



人工智能与医学

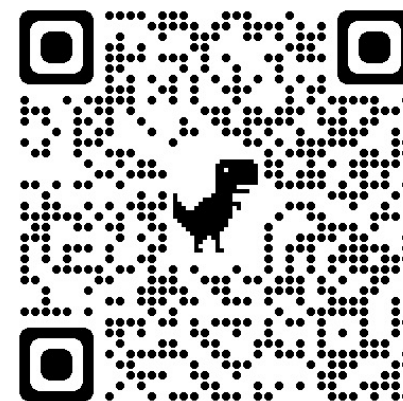
-计算机视觉算法

向孙程

主页: <https://JeremyXSC.github.io/>

上海交通大学医学院课程

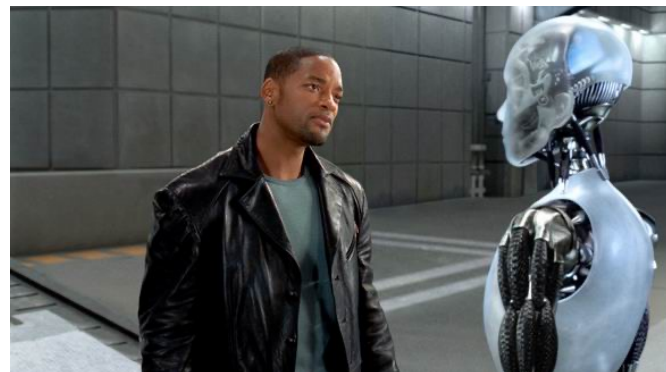
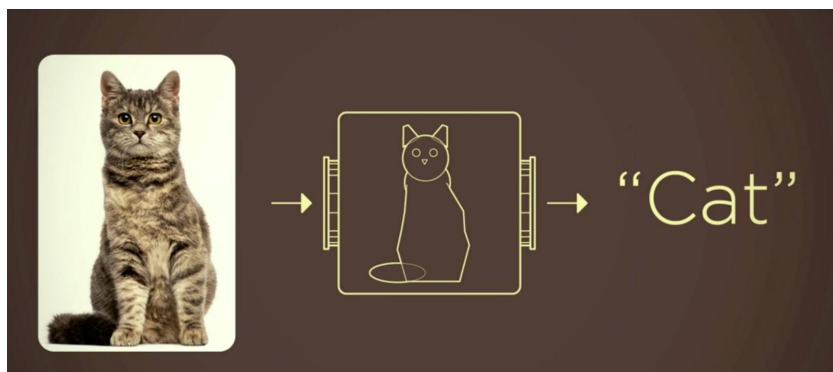
2025春季





计算机视觉定义

- 计算机视觉是使用计算机及相关设备对生物视觉的一种模拟。它的主要任务就是通过对采集的图片或视频进行处理以获得相应场景的认知信息——让机器代替人感知世界。





工业界应用领域





工业界应用领域





学术界受关注程度（数据库和竞赛）

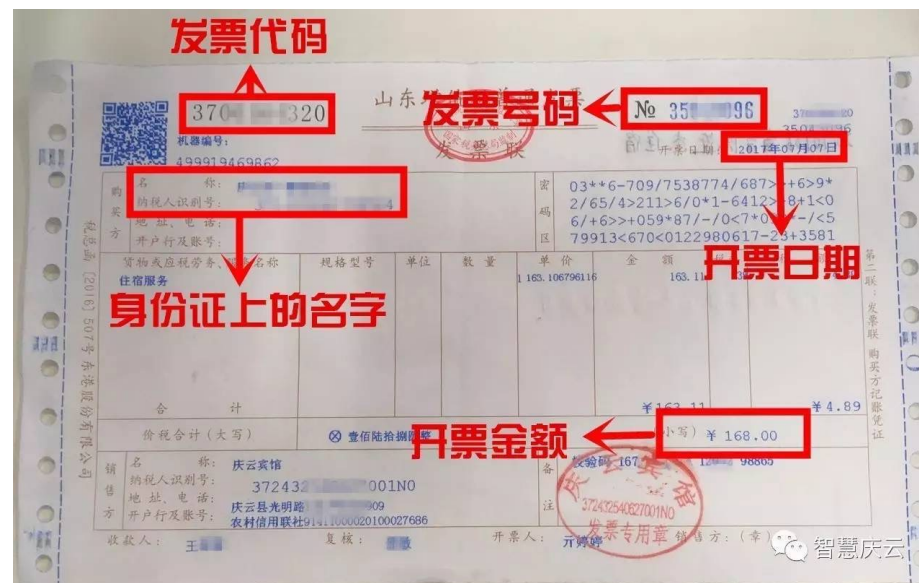
- ImageNet: 斯坦福计算机科学家李飞飞模拟人类视觉系统建立的数据库
<https://image-net.org/>
- ILSVRC Challenge
<https://image-net.org/challenges/LSVRC/>
- MIMIC: 医学报告多模态数据库
<https://physionet.org/content/mimic-cxr-jpg/2.0.0/>
- HUST-19: 新冠肺炎CT影像和临床信息数据库
<http://ictcf.biocuckoo.cn/HUST-19.php>



计算机视觉应用：光学字符识别 OCR



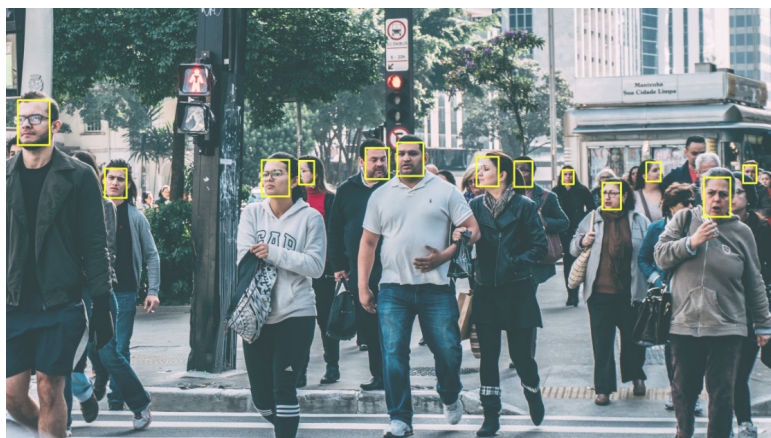
车牌识别



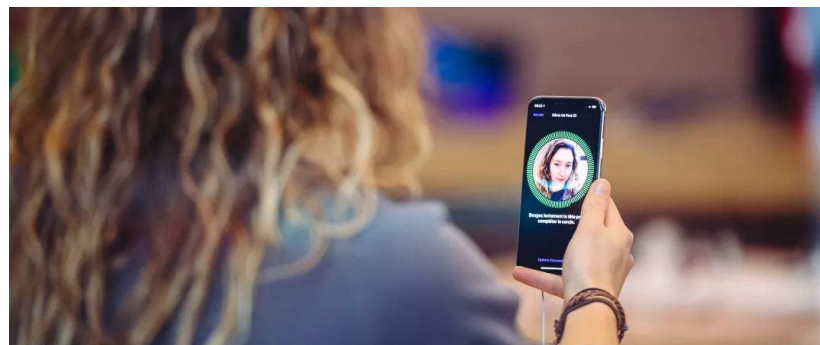
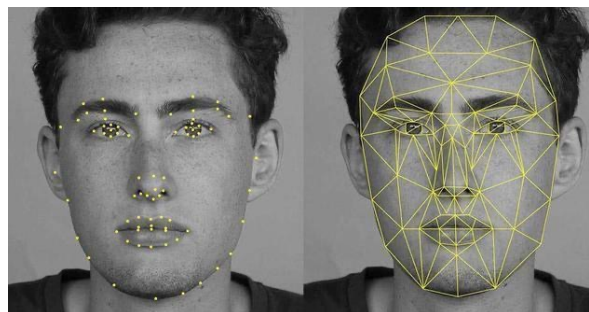
发票识别



计算机视觉应用：人脸检测和识别



人脸检测



人脸识别



计算机视觉应用：图像修复



图像自动修复和上色



计算机视觉应用: 妆容迁移



图像自动修复和上色



计算机视觉应用: 图像高分重建

bicubic
(21.59dB/0.6423)



SRResNet
(23.53dB/0.7832)



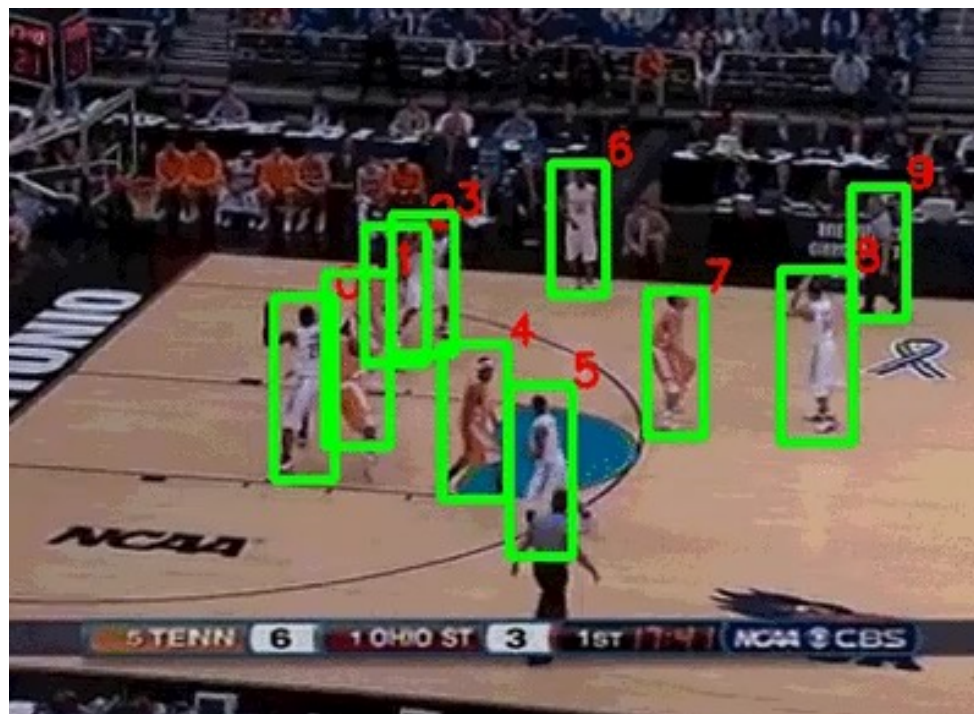
SRGAN
(21.15dB/0.6868)



超分辨率重建



计算机视觉应用：体育解说



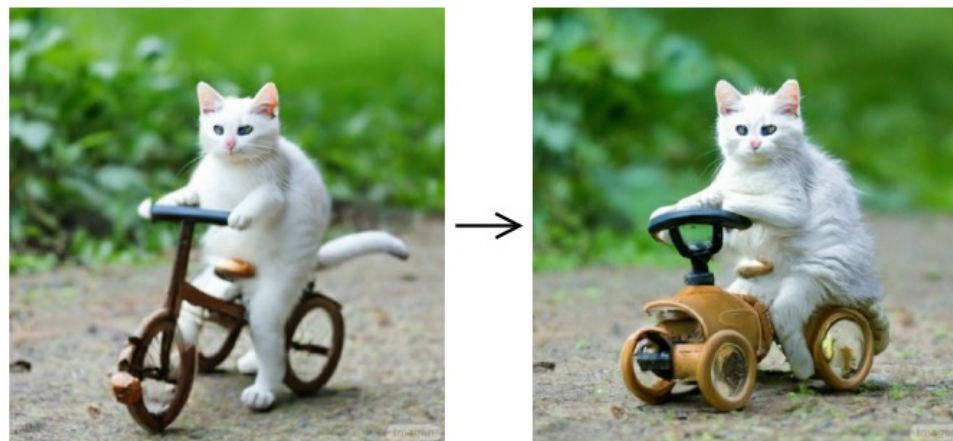
NBA赛事自动解说



计算机视觉应用：图像生成



AI试衣



“Photo of a cat riding on a ~~bicycle~~.
car”

