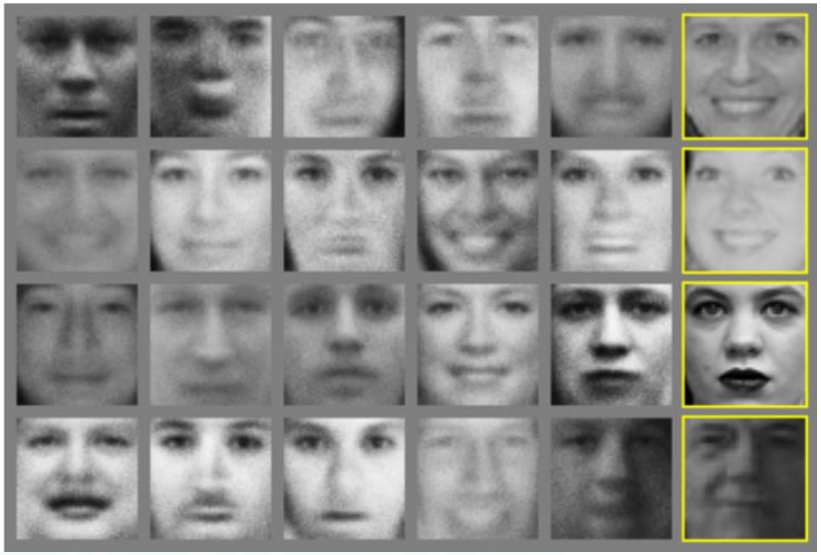
群模乱舞之GAN：初始模型

- 目标函数

真实数据 ↓ 生成的数据 ↓

$$\min_{G_\theta}\max_{D_w}\{\mathbb{E}_{x\sim P_r}[\log D_w(x)]+\mathbb{E}_{z\sim P(z)}[\log(1-D_w(G_\theta(z)))]\}$$