AI绘画



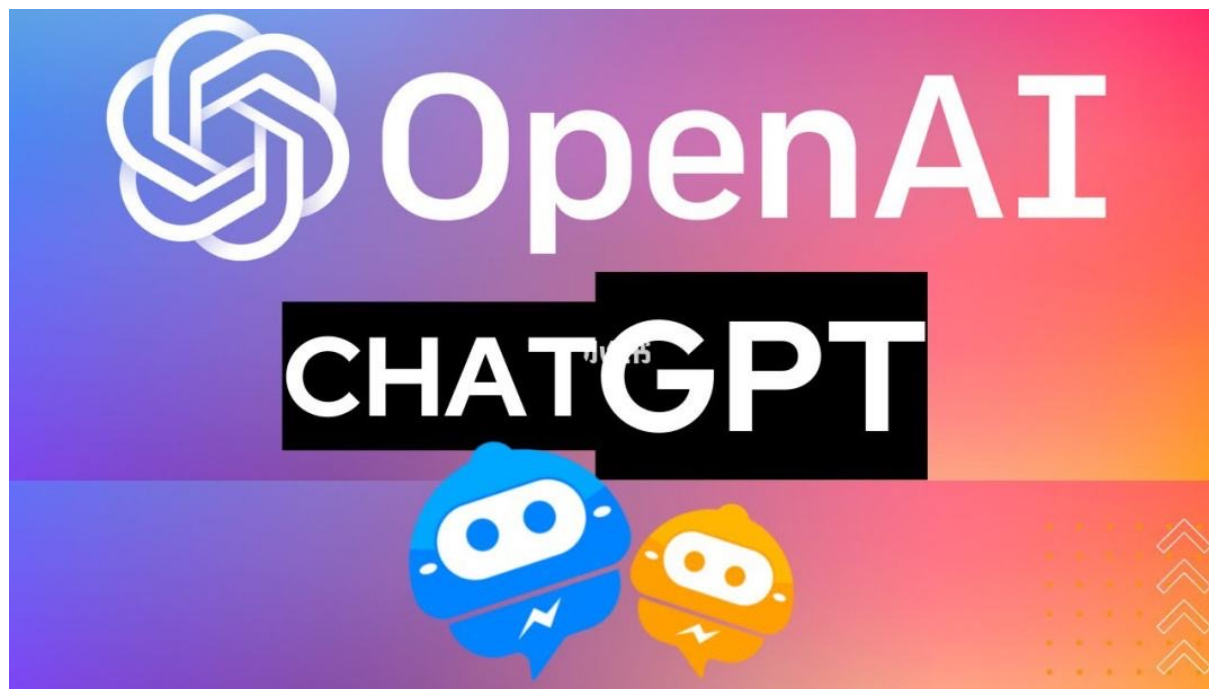
计算机视觉应用:动作识别和跟踪



动作实时检测和跟踪



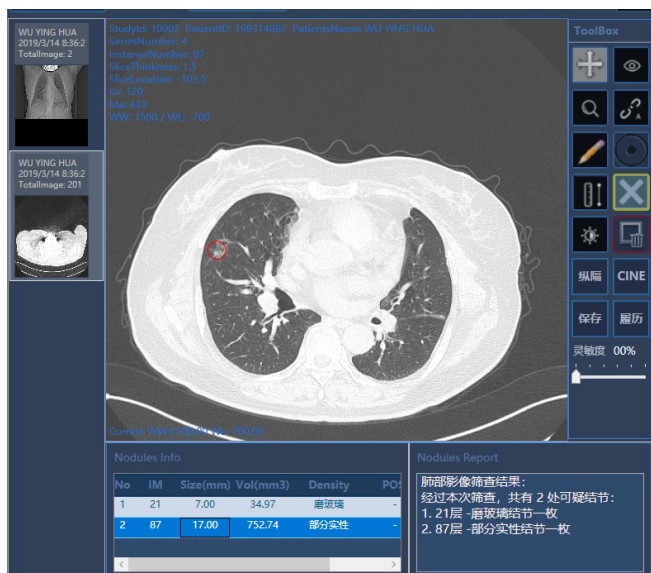
智能聊天机器人



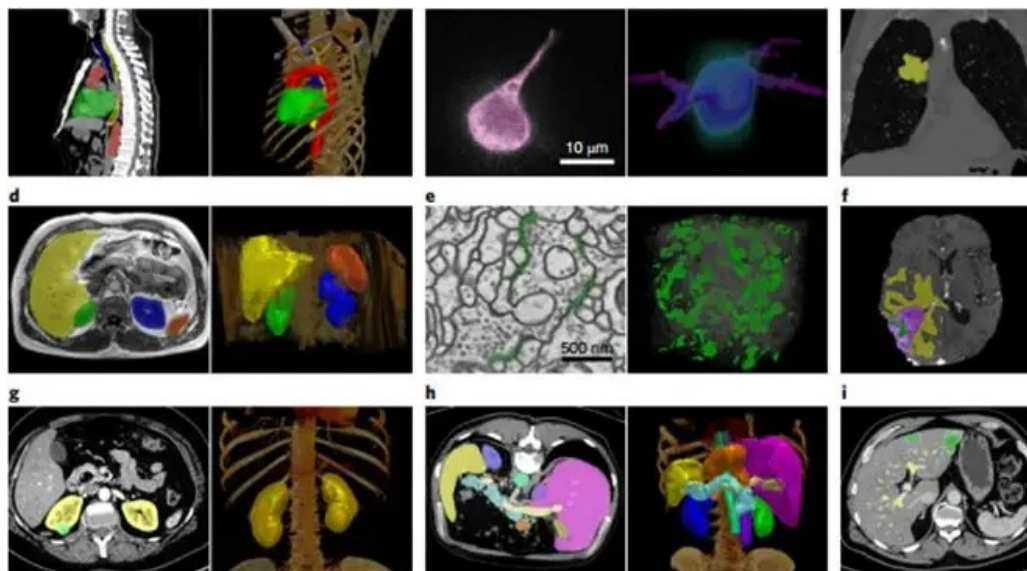
ChatGPT



计算机视觉应用：医学检测与分割任务



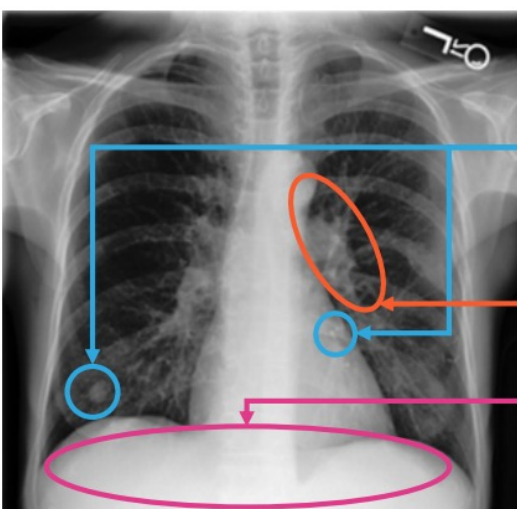
肺结节检测



多器官分割



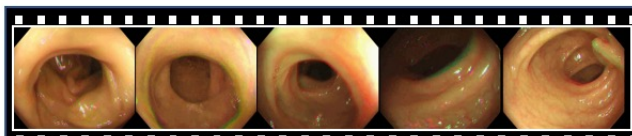
计算机视觉应用：诊断报告生成

	<u>Findings</u> <i>There is no focal consolidation, pleural effusion or pneumothorax. <u>Bilateral nodular</u> opacities that most likely represent nipple shadows. The cardiomeastinal silhouette is normal. Clips project over the <u>left lung</u>, potentially within the breast. The imaged <u>upper abdomen</u> is unremarkable.</i> <u>Impression</u> <i>No acute cardiopulmonary process.</i>
---	---



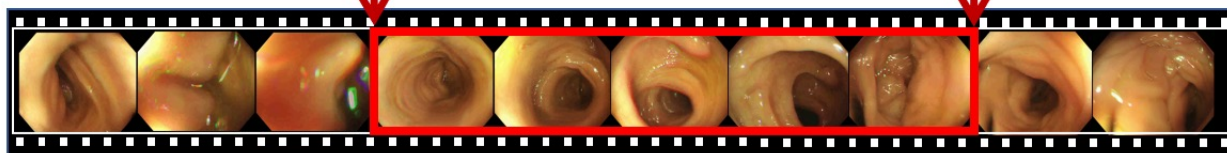
计算机视觉应用:结肠息肉检索

查询视频片段（取自第一次肠镜检查） 文本（节选自肠镜诊断报告单）



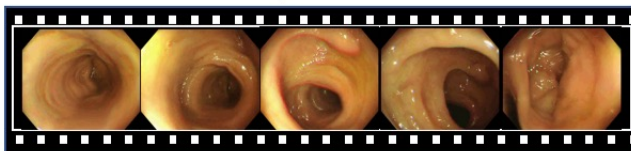
退镜见乙状结肠距肛25cm处一个
息肉，直径1.2cm，表面粗糙。

重定位



第二次肠镜检查视频

匹配结果



匹配视频片段（取自第二次肠镜检查）



机器学习/深度学习 (Machine Learning/Deep Learning)



学习资料推荐

经典书籍：

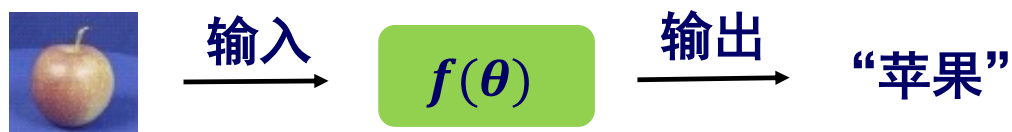
- ① 周志华 《机器学习理论导引》 机械工业出版社
- ① Richard O. Duda 《模式分类》 机械工业出版社
- ① Ian Goodfellow 《深度学习》 人民邮电出版社

在线资源：

- ① 斯坦福大学计算机视觉课程cs231n
<http://cs231n.stanford.edu/schedule.html>
- ① 李宏毅免费机器学习课程
<https://aistudio.baidu.com/aistudio/lhycourse>
- ① 深度学习CV从入门到实战
<https://aistudio.baidu.com/aistudio/course/introduce/789>



什么是机器学习

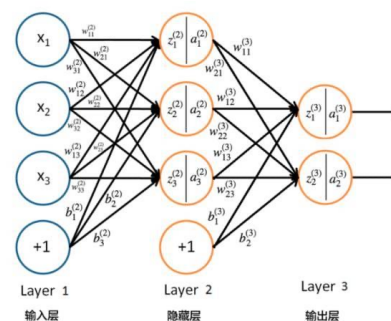


机器学习的目的：让计算机从一系列函数中找出一个最优函数，用于实现特定的任务，比如分类和回归。

这个函数一般称为模型(Model)，由架构（线性、网络）和参数构成。

$$y = wx + b$$

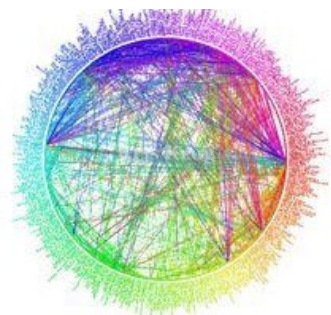
简单的线性模型



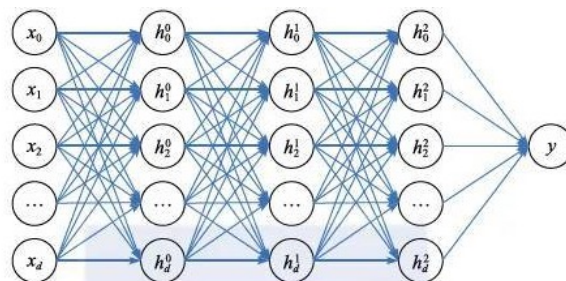
复杂的神经网络模型



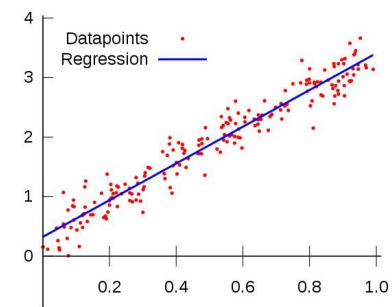
基本任务：回归和分类



数据/特征

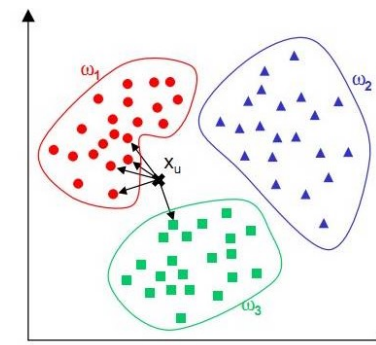


机器学习模型
(SVM, RF, ANN, CNN)