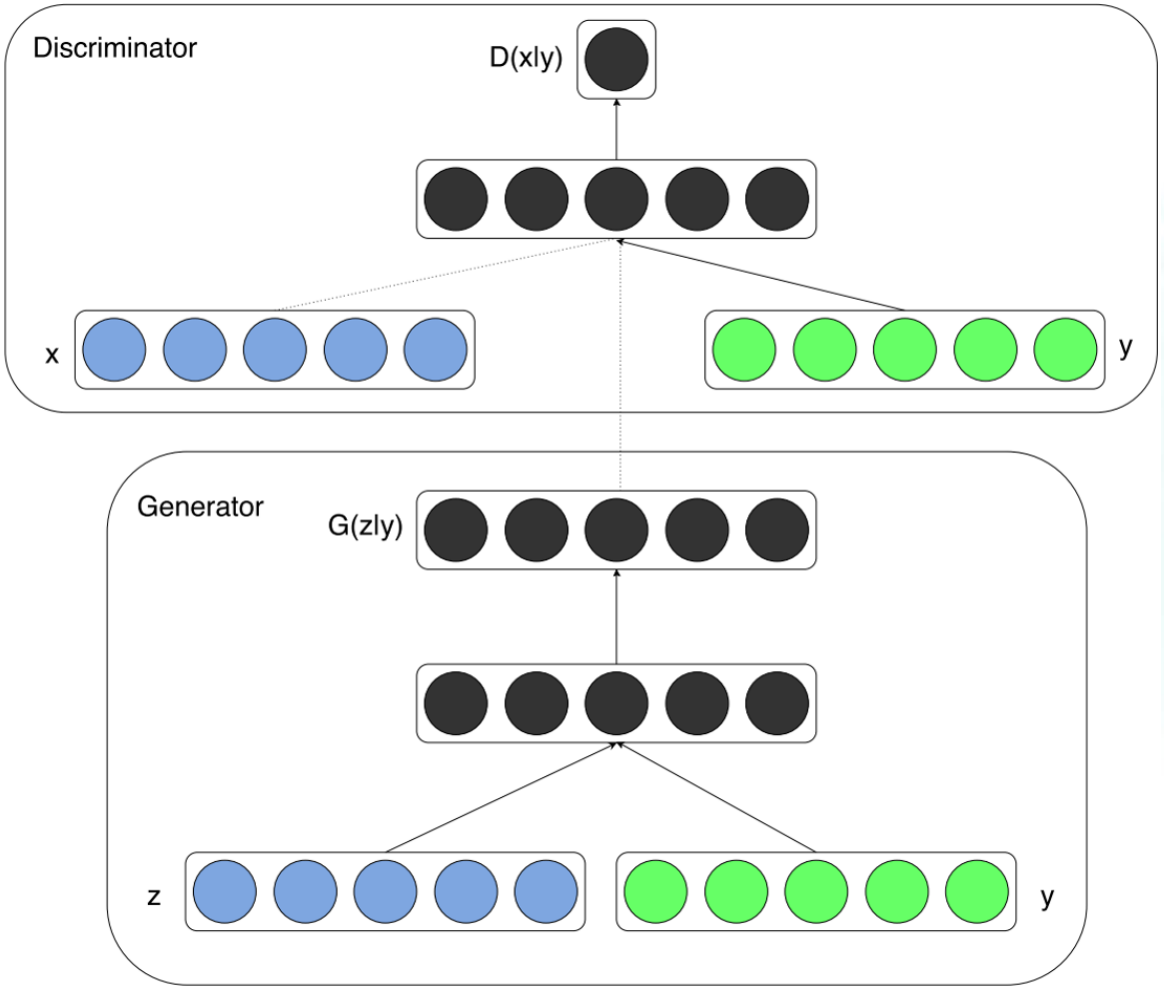




黄色框中图像为训练好的GAN生成的数据；
每行其它数据是与黄色框数据最相似的几个真实数据

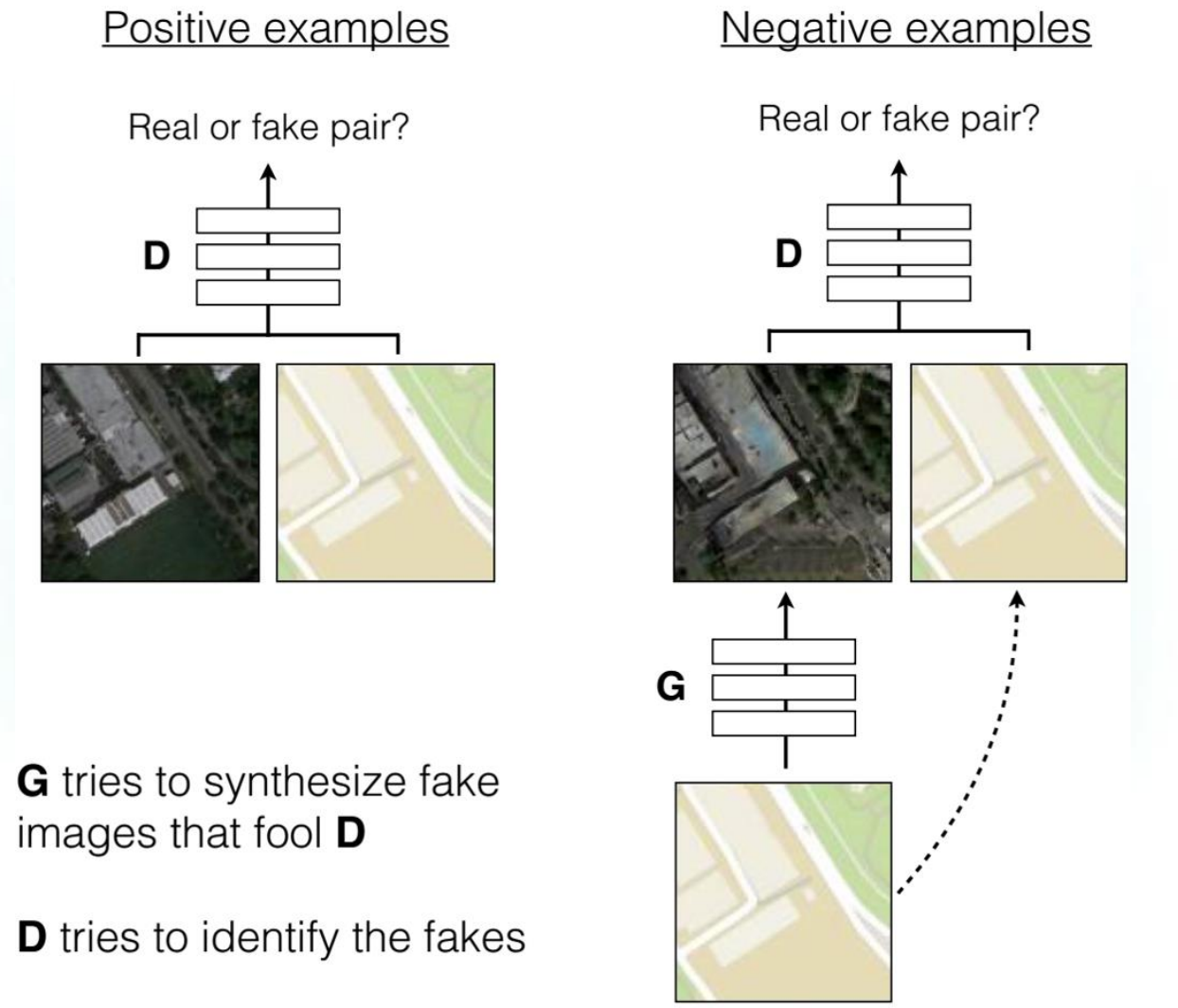
群模乱舞之GAN： Conditional GAN

- Condition ‘y’ 作为生成器和判别器输入的一部分
- ‘y’可以是向量，也可以是图像等更复杂的表示



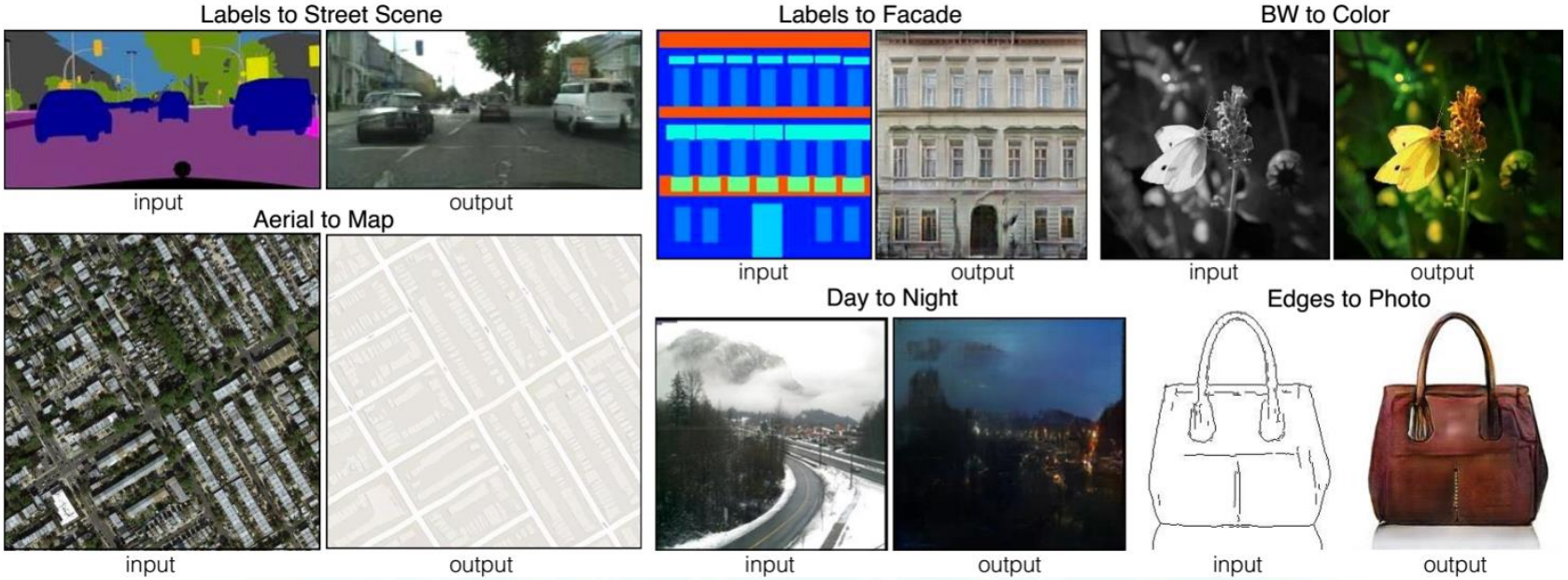
群模乱舞之GAN： Conditional GAN

- CGAN用于图像翻译



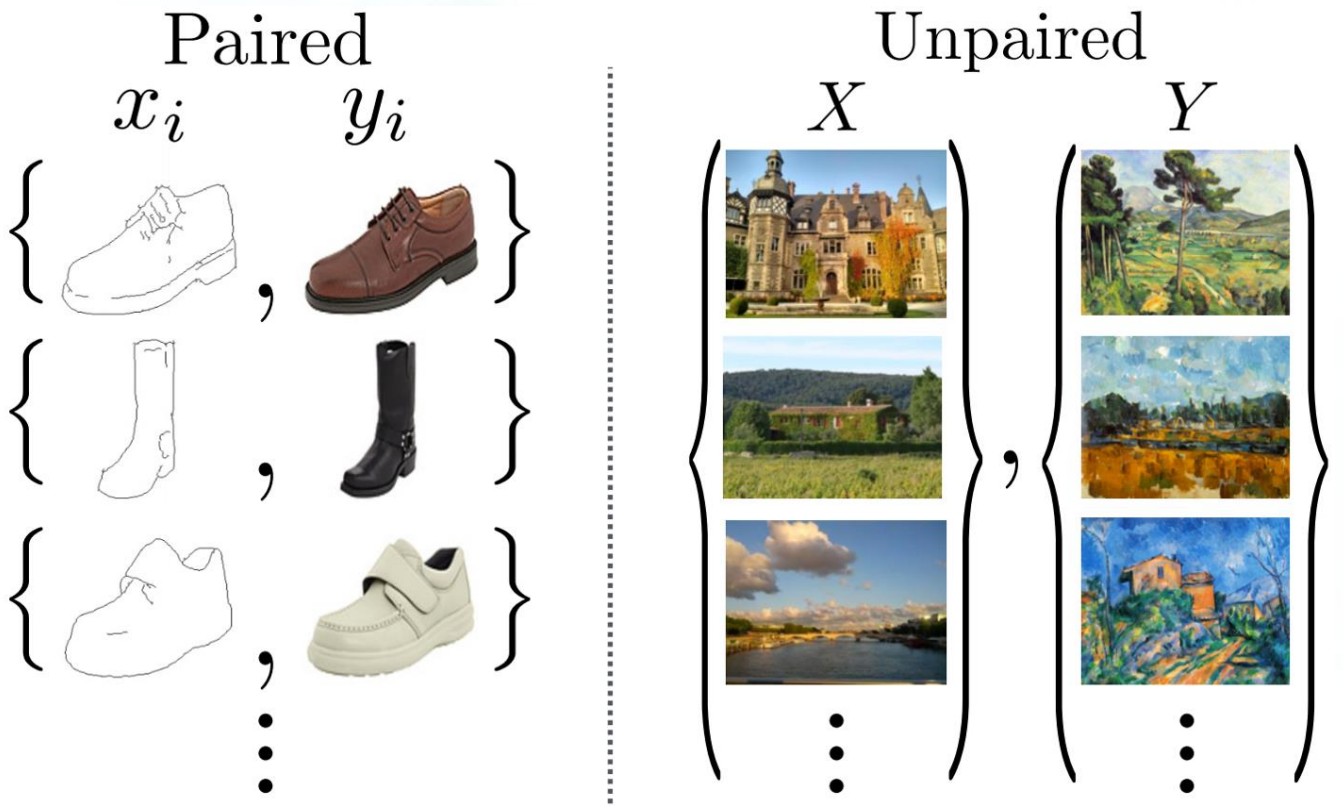
群模乱舞之GAN：Conditional GAN

- CGAN用于不同功能的图像翻译：黑白->彩色；素描->实物，等



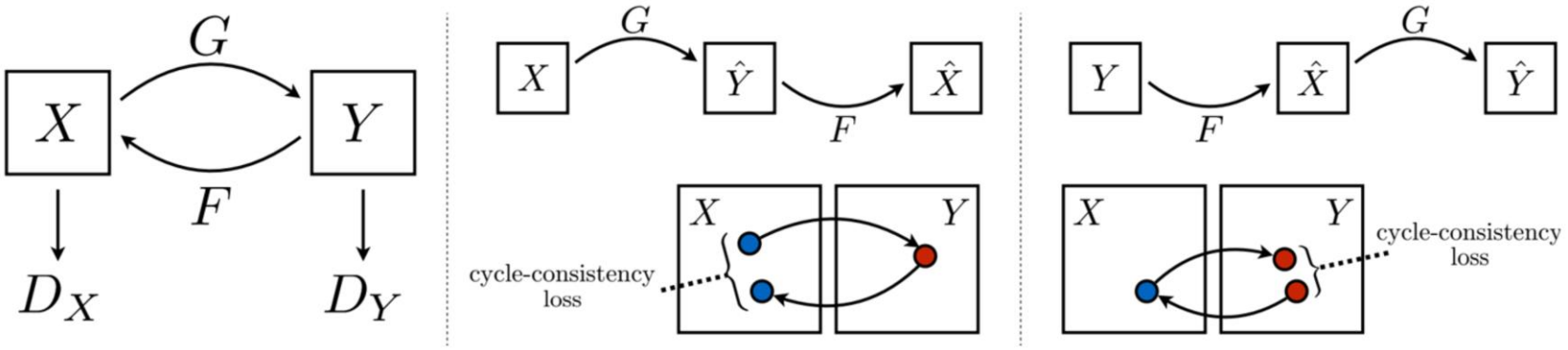
群模乱舞之GAN：CycleGAN

- 更加挑战情况下图像翻译：不存在一一对应的训练数据（右图）



群模乱舞之GAN：CycleGAN

- 思路：将图像从一个domain转换到另一个domain，然后再转换回来，得到的图像与原图像尽量一样；两个GAN模型

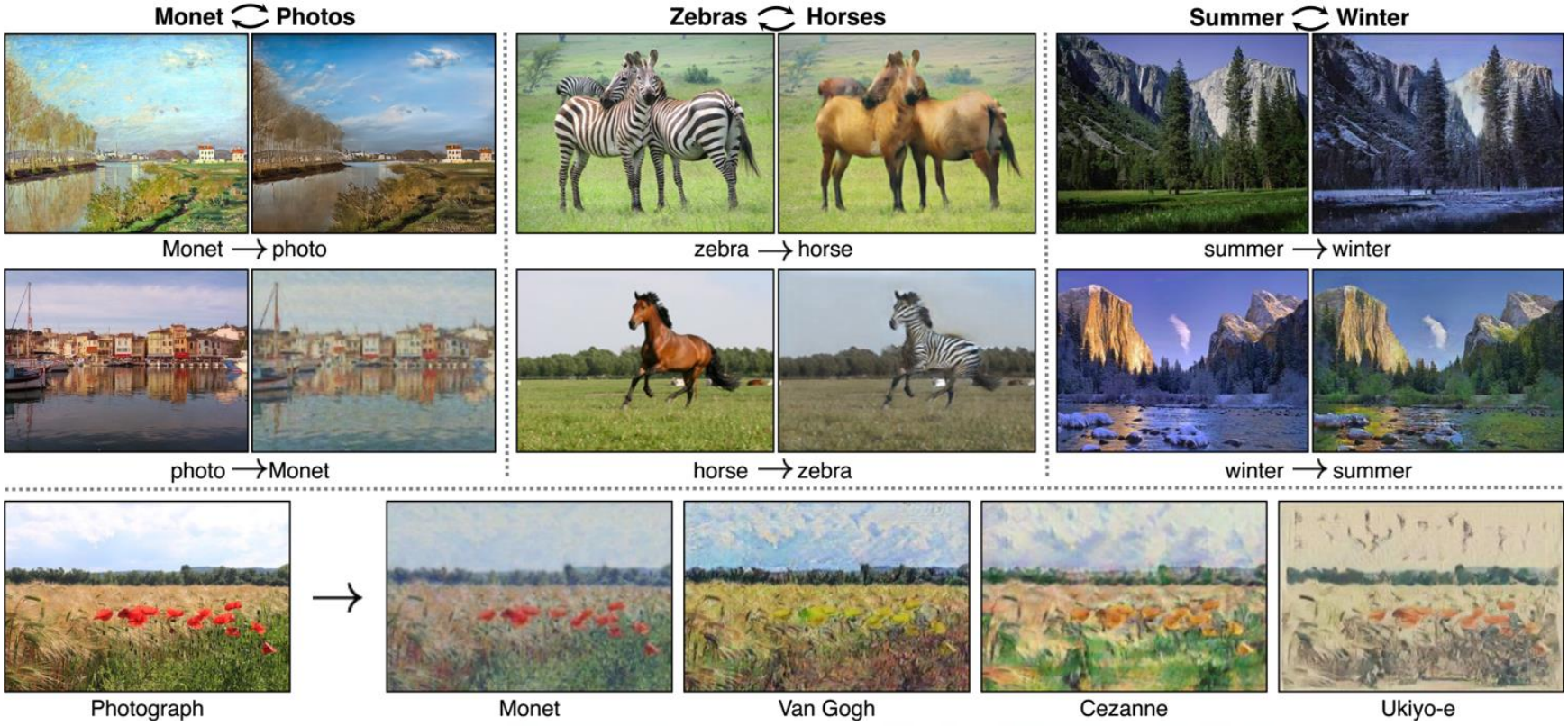


$$\mathcal{L}_{\text{cyc}}(G,F) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\|F(G(x)) - x\|_1] + \mathbb{E}_{y\sim p_{\text{data}}(y)}[\|G(F(y)) - y\|_1]$$

$$\mathcal{L}(G,F,D_X,D_Y) = \mathcal{L}_{\text{GAN}}(G,D_Y,X,Y) + \mathcal{L}_{\text{GAN}}(F,D_X,Y,X) + \lambda\mathcal{L}_{\text{cyc}}(G,F),$$

群模乱舞之GAN：CycleGAN

- CycleGAN用于图像风格转移



群模乱舞：自然语言处理（NLP）任务

- 文献分类
- 观点分析
- 机器翻译
- 阅读理解
- 聊天机器人
- 个人助手
- 文章总结
- ...