回归

输出连续的
结果
(房价)



分类

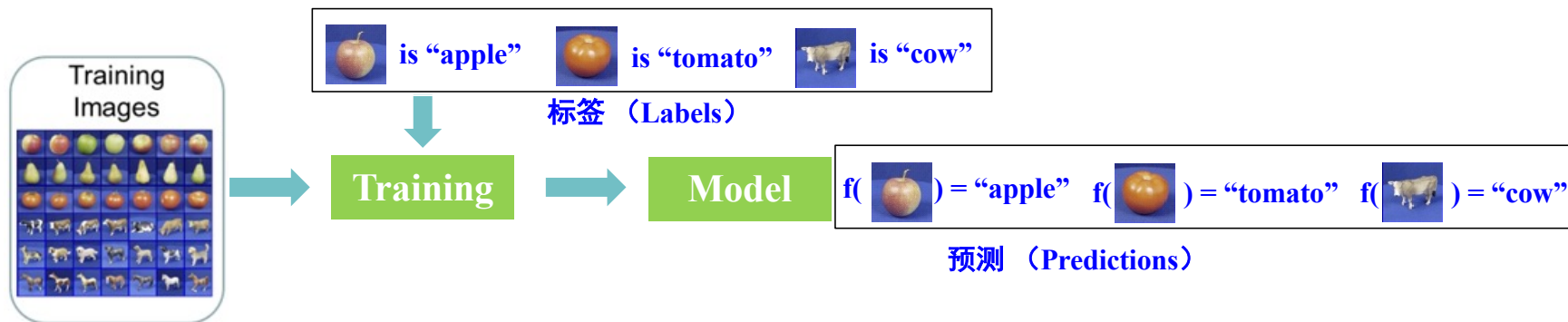
输出离散的
结果
(类别)

图像模型主要是分类任务，分割、检测本质上也是分类的应用

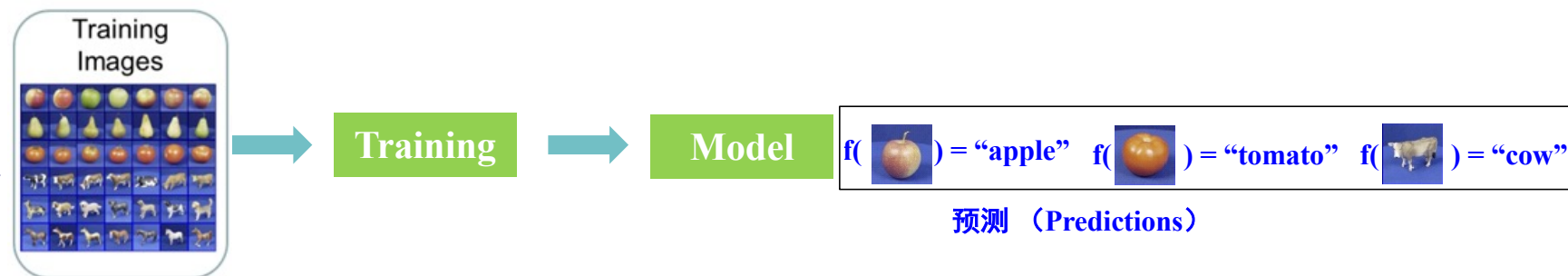


有监督和无监督学习

有监督
supervised



无监督
unsupervised



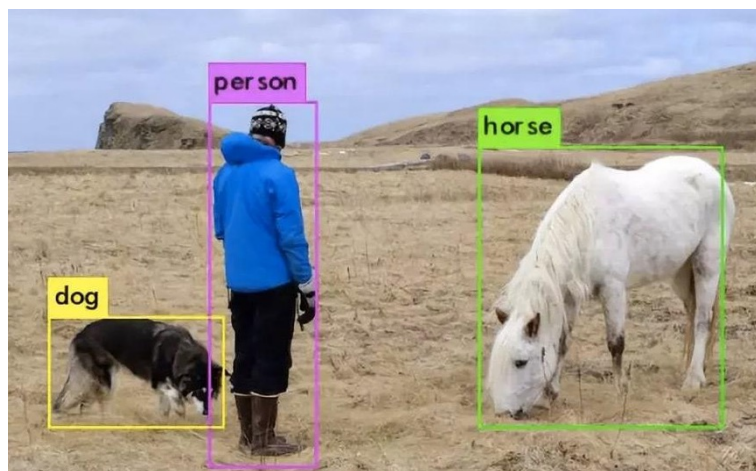
- 有监督学习：数据集需要有标签信息。主流。
- 无监督学习：数据集不需要标签信息。难，应用场景简单。



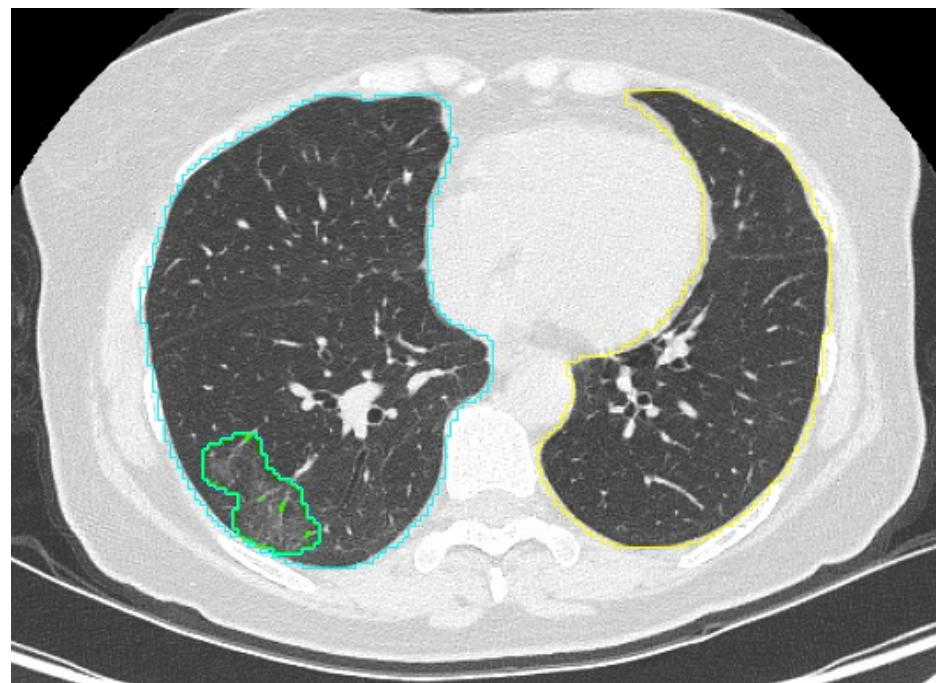
数据标注



图像级标注 (Image-level annotations)



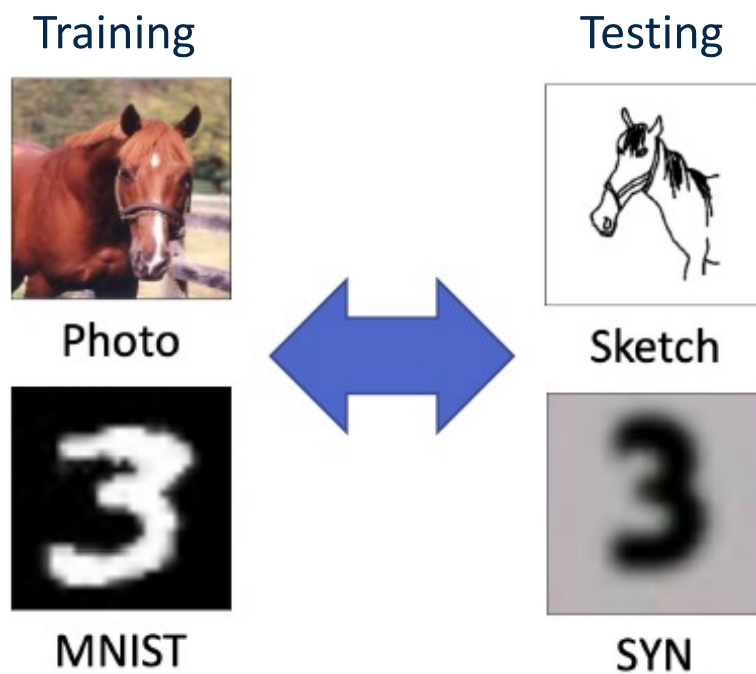
目标级标注 (RoI-level annotations)



像素级标注 (Pixel-level annotations)



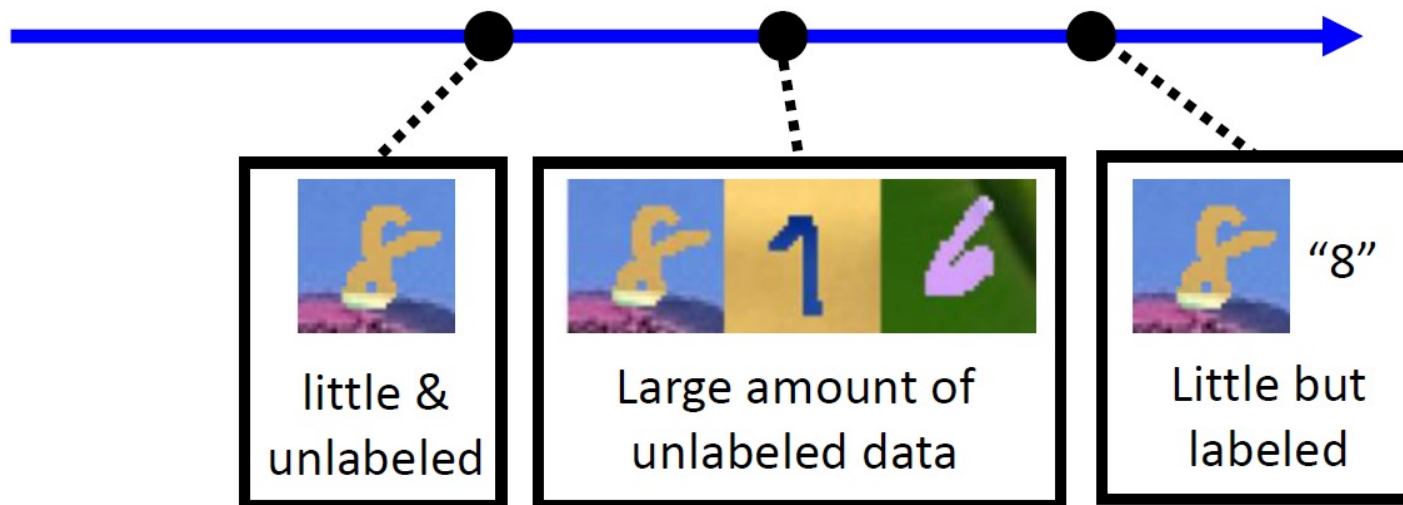
源域和目标域



- ④ 域偏差：训练数据和测试数据有不同的特征分布
- ④ 针对多中心数据的模型训练策略



目标域的知识

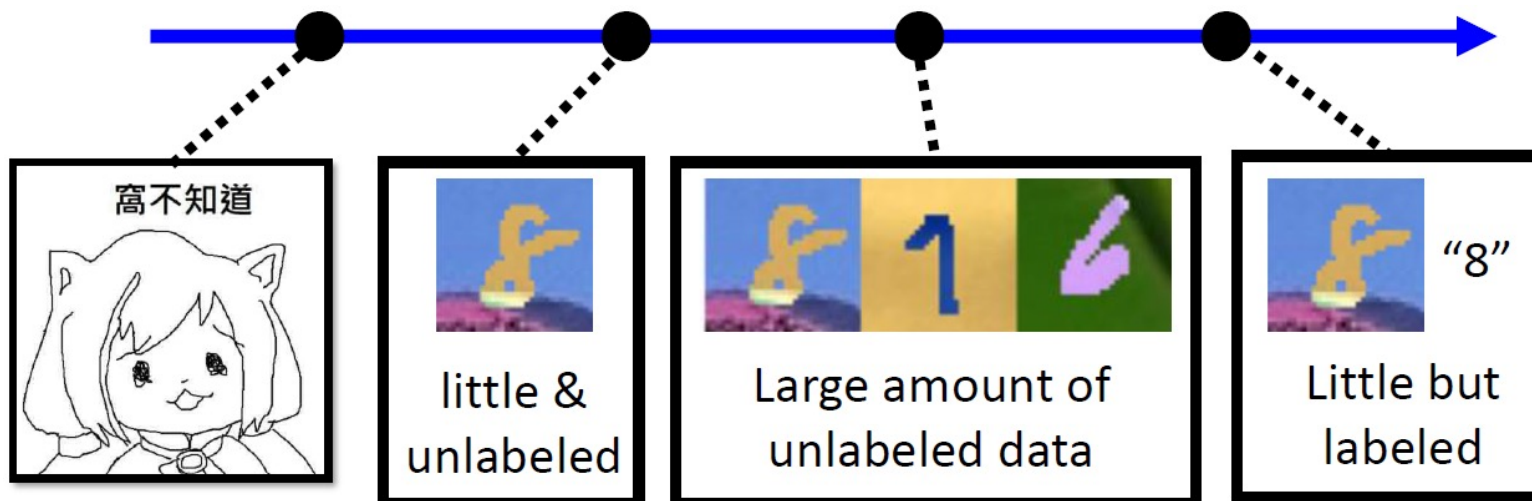


This picture is from:

https://speech.ee.ntu.edu.tw/~hylee/ml/ml2021-course-data/da_v6.pdf



目标域的知识



This picture is from:

https://speech.ee.ntu.edu.tw/~hylee/ml/ml2021-course-data/da_v6.pdf

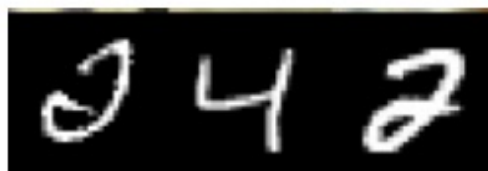


域适应 (Domain Adaptation)

Training
Data



Testing
Data



99.5%

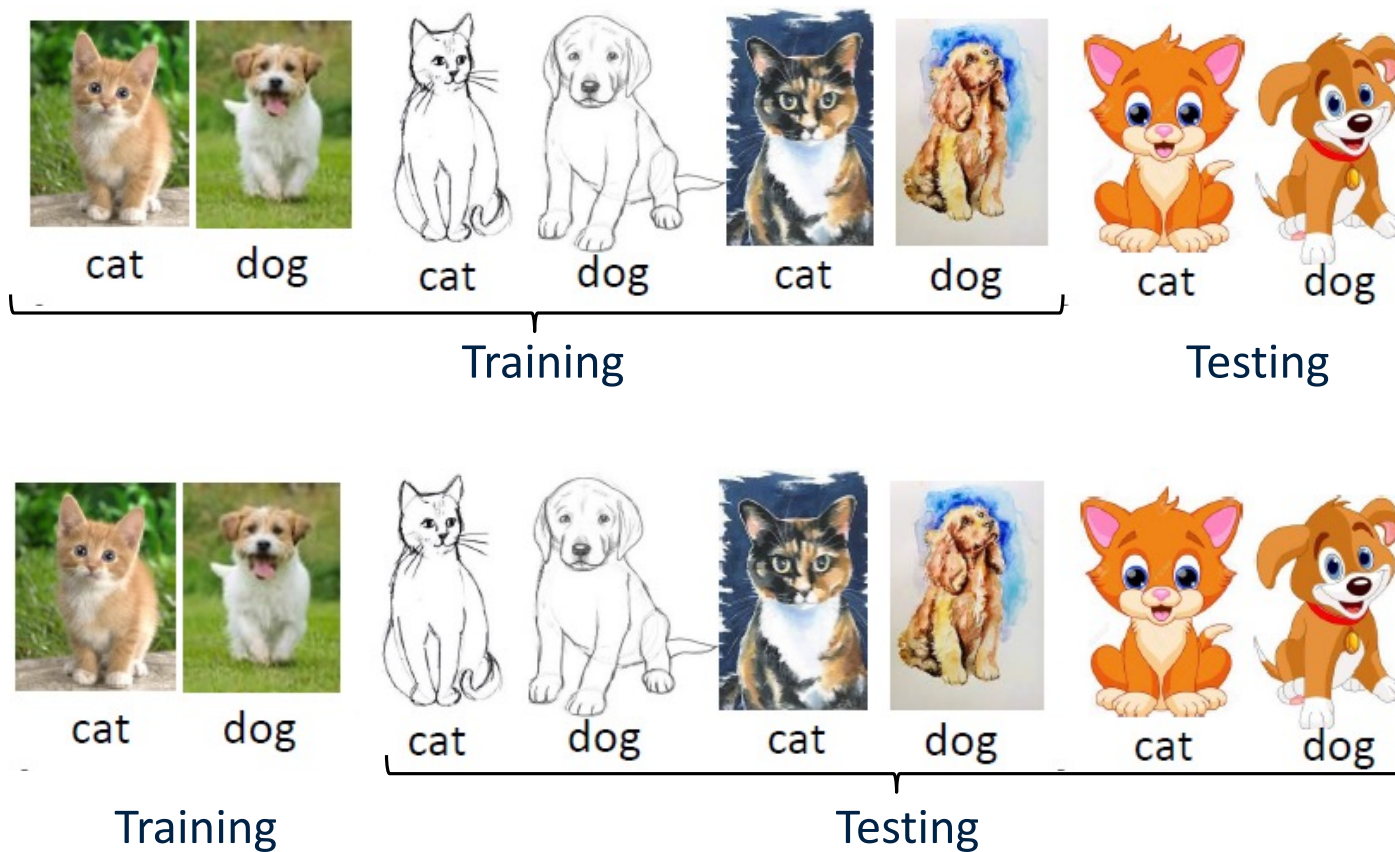


57.5%

The results are from: <http://proceedings.mlr.press/v37/ganin15.pdf>

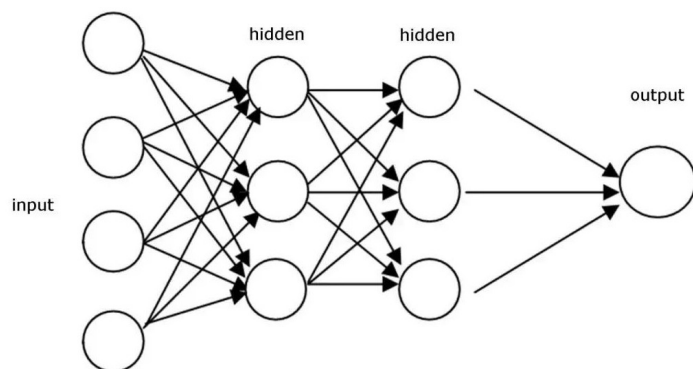


域泛化 (Domain Generalization)