循环神经网络（Recurrent Neural Network）
Transformer模型

寻万能之源：为什么超越其它方法？

- 特定任务下的学习：自动学到任务相关的特征
- 深度模型 -> 多层次特征，更复杂非线性关系

为什么最近几年深度学习模型才表现更好？

- 大数据
- 硬件运算速度，多核并行
- 模型结构创新：ResNet,GAN,Memory N,GCN,...
- 模块/操作创新：BN,ReLU,dropout,skip, deconv, attention,...
- 算法创新：初始化，优化，损失函数，强化学习，...
- 研发平台：TensorFLow, PyTorch, ...
- 应用驱动：无人驾驶，安防，个人助手，智慧城市与医疗

深度学习可以解决所有任务了？

既然DL这么厉害，那我们还研究什么？

只是比传统方法相对好点而已！并不完美！

深度学习并不Perfect

需要大数据量 —数据少时怎么办?→ 小样本学习

数据标注费时费力 —无标注数据有用么?→ 无/自监督学习

训练测试计算量大 —手机等终端怎么用?→ 模型轻量化

测试场景单一 —场景变化时怎么办?→ 模型鲁棒性

预测过程不清楚 —要求预测证据怎么办?→ 模型可解释性

人工设计模型结构 —能否减少人工因素?→ 自动机器学习

……

追研究前沿：小样本下深度学习

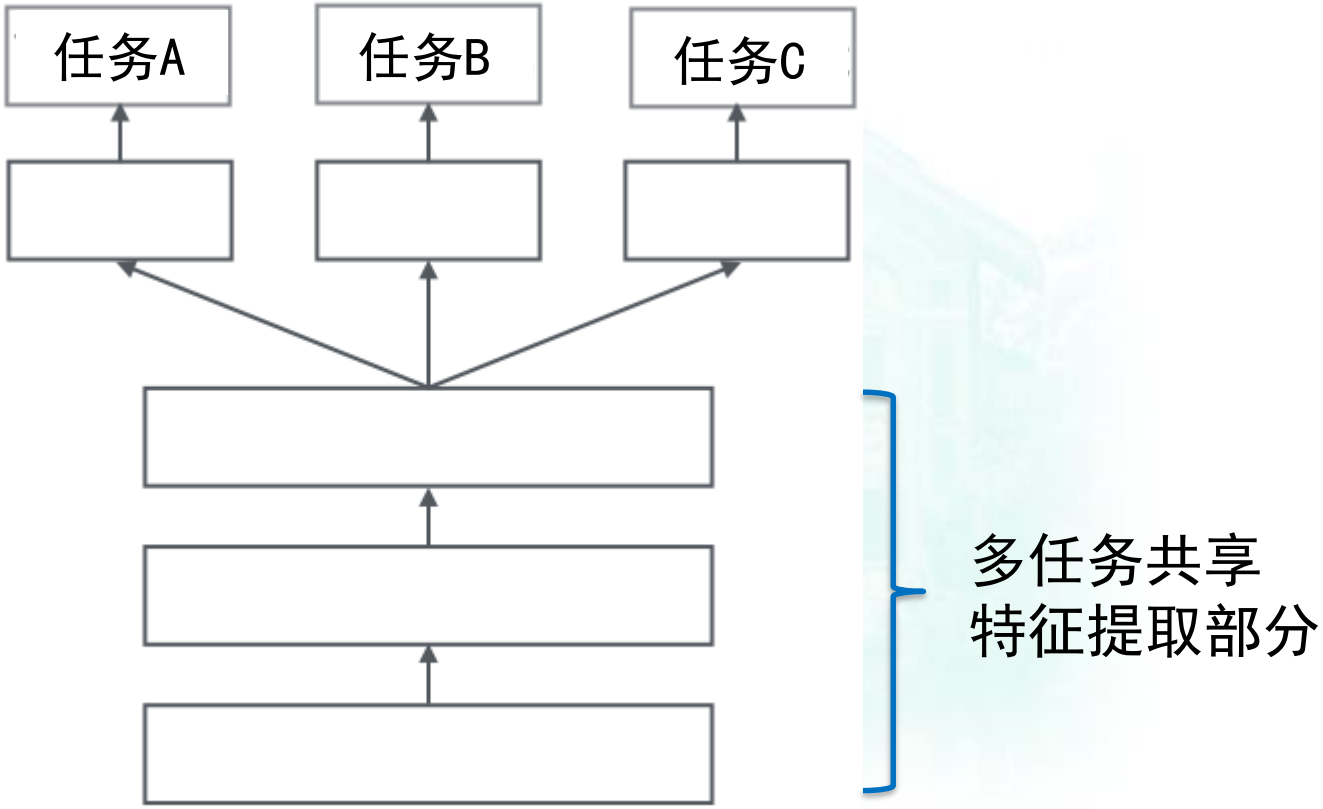
- 迁移学习（一般步骤）
  - 在其它大数据集上预训练一个模型（如分类器）
  - 保留中低层（特征提取）部分，输出端换为与当前任务相关
  - 在当前小样本数据集上对模型进行训练微调

- 迁移学习简单有效，在大量应用中得到证实

- 为什么迁移学习有效？
  - 模型低层部分提取的（纹理、形状）特征具有跨任务特性
  - 需要更新的模型参数较少，减少过拟合

追研究前沿：小样本下深度学习

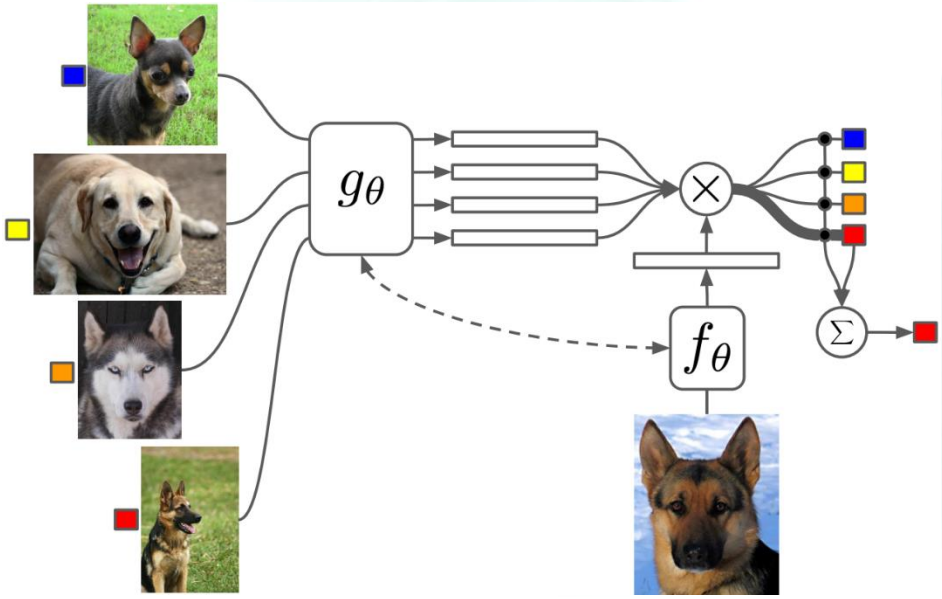
- 多任务学习，使得特征提取部分更具有普适性（对照迁移学习）



追研究前沿：小样本下深度学习

- 元学习(Meta-learning): 训练能生成模型的模型

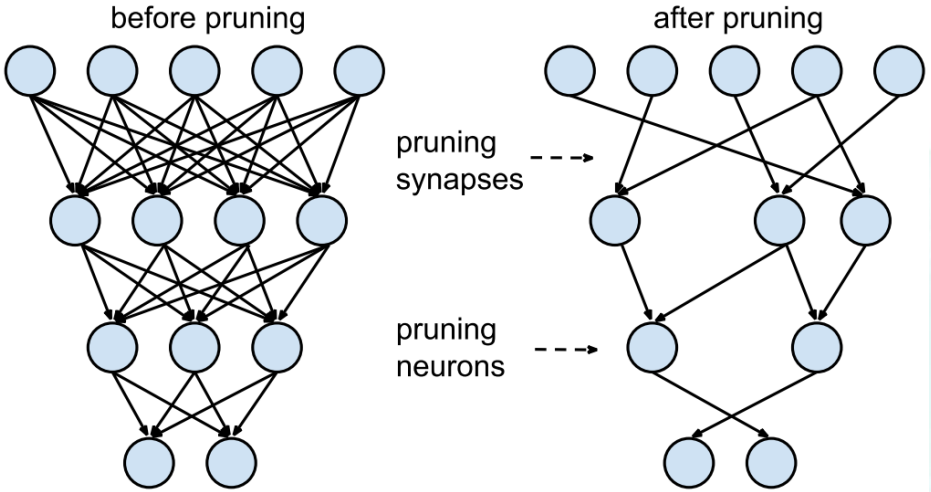
- 传统分类器：输入一个数据，输出数据的类别
- 元分类器：输入一组数据，输出一个分类器

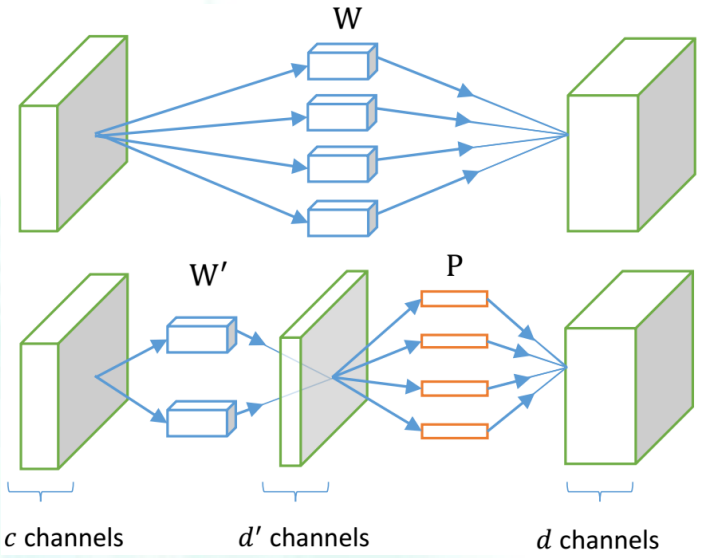


$$\hat{y} = \sum_{i=1}^{k} a(\hat{x}, x_i) y_i$$

追研究前沿:轻量化神经网络模型

- 模型小和/或运算速度快
- 参数少，耗电少，易于部署在移动/嵌入式终端设备

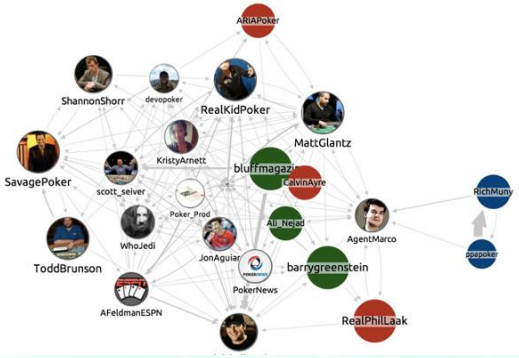


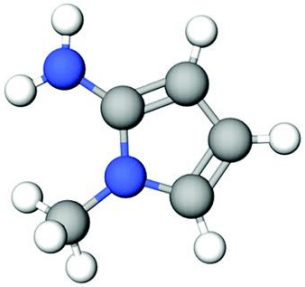


将在模型压缩课程部分作详细介绍！

追研究前沿：图卷积神经网络GCN

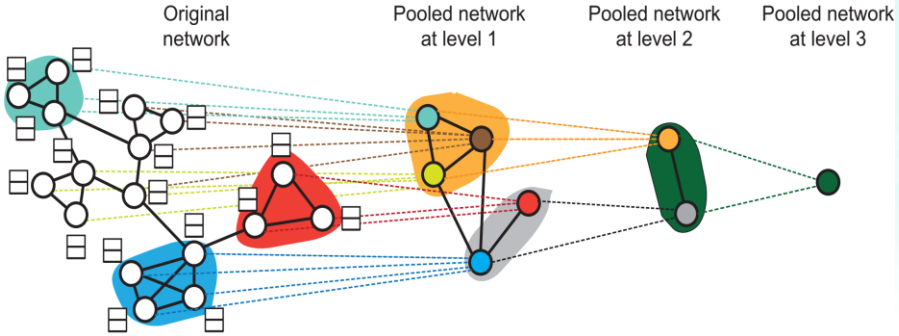
- 数据为图(Graph)；应用于社交网络、推荐系统、分子结构等

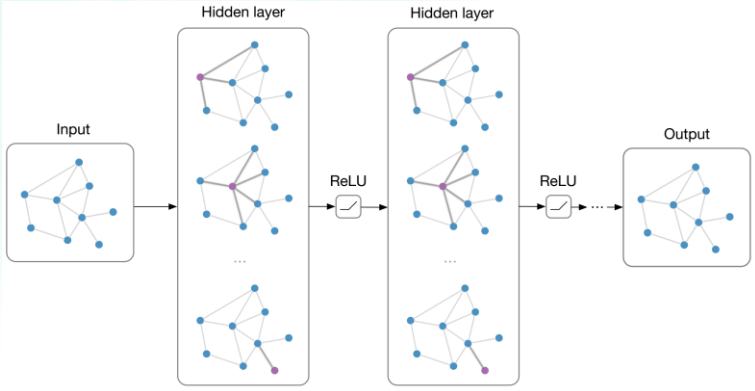




Molecular or crystal structure

- 挑战：如何处理非结构化/不规则下的局部运算/卷积

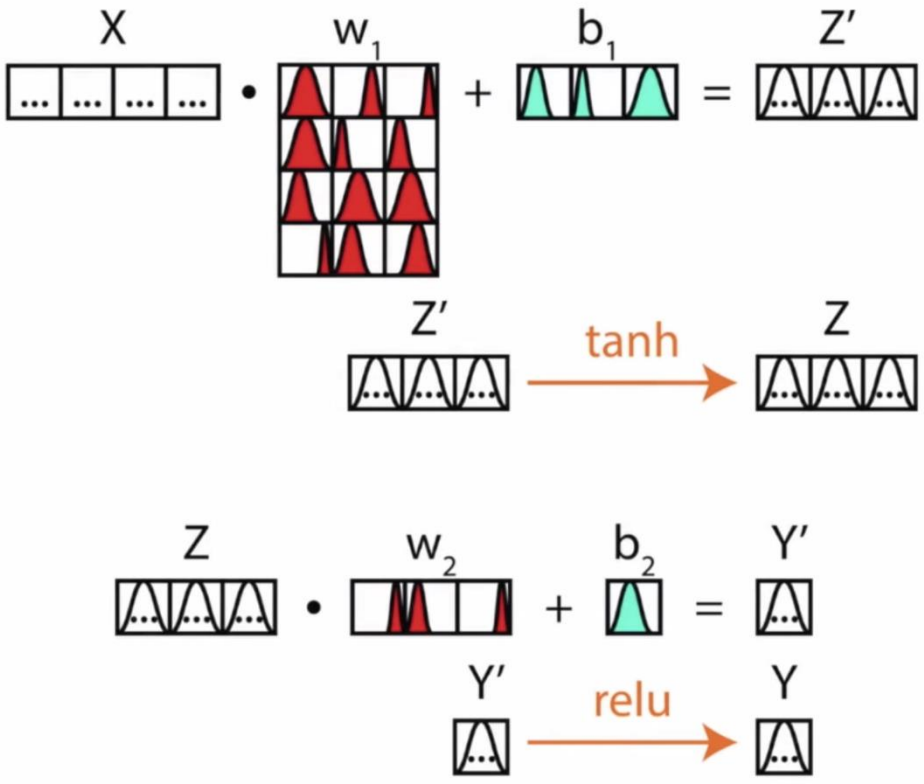




Ying et al., Hierarchical graph representation learning with differentiable pooling, NeurIPS, 2018.
Kipf & Welling, Semi-supervised classification with graph convolutional networks， ICML, 2017

追研究前沿:贝叶斯深度学习

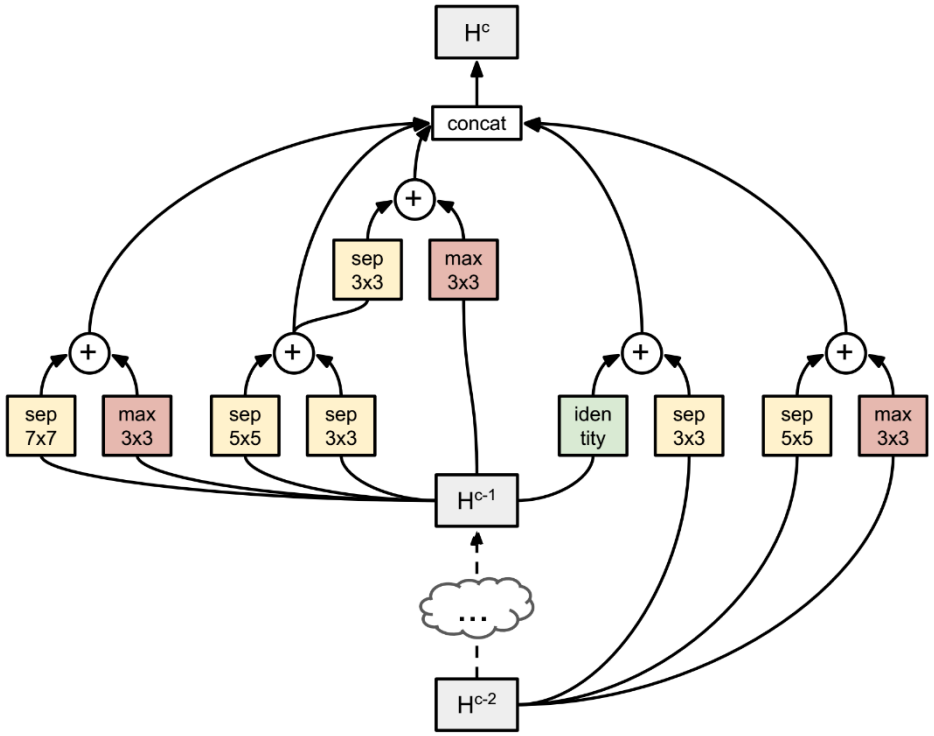
- 学习和预测的不是模型每个参数的具体值，而是分布
- 可以提供预测结果的不确定性（注：softmax输出是预测结果）
- 比非贝叶斯深度学习更加鲁棒（针对输入或参数微小改变）

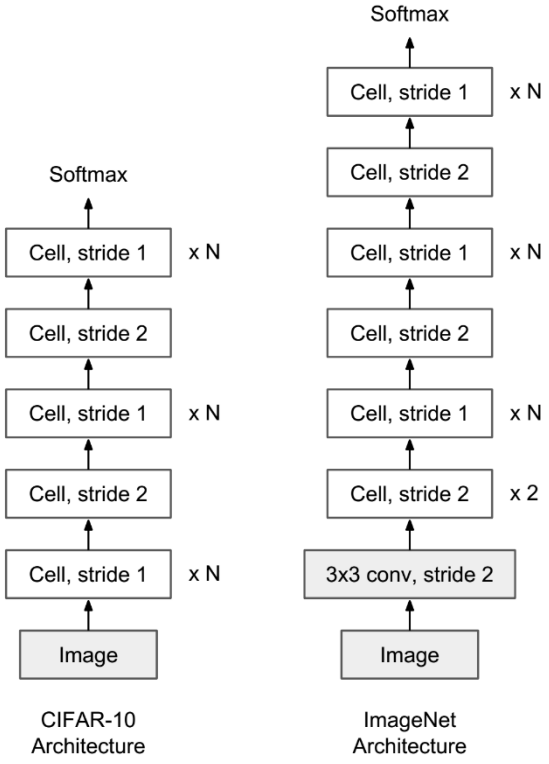


图片来自于https://www.cnblogs.com/wuliytTaotao/p/10281766.html

追研究前沿:自动机器学习AutoML

- 深度学习实现了特征提取自动化，但是模型结构需要人设计
- AutoML更加智能:自动搜索/设计更优的模型结构等超参数
- E.g., NAS, DARTS





当前AutoML的基本操作模块仍然需要人设计，所以下一步...?

追研究前沿:自监督学习

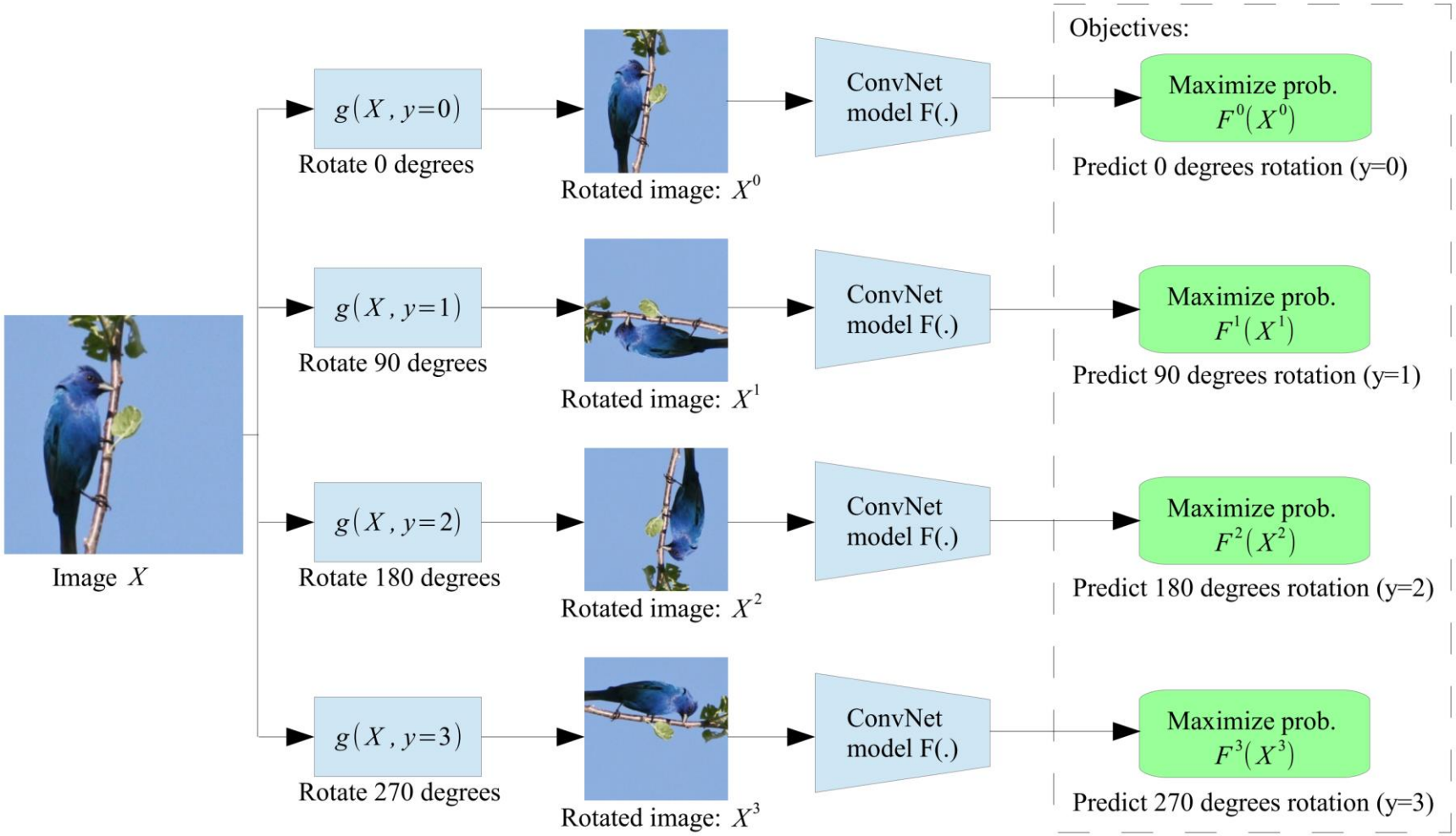
- 一般步骤:
  - 从每一个无标签数据中，通过某种方式自动生成一个/组类别标签和对应的数据;
  - 用这些伪标签数据训练神经网络模型;
  - 用训练好的模型实现对原始数据的有效特征提取和表示