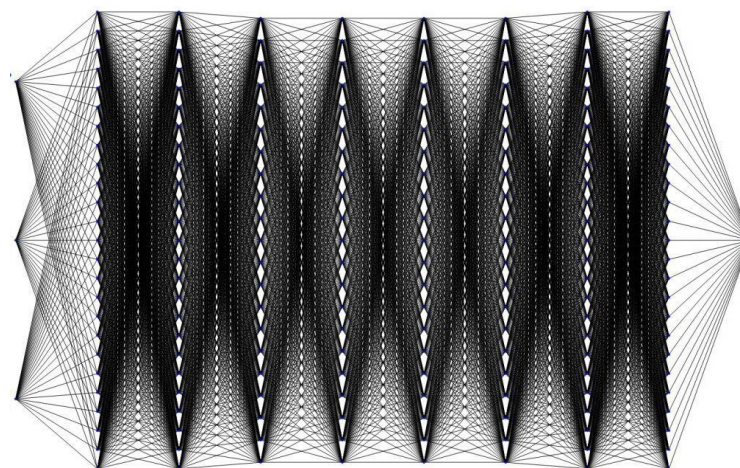




深度学习



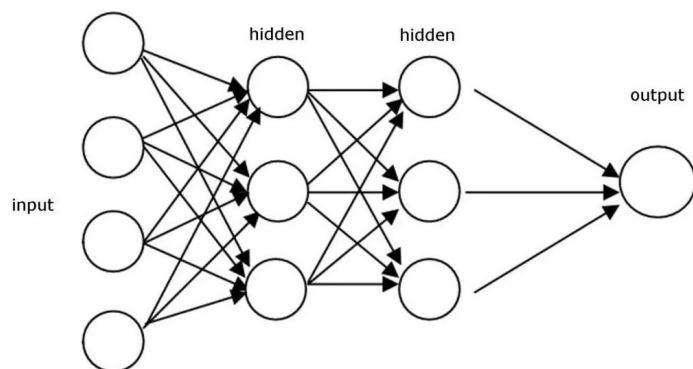
机器学习
网络比较浅，神经元比较少



深度学习
网络更深，神经元非常多

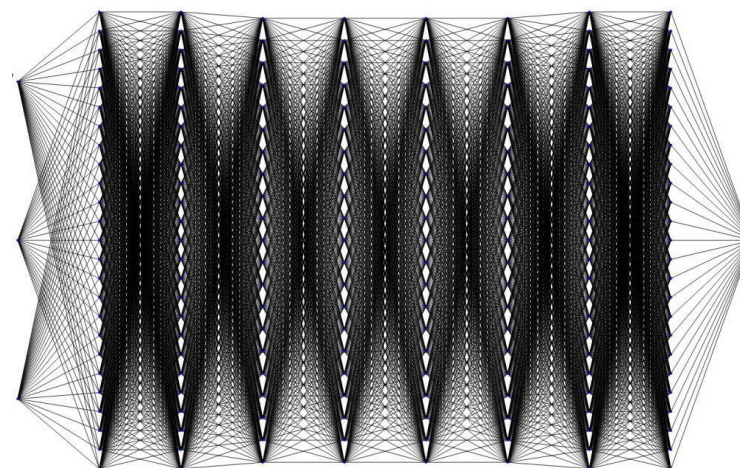


深度学习



机器学习

网络比较浅，神经元比较少



深度学习

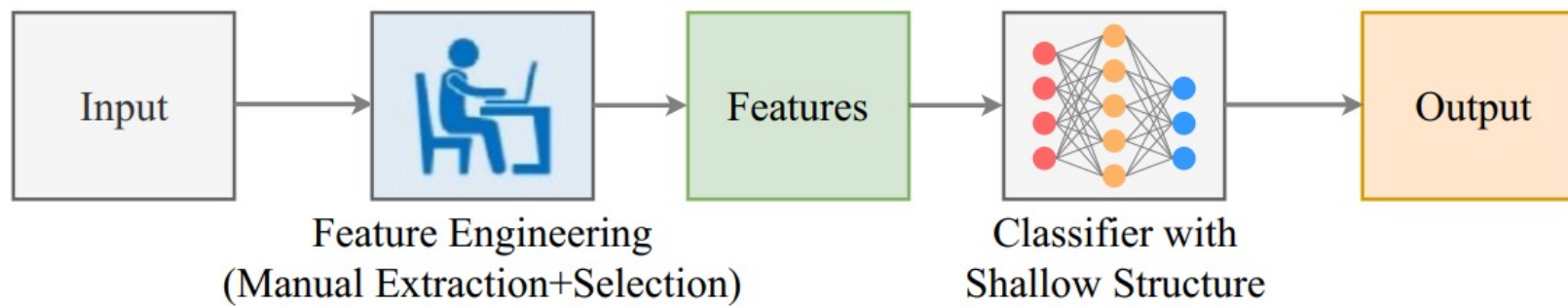
网络更深，神经元非常多

- 存储、GPU并行加速等硬件的发展
- 梯度消失问题的有效解决

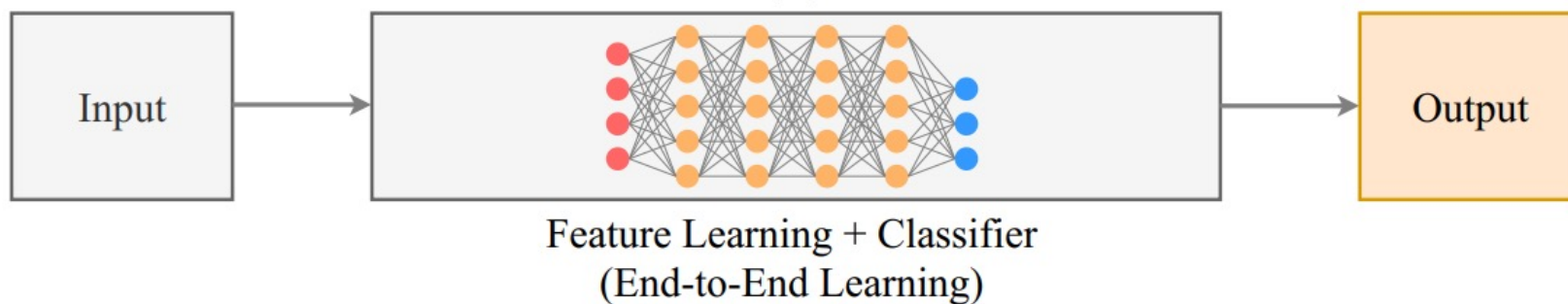


手工特征 vs. 深度特征

传统视觉任务框架：

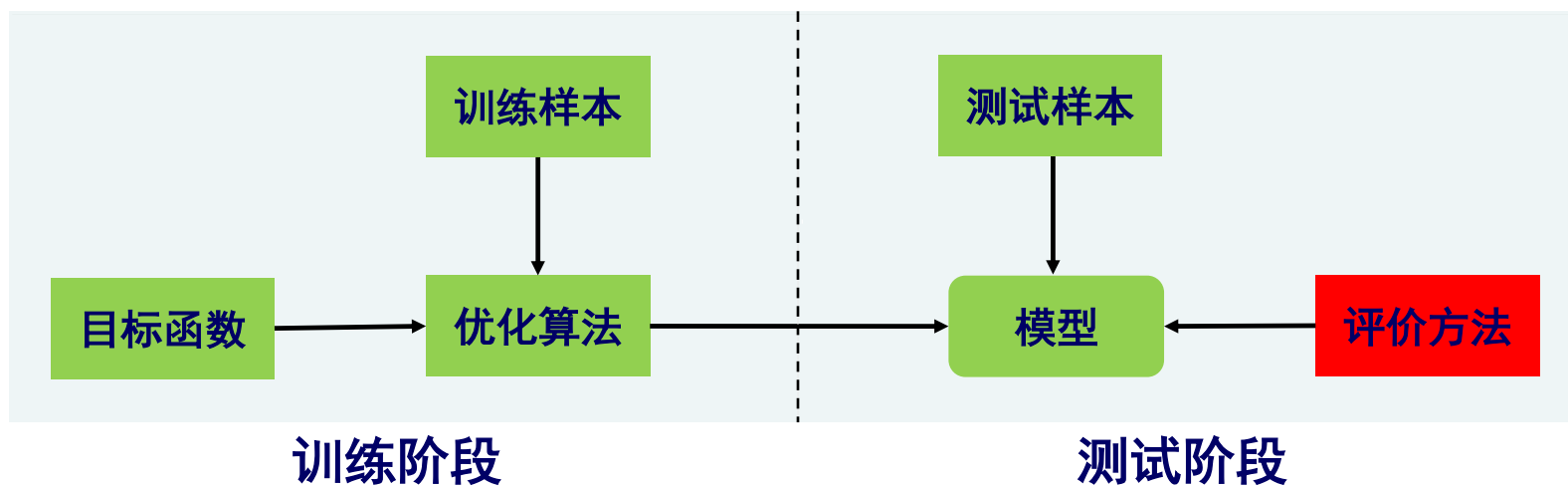


深度学习视觉任务框架：





如何构建模型





如何去衡量模型的性能？





模型泛化性能

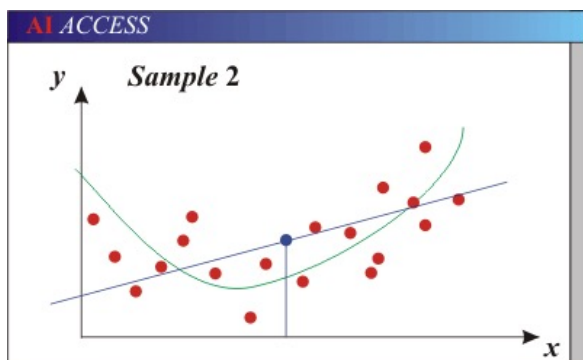
用训练集训练好的模型在未知数据上的预测好坏（性能表现高低），通常叫做泛化能力。

$$E(\text{泛化误差}) = \text{noise (噪声)} + \text{bias (偏差)} + \text{variance (方差)}$$

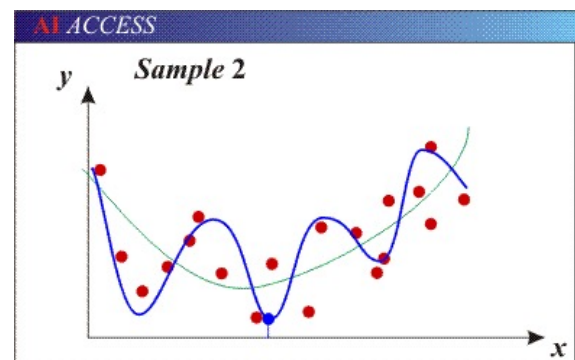
刻画了学习问题本身的难度

刻画预测和真实的偏离程度

刻画数据扰动对性能影响



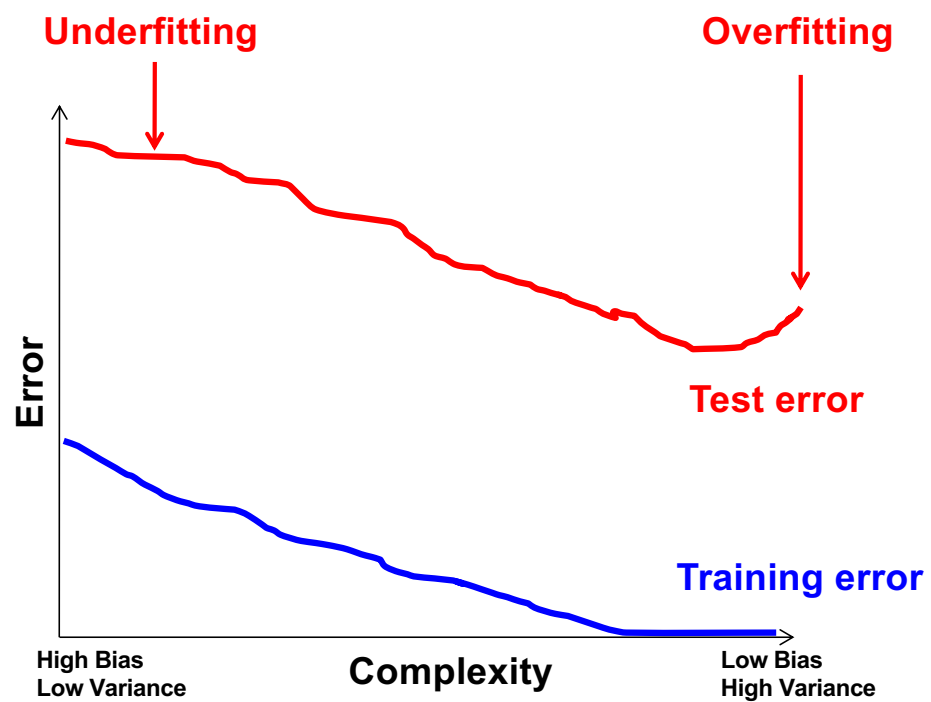
欠拟合：模型假设过于简单，参数量少，无法充分拟合真实期望模型
(高偏差：模型连训练数据都不能很好识别)



过拟合：模型假设过于复杂，参数量大，过度拟合数据
(高方差：模型只能识别训练数据，扰动影响很大)

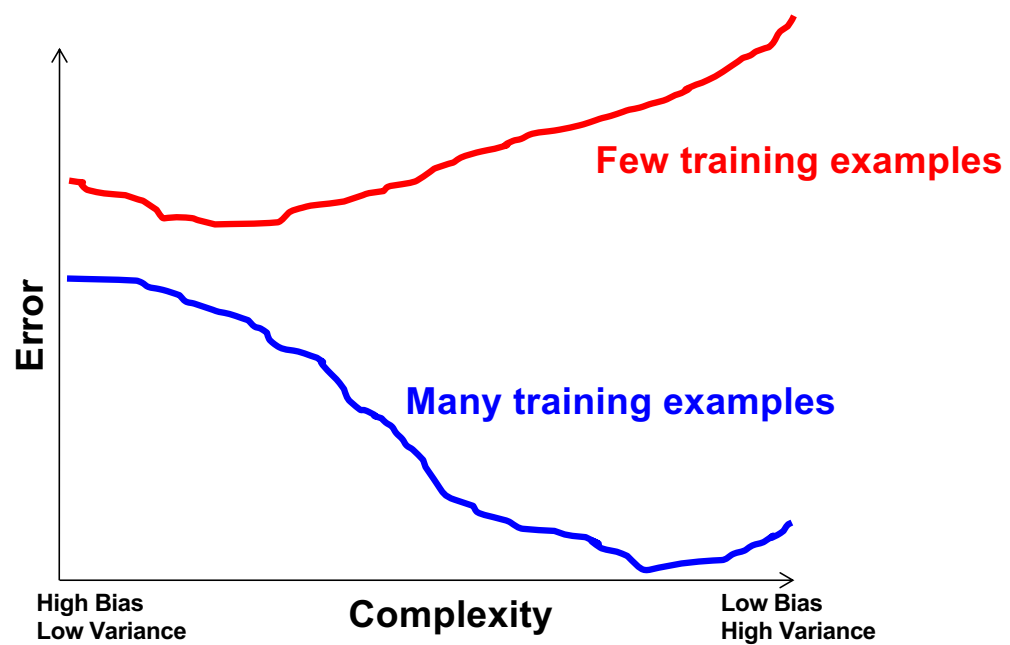


欠拟合和过拟合





数据量的影响



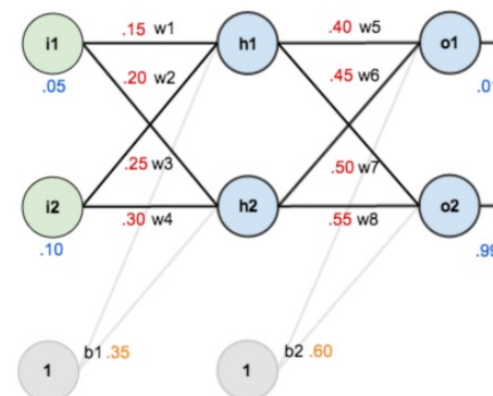


缓解过拟合常用方法

- 尽量获得更多的训练数据
- 数据扩增手段
- 简化模型结构和减少参数量
- Dropout: 随机丢弃部分参数（设为0）
- 正则化: Loss函数加入L1或者L2