- 为什么有效?
  - 模型必须学习到数据中的特征才能实现较好的伪标签预测

- 为什么要进行自监督学习?
  - 为后续（下游）任务提供较好的数据（初始）特征表示
  - 自监督学习可看作迁移学习的第一步进行模型预训练
  - 后续任务所需标注的数据量大大减少

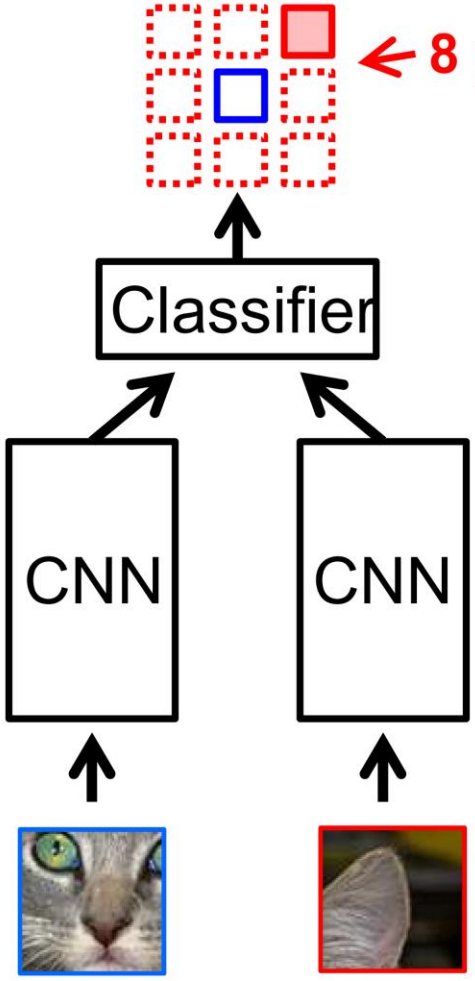
追研究前沿:自监督学习

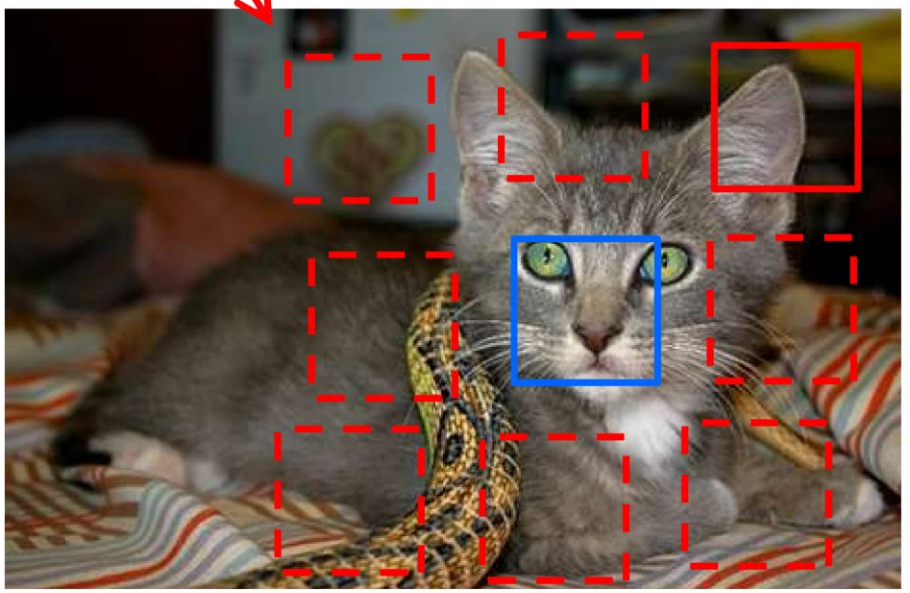
- Rotation net: 训练CNN分类器自动判断图象旋转角度



追研究前沿:自监督学习

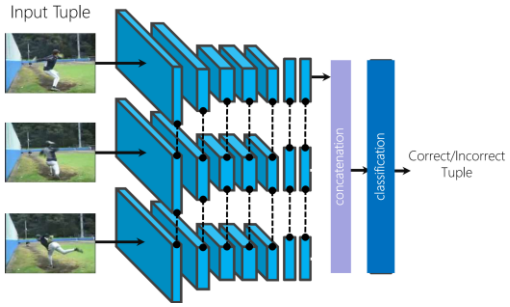
- Location net: 训练CNN预测两个图像块在原图中的相对位置

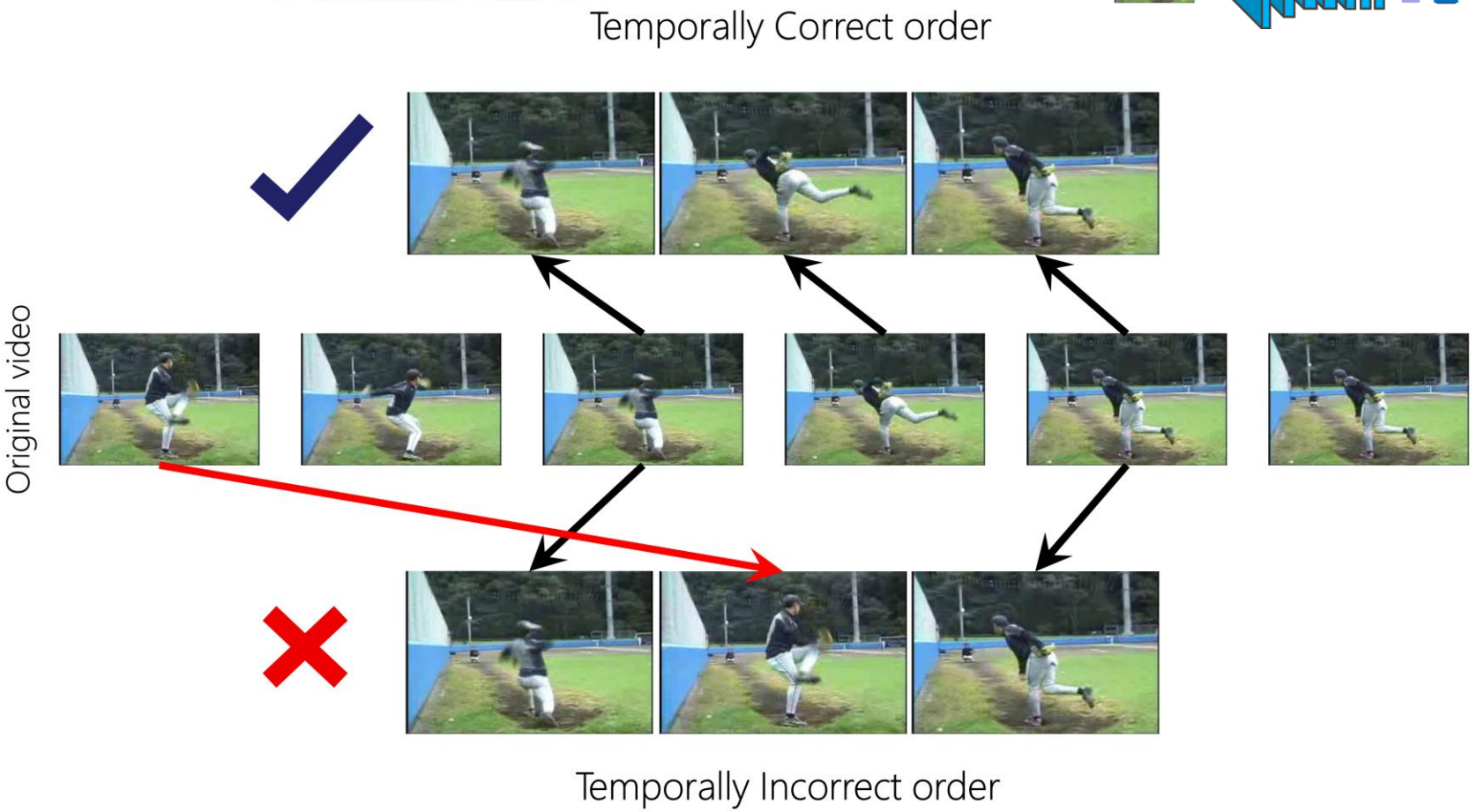




追研究前沿：自监督学习

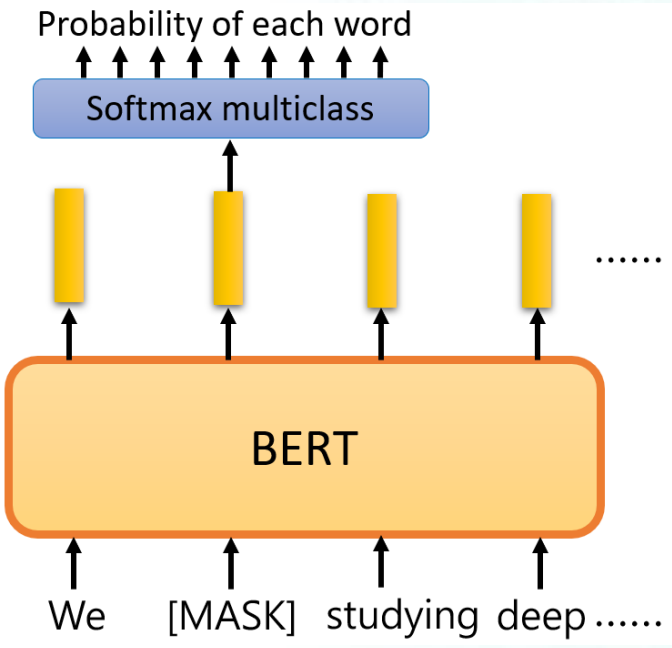
- Shuffle & Learn: 训练模型判断几张图象出现顺序是否正确

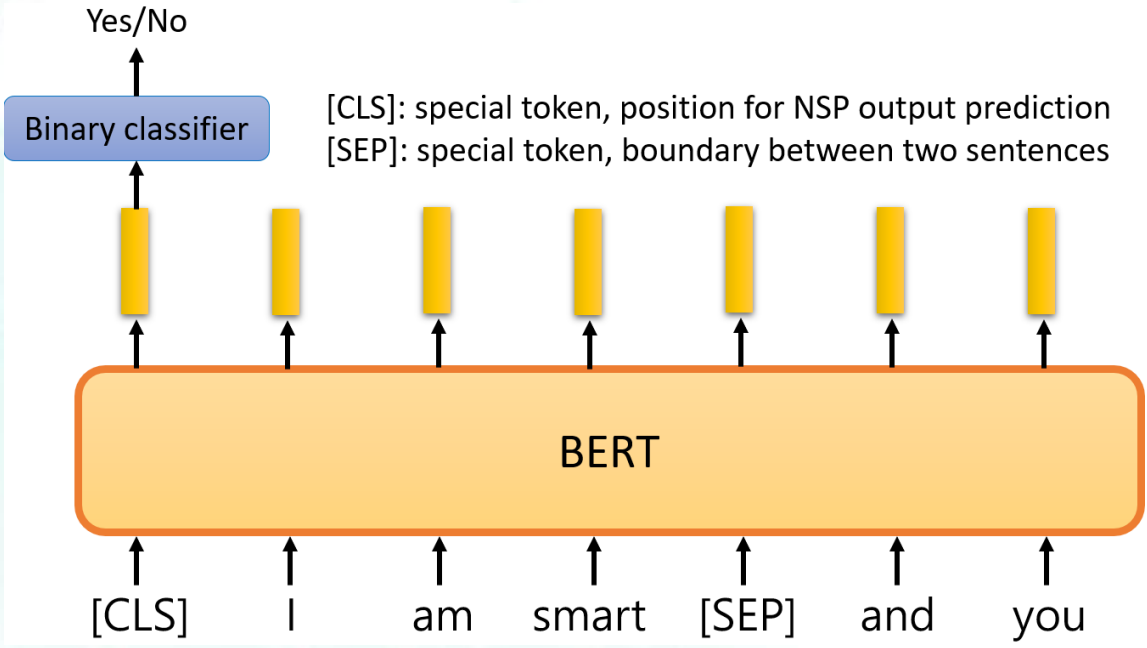




追研究前沿:自监督学习

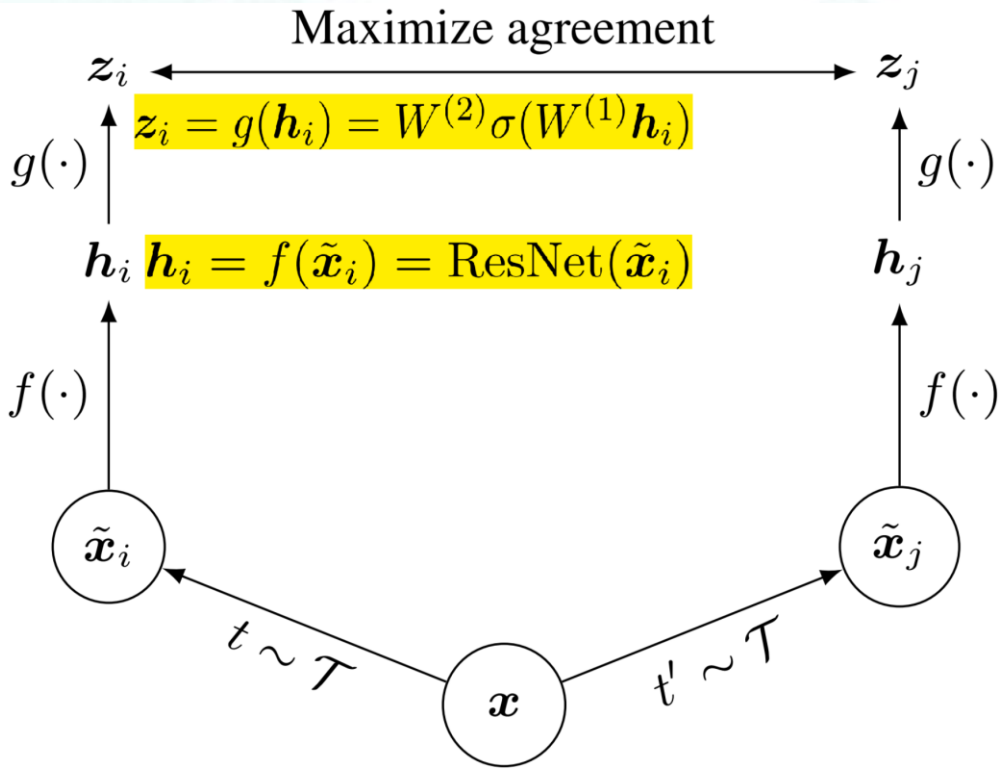
- BERT任务1:训练模型预测句子中被遮挡的单词
- BERT任务2:预测第二句话紧接第一句话出现是否合理





追研究前沿:自监督学习

- 基于对比学习：‘正样本对’ vs ‘负样本对’的特征相似度
- SimCLR:同一张图象变换后所提取的特征相互更相似

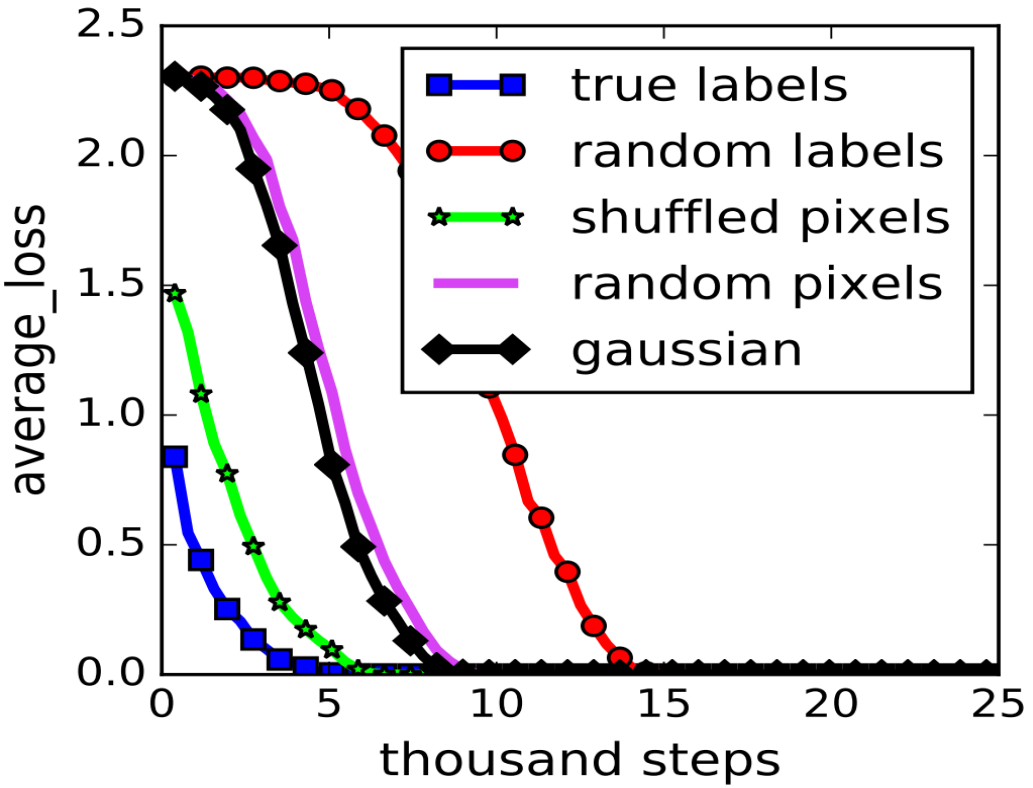


探未解之谜

还有很多问题没答案！

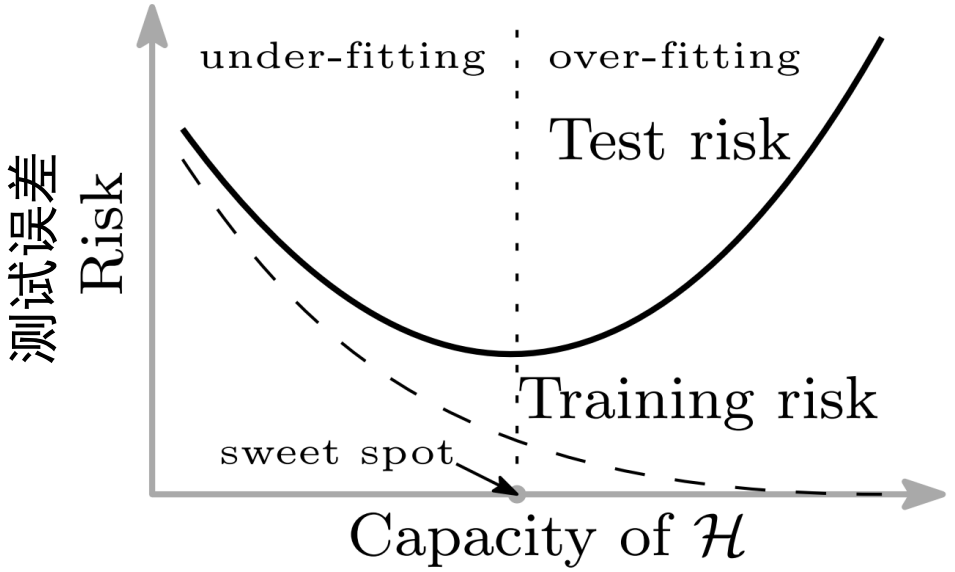
探未解之谜：为什么神经网络没有过拟合？

- CNN在标签随机打乱的数据甚至噪音数据上能被训练到误差为0.
- 即使加上正则项或dropout等操作，CNN也有类似表现！



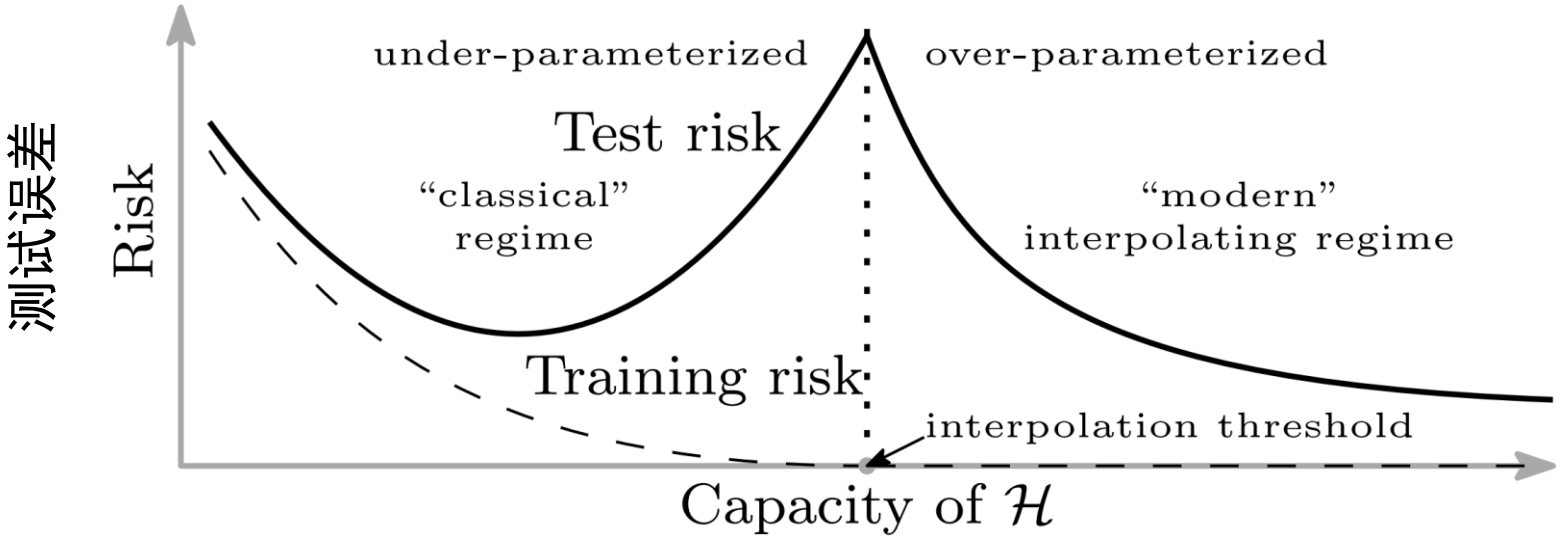
揭示：CNN在分类任务中表现好，可能不是因为各种正则化，而是来自于CNN对所有训练数据的记忆。