L2正则化: loss函数后面加上权重参数的平方
效果: 权重参数值趋向于0, 也称为权重衰减。

$$L + \lambda \sum w^2$$





计算机视觉深度网络 (Deep Network in Computer Vision)

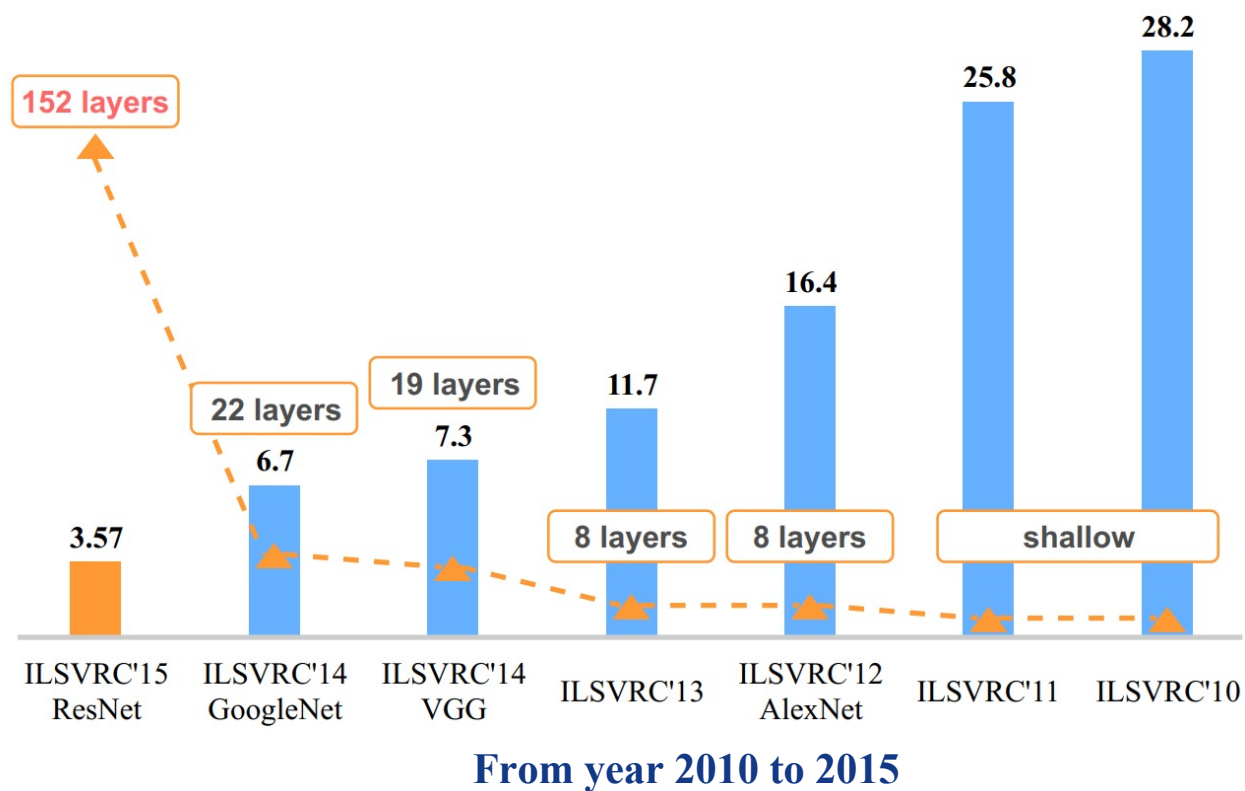


计算机视觉基础贡献奖 - CVPR 2019

[Deng, CVPR, 2009]

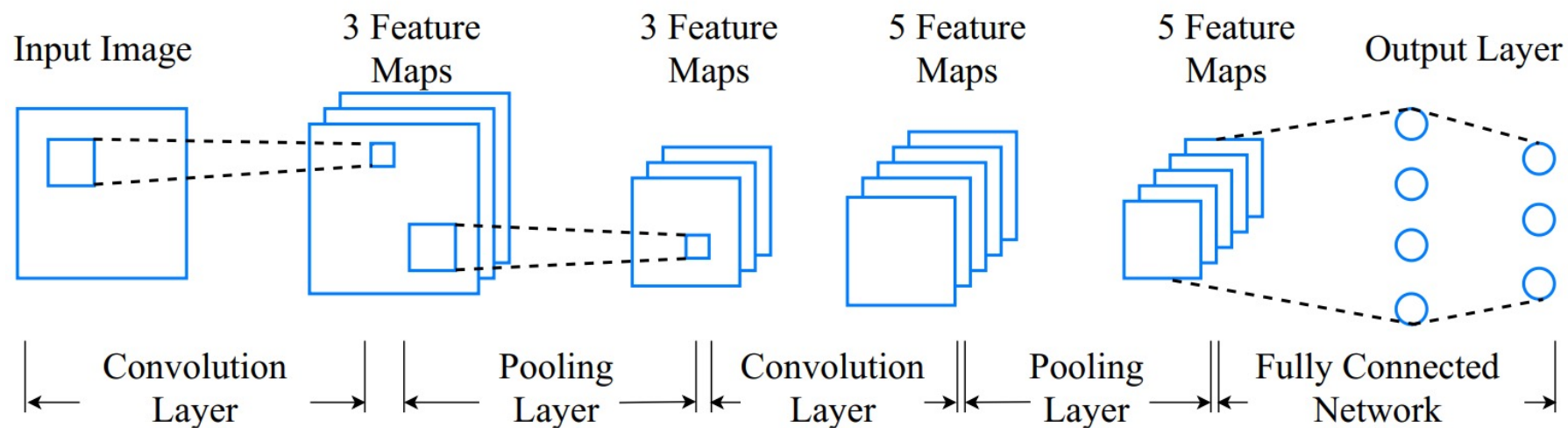


ImageNet分类Top-5错误率





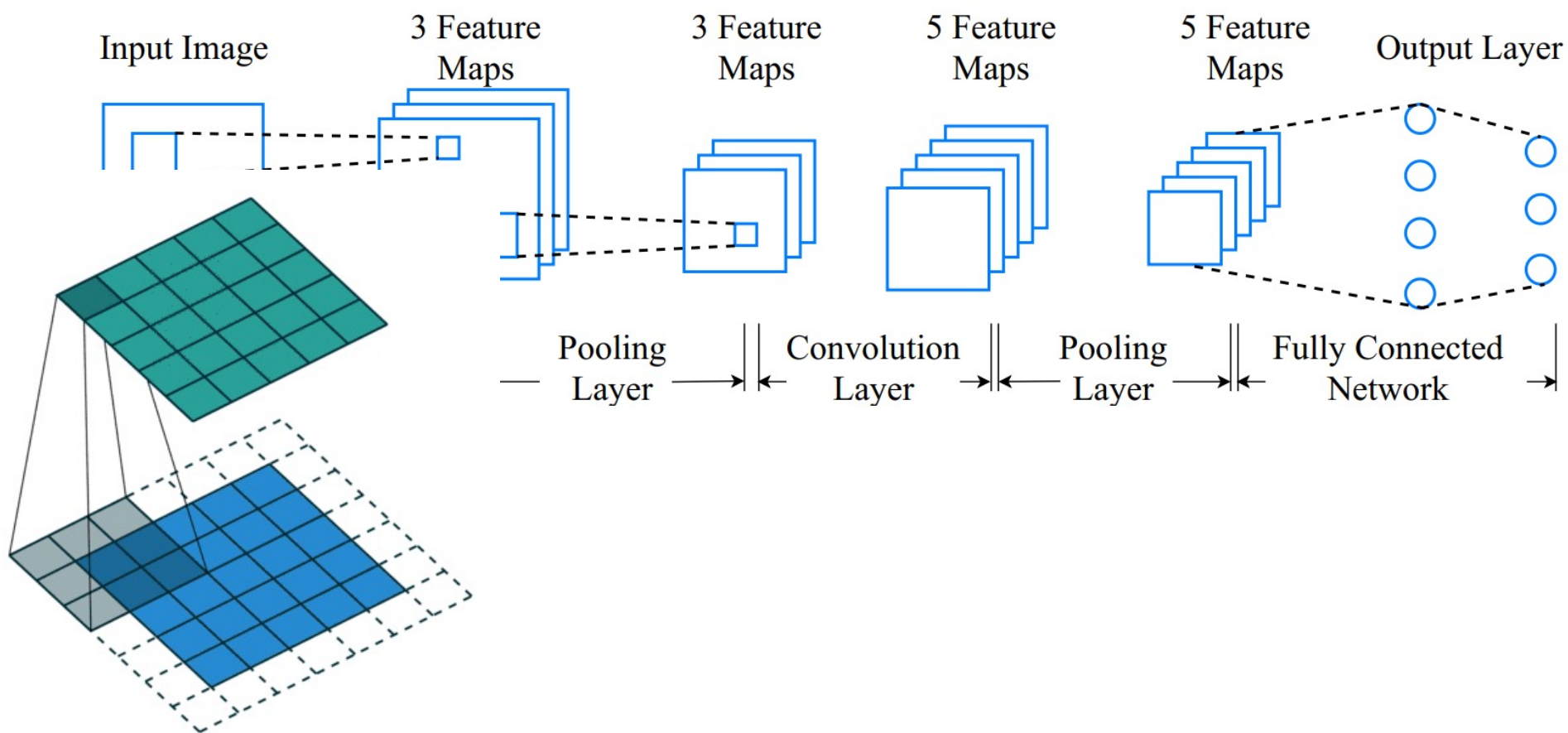
CNN网络的主要结构



CNN: 善于提取图像特征

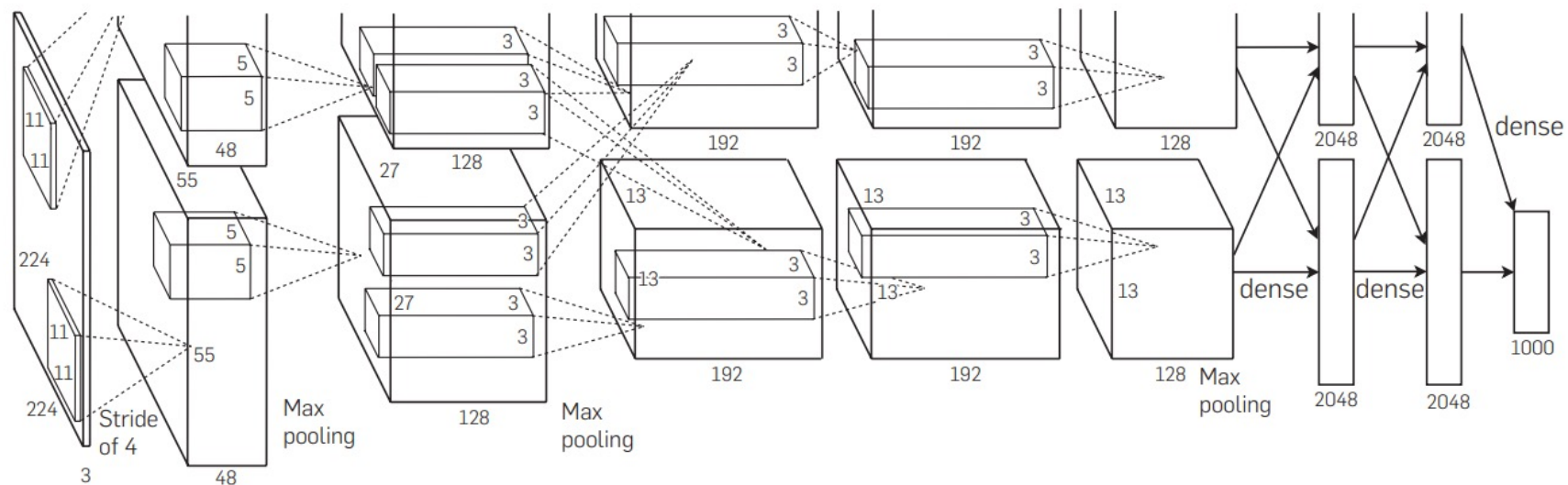


CNN网络的主要结构





经典卷积神经网络-AlexNet



[Krizhevsky, Communications of the ACM, 2017]