探未解之谜：传统机器学习理论不再适用？



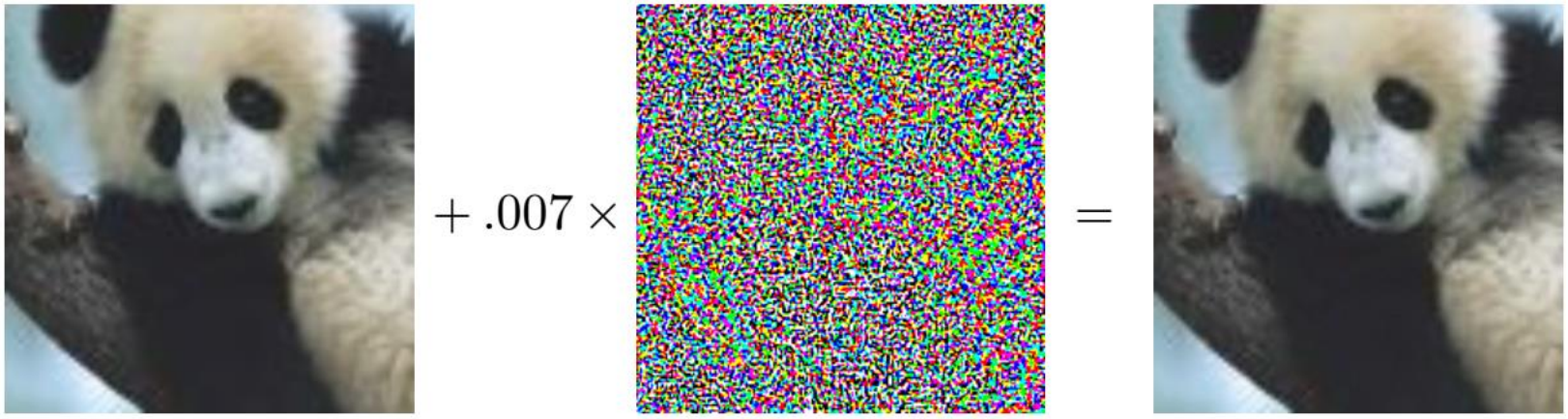
传统机器学习理论

模型真实表现



探未解之谜：鲁棒性/安全性

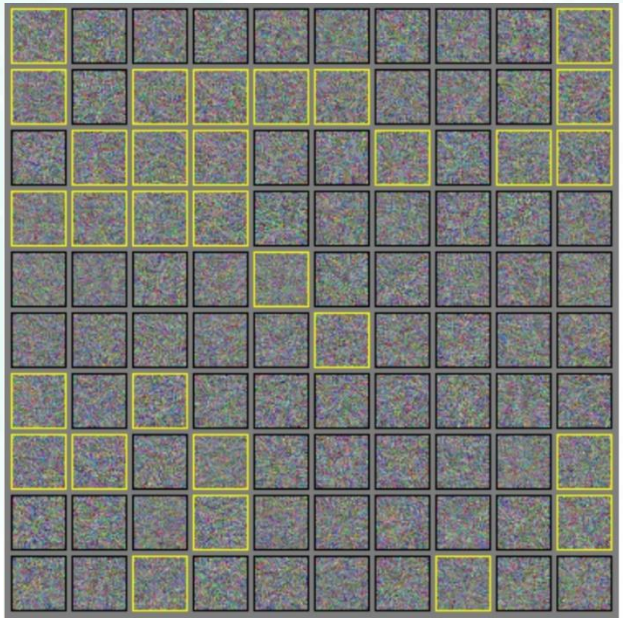
- 模型的安全性很差：在任一图像中加入特定的人眼感觉不到的特定噪音（对抗噪音），会导致模型的性能急剧下降！



正常图像　　对抗噪音　　对抗样本

探未解之谜：鲁棒性/安全性

- 模型攻击基本思路：如果知道模型参数，通过梯度上升法改变原始输入图像，使模型对更改后的图像（对抗样本）预测错误
- 对模型攻击，可以让分类器准确率降到10%以下！
- 也可以让模型将噪音图像误认为特定一类物体！



FGSM方法让模型将以上黄色框内图像识别为“Airplane”！

探未解之谜：鲁棒性/安全性

- 即使不知道待攻击模型的参数，也可以通过攻击其它（已知模型参数）模型产生对抗样本，用这些样本攻击目标！黑盒攻击！
- 黑盒攻击更常见；白盒攻击（知道待攻击模型参数）更厉害！

如何防御这些攻击?

- 训练模型时将对抗样本纳入训练数据 → 难于有效防御多种不同的攻击方法
- 设计模型使得对抗样本难于产生 → 不能对训练好的模型进行防御；容易被新方法攻击
- 设法去除对抗样本中的（对抗）噪音

↓

不用考虑分类器的结构与训练；
不用考虑攻击方法的类型

探未解之谜：模型可解释性

- 深度学习模型是个“黑盒子”？
  - 只有预测结果，没有预测详细过程
  - 原因：模型是一个复杂非线性变换

- 为什么要打开这个黑盒子？
  - 低于人类表现时：帮助人们发现模型的不足
  - 与人类表现持平：增加人们对模型预测的信任感
  - 超过人类表现：教人类如何更好地预测

探未解之谜：解释模型特定输出神经元

- 从输入端寻找可解释性：对于训练好的模型，寻找最佳输入图像 $\mathbf{I}$，使得特定pre-softmax输出$f_c(\mathbf{I})$最大

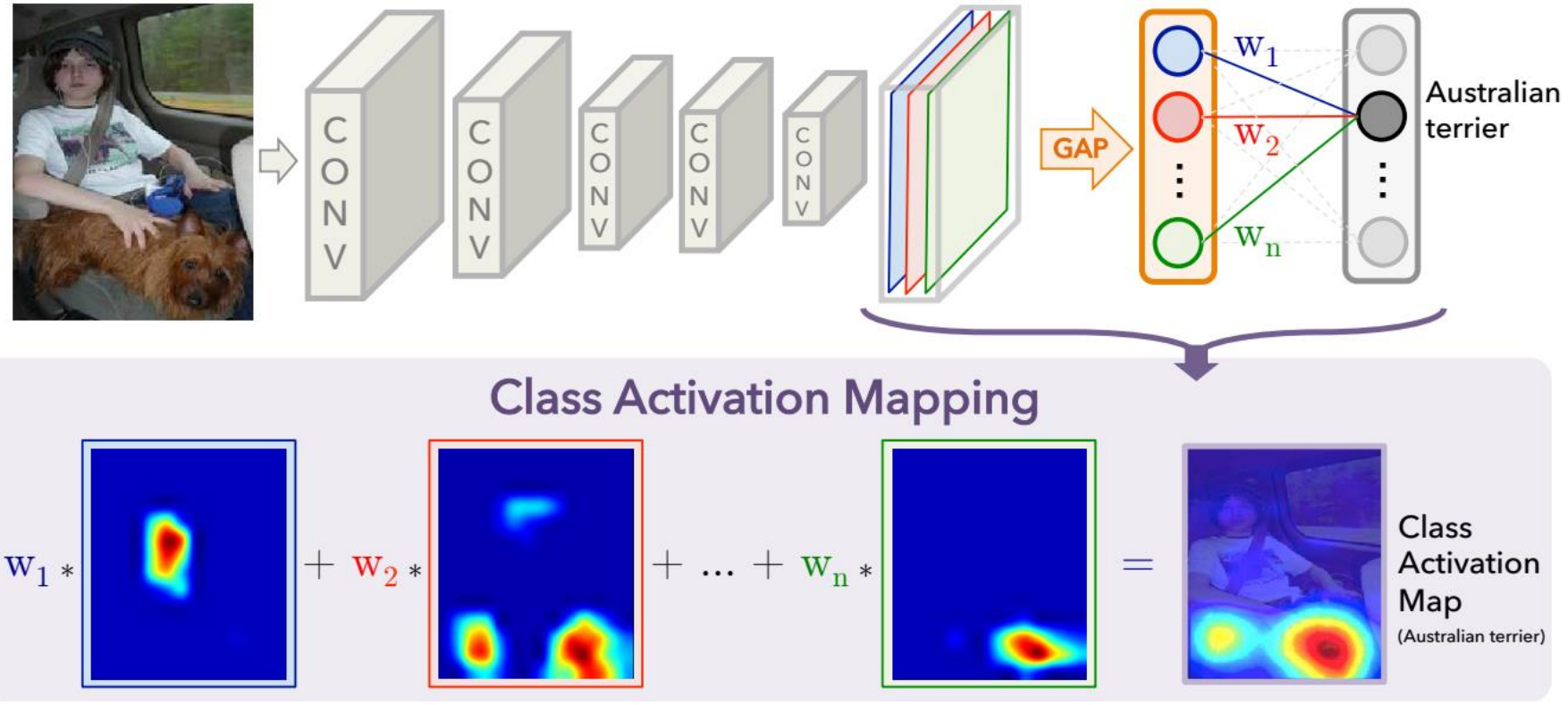
$$\arg\max_{\mathbf{I}} f_c(\mathbf{I}) - \lambda \|\mathbf{I}\|^2$$

dumbbell　　cup　　dalmatian

一定程度上帮助人们理解每个输出神经元在关注什么样的视觉特征，但无法解释对真实图像预测时模型在关注什么。

探未解之谜：解释模型的每次预测

- 从靠近输出端寻找可解释性：最后卷积层输出特征图的特定线性组合可解释模型预测时关注的图像区域



探未解之谜：解释模型的每次预测

- LRP (Layer-wise relevance propagation):将特定输出一层层分解到输入端，得到每个像素对最终预测的贡献大小

input x

Classifier

Rooster

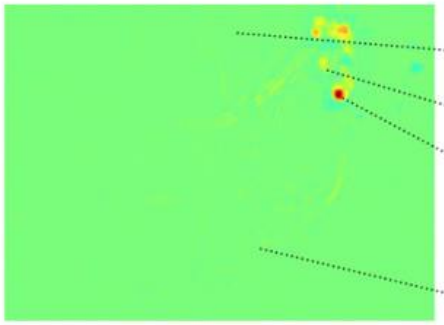
prediction $f(x)$

Explain prediction

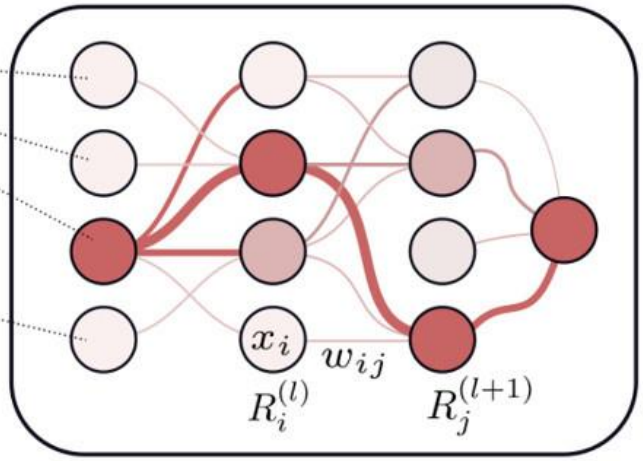
(how much each pixel contributes to prediction)