2012年ImageNet竞赛冠军

使用GPU进行运算加速



GeForce GTX 690 - year 2012



This picture is from:

<https://baijiahao.baidu.com/s?id=1757972284429992661&wfr=spider&for=pc>

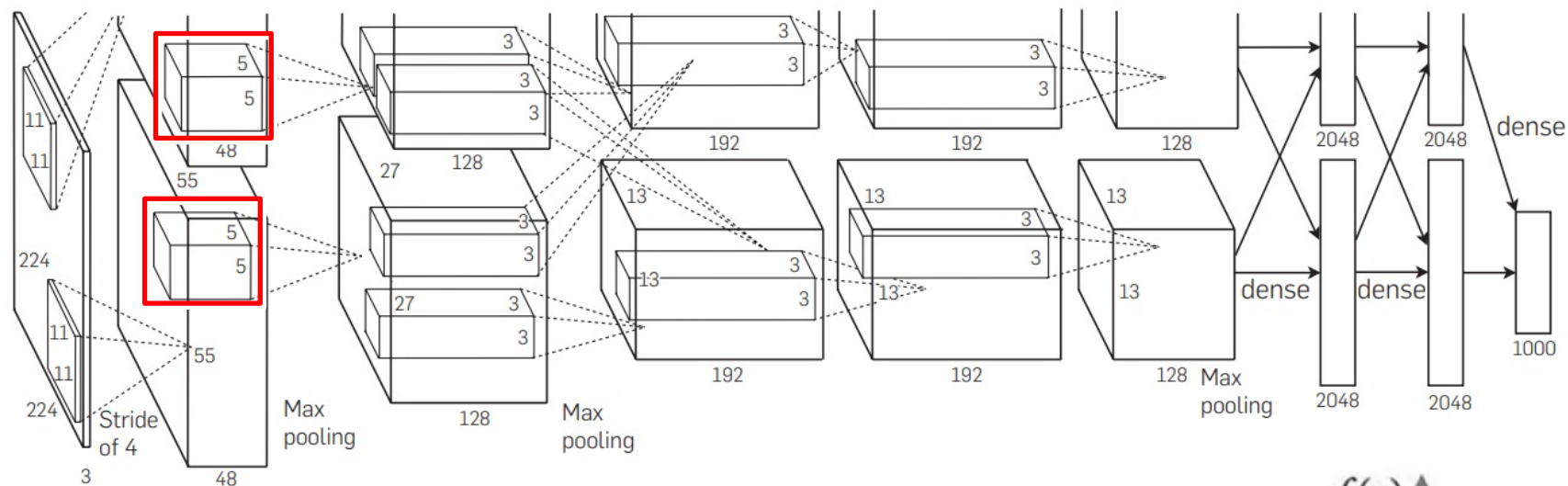
28nm制程工艺 & Kepler架构 (\$1,000)

显存容量: 4096MB

显存频率: 6000MHz



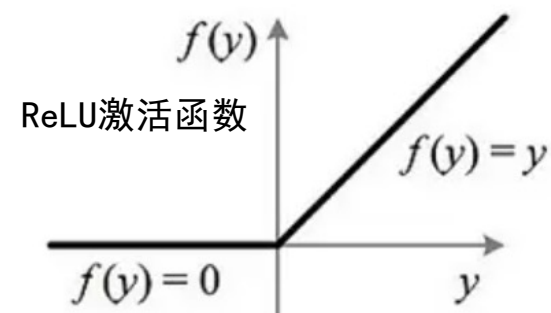
经典卷积神经网络-AlexNet



[Krizhevsky, Communications of the ACM, 2017]

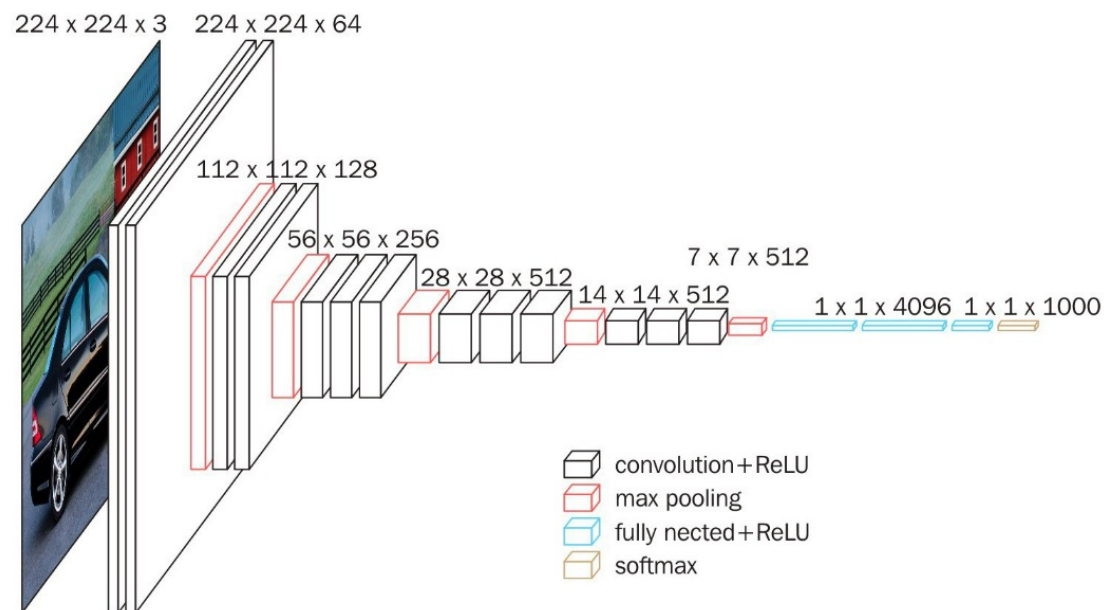
2012年ImageNet竞赛冠军

使用GPU进行运算加速





经典卷积神经网络-VGG



[Simonyan, ICLR, 2015]

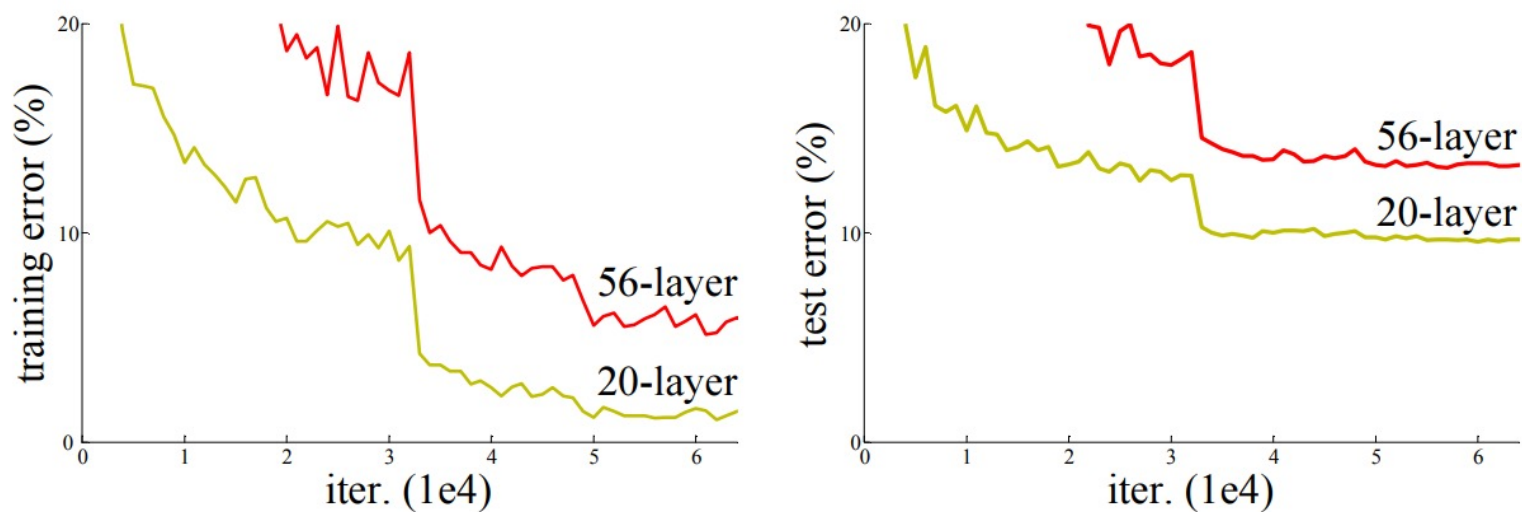
2014年ImageNet竞赛 (Localization Task第一名 & Classification Task 第二名)

小卷积核&多尺度 

加深网络深度 



经典CNN模型-ResNet



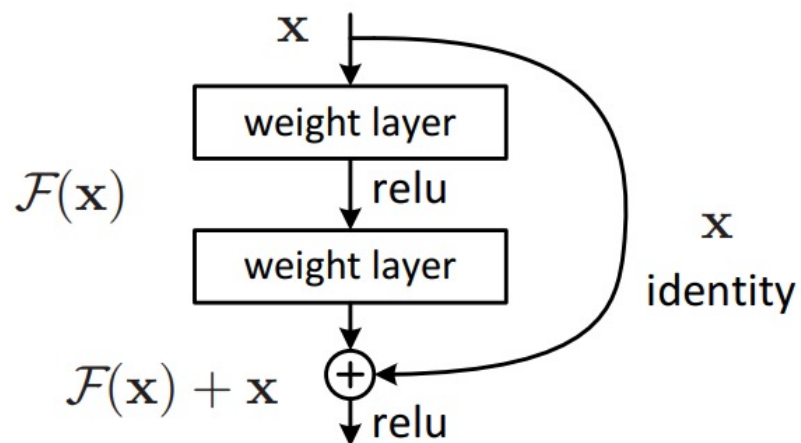
Training / Test error on CIFAR-10 with 20-layer and 56-layer “plain” networks

This picture is from:

https://openaccess.thecvf.com/content_cvpr_2016/papers/He_Deep_Residual_Learning_CVPR_2016_paper.pdf

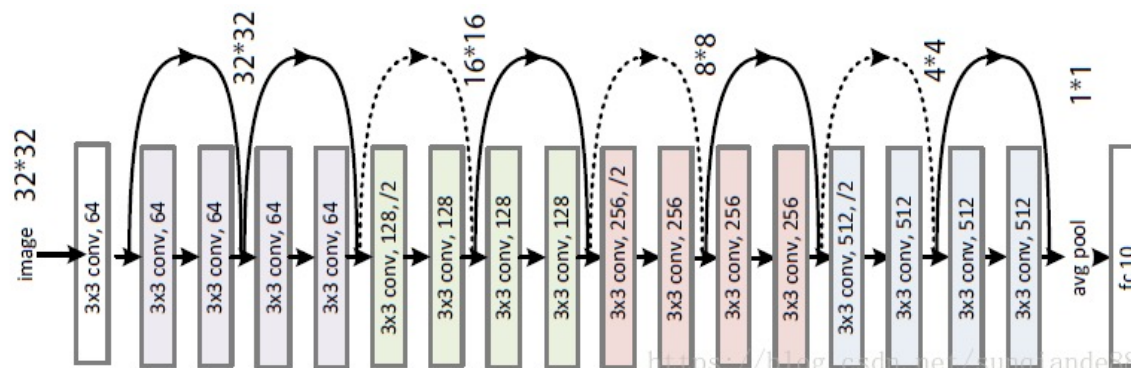


经典CNN模型-ResNet



残差学习单元

可以有效缓解
梯度消失问题

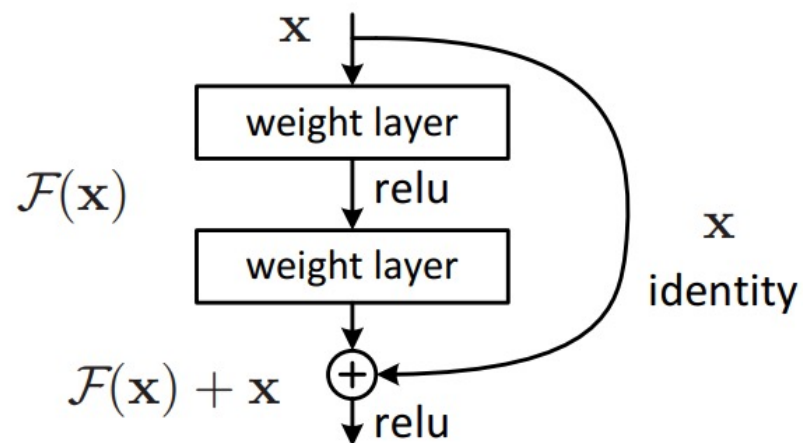


ResNet-18

[He, CVPR, 2016]