Idea: Decompose function

$$\sum_i R_i = f(x)$$



heatmap

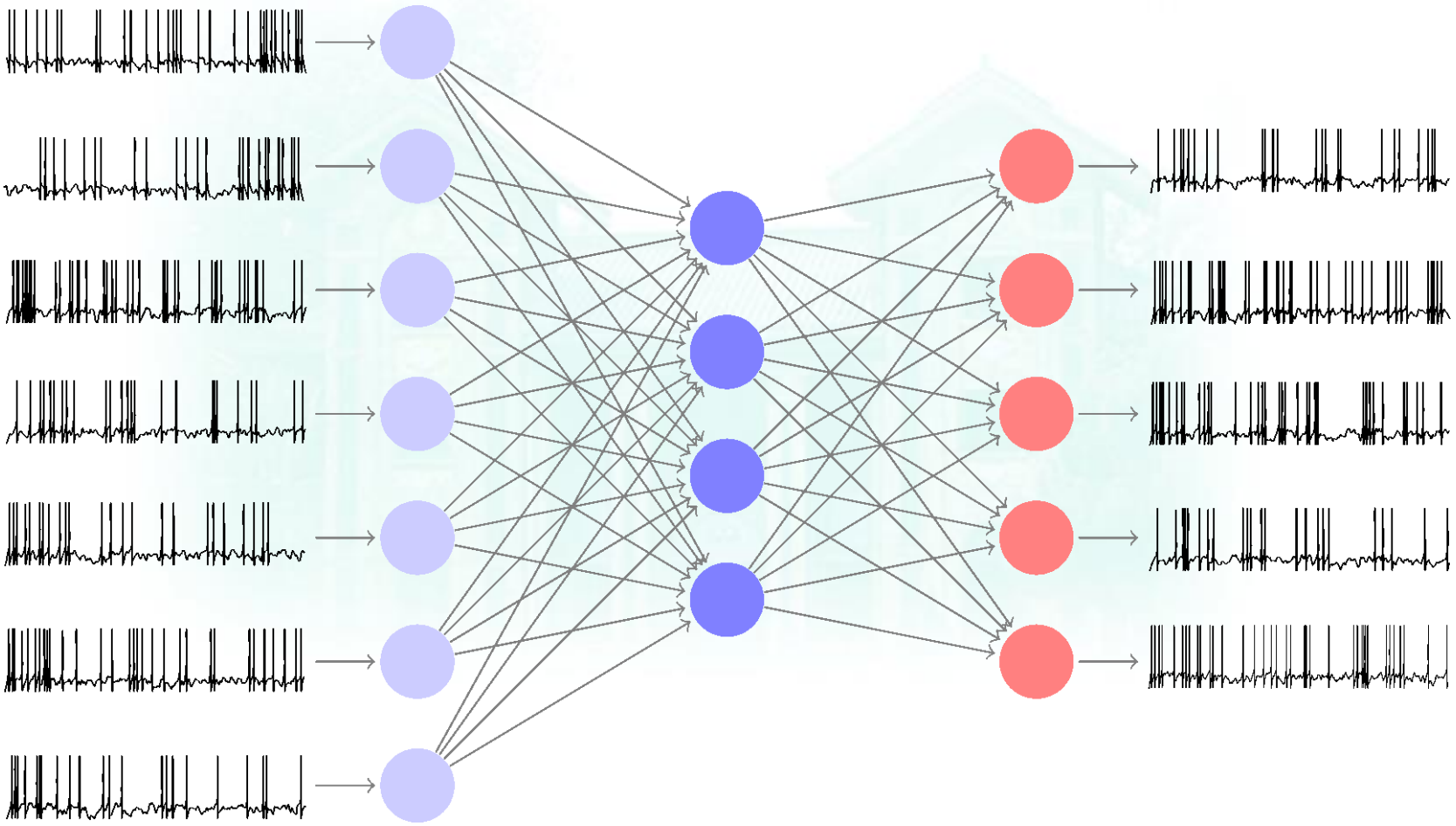


"every neuron gets it's share of relevance depending on activation and strength of connection."

redistribute $f(x)$

探未解之谜：脉冲神经网络

- 人类大脑神经元活动：脉冲序列；如何学习和更新？
- Spiking neural network: 更像大脑神经元，但有监督训练较困难



探未解之谜：认知与推理

- 深度学习目前解决的是Perceptual AI问题
- 还不具有真正认知和推理能力（Cognitive AI）
  - 目前：基于大数据建立的输入–输出统计关系
  - 失效：少见/极端情况下犯简单错误

- 探索方向
  - 基于符号的逻辑推理?
  - 将推理过程嵌入模型？
  - 如何连接认知与推理？

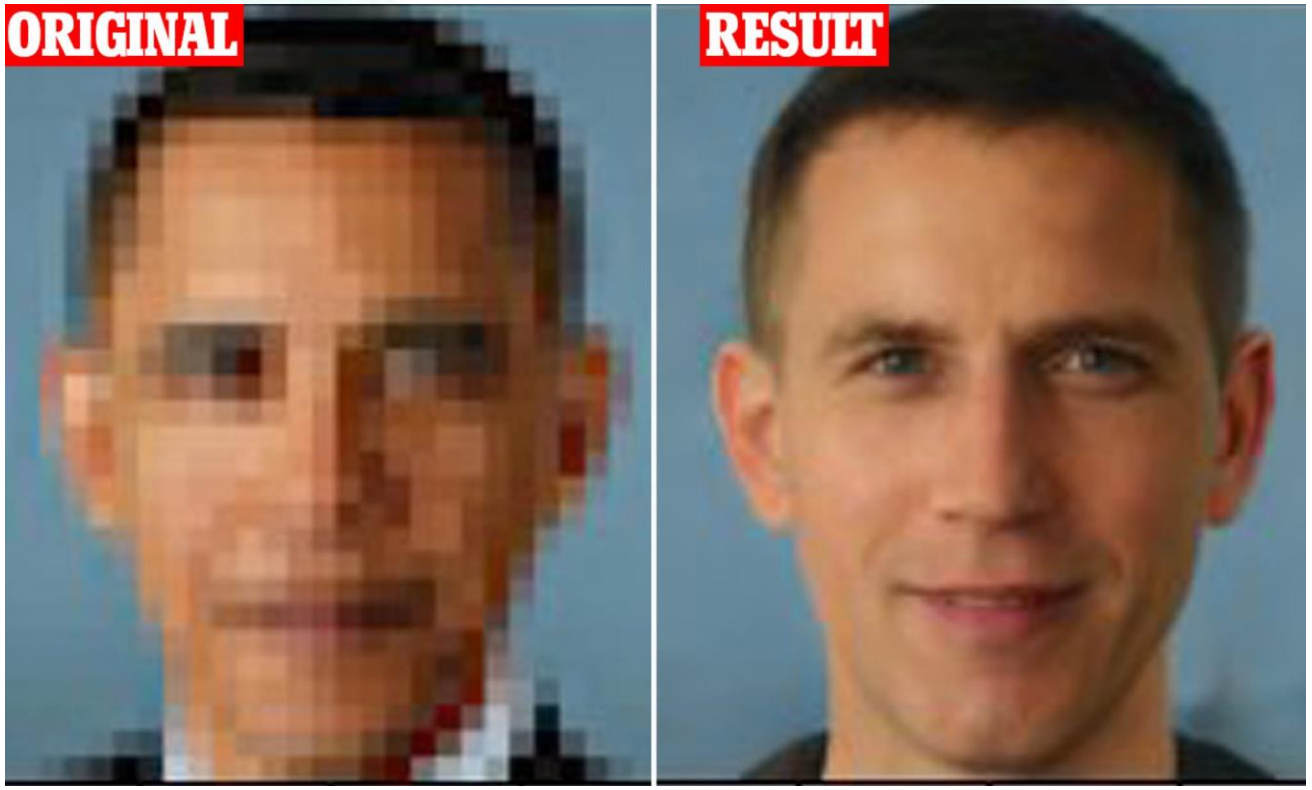
探未解之谜：伦理与公平

- AI伦理
  - 安全可控(AI-人类)：如何确保AI机器人/算法绝对可控？
  - 人机融合：半人半机，思想决策被AI影响/控制
  - AI武器(国家-国家/团体/个体)：谁决定谁能/不能使用，如何有效限制；AI蜜蜂寻找攻击特定目标…
  - 自动驾驶事故（公司-个体）：AI算法提供商，汽车制造商，出租公司，乘客各负什么责任？保护乘客还是行人？
  - AI导致失业：谁之错？如何用AI帮助失业者？

超越了技术本身，需要多学科、多职能部门参与

探未解之谜：伦理与公平

- AI公平
  - 算法偏见：人脸识别算法对白种人性能好，对黑种人差！
  - 如何保证不同性别、地区、民族、国家的人们能同等接受AI的教育或利用AI帮助学习、工作和生活？



评大众观点

- 深度学习只不过是在调参
- 没新意，只是数据更多、计算更快
- 深度学习只是工具和工程方法
- 深度学习需要大数据和人工标注
- 深度学习发展已经到顶
- 深度学习真正落地应用的例子很少

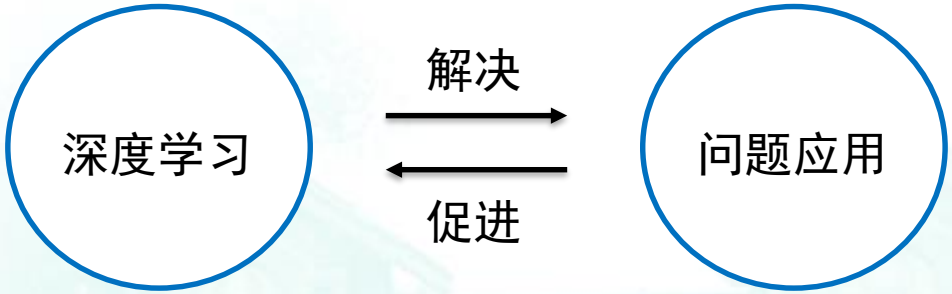
模型算法创新
思想百花齐放
安全解释之谜
前沿挑战多样
产业应用井喷
莫学无畏儿郎

- AI理解你的所思所想
- 机器翻译已超越人类翻译
- AI将很快超越人类智能

哗众取宠
以偏概全
难越天堑

结论

- 深度学习与应用



- 深度学习的研究大大促进了AI的发展与应用
- 深度学习是一种思想，而不仅仅是一类方法
- 还有很多问题和挑战需要被攻克

最终：人类的还是AI的？

人-机鸿沟（The Human-AI Gaps）：

真认知（Conscious）

真情绪（Emotional）

真体验（Embodiment）

真生死（Limited life）