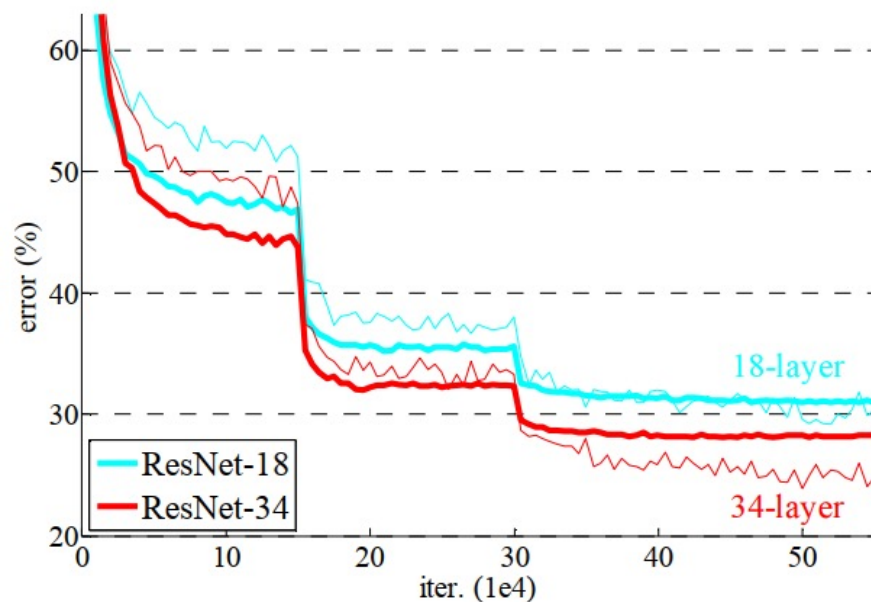


经典CNN模型-ResNet



残差学习单元

可以有效缓解
梯度消失问题



Validation error of ResNet-18 and ResNet-34



经典CNN模型-ResNet

layer name	output size	18-layer	34-layer	50-layer	101-layer	152-layer
conv1	112×112	7×7, 64, stride 2				
conv2_x	56×56	3×3 max pool, stride 2				
		$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 64 \\ 3 \times 3, 64 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 64 \\ 3 \times 3, 64 \\ 1 \times 1, 256 \end{bmatrix} \times 3$
conv3_x	28×28	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 128 \\ 3 \times 3, 128 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 4$	$\begin{bmatrix} 1 \times 1, 128 \\ 3 \times 3, 128 \\ 1 \times 1, 512 \end{bmatrix} \times 8$
conv4_x	14×14	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 256 \\ 3 \times 3, 256 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 6$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 23$	$\begin{bmatrix} 1 \times 1, 256 \\ 3 \times 3, 256 \\ 1 \times 1, 1024 \end{bmatrix} \times 36$
conv5_x	7×7	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 2$	$\begin{bmatrix} 3 \times 3, 512 \\ 3 \times 3, 512 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$	$\begin{bmatrix} 1 \times 1, 512 \\ 3 \times 3, 512 \\ 1 \times 1, 2048 \end{bmatrix} \times 3$
	1×1	average pool, 1000-d fc, softmax				
FLOPs		1.8×10^9	3.6×10^9	3.8×10^9	7.6×10^9	11.3×10^9

ResNet 系列



经典CNN模型-ResNet

CheXNet: Radiologist-Level Pneumonia Detection on Chest X-Rays with Deep Learning

Pranav Rajpurkar^{*1} Jeremy Irvin^{*1} Kaylie Zhu¹ Brandon Yang¹ Hershel Mehta¹
Tony Duan¹ Daisy Ding¹ Aarti Bagul¹ Robyn L. Ball² Curtis Langlotz³ Katie Shpanskaya³
Matthew P. Lungren³ Andrew Y. Ng¹



Input
Chest X-Ray Image

CheXNet



Output

Pneumonia Positive (85%)



肺炎诊断应用

[Rajpurkar, arXiv, 2017]



经典CNN模型-DenseNet

- 特征重用（性能好）
- 参数少
- 计算量小
- 内存消耗大上采样

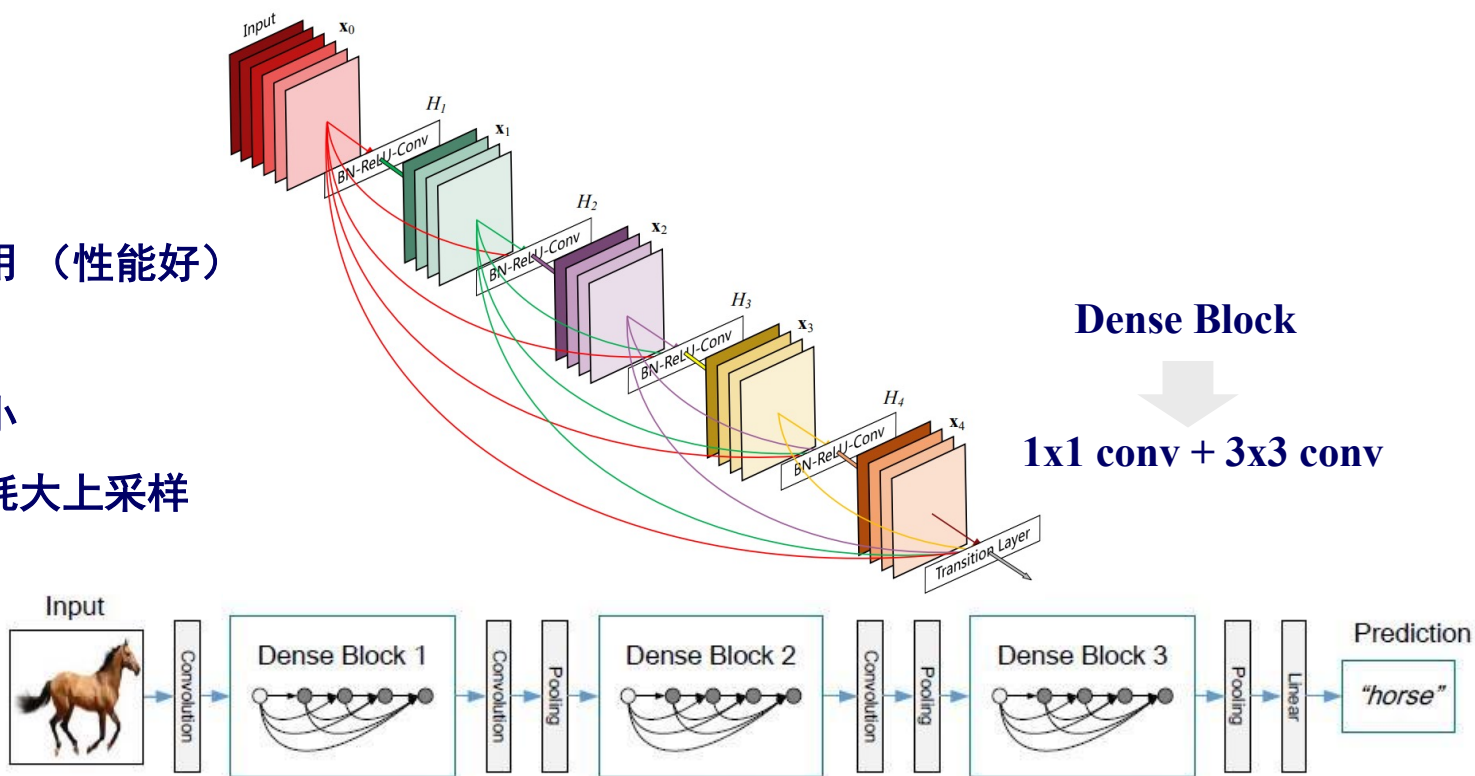


Figure 2. A deep DenseNet with three dense blocks. The layers between two adjacent blocks are referred to as transition layers and change feature map sizes via convolution and pooling.

DenseNet

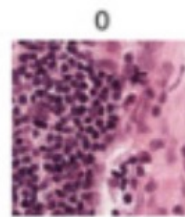
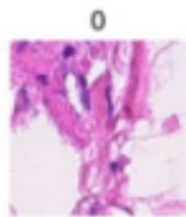
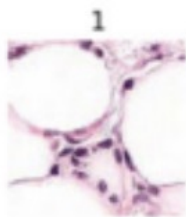
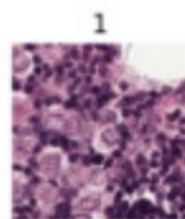
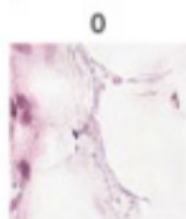
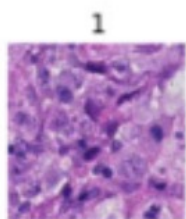
[Huang, CVPR, 2017]



经典CNN模型-DenseNet

Cancer image classification based on DenseNet model

Ziliang Zhong¹, Muhang Zheng¹, Huafeng Mai², Jianan Zhao³ and Xinyi Liu⁴

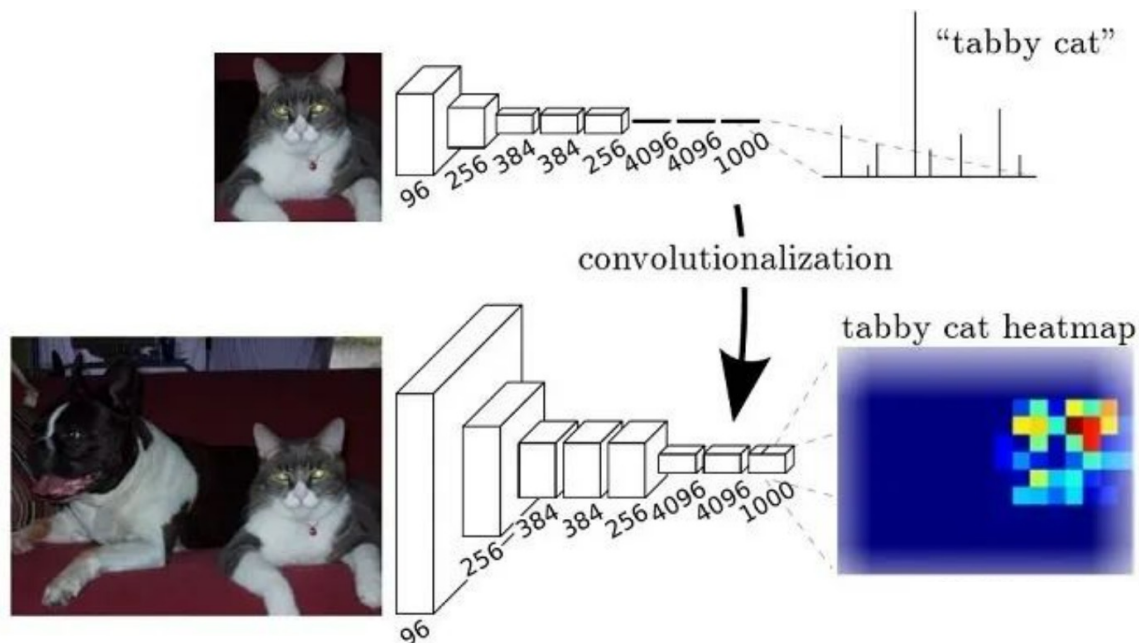


[Zhong, ICAITA, 2020]

癌症分类应用



经典CNN模型 -- Fully Convolutional Networks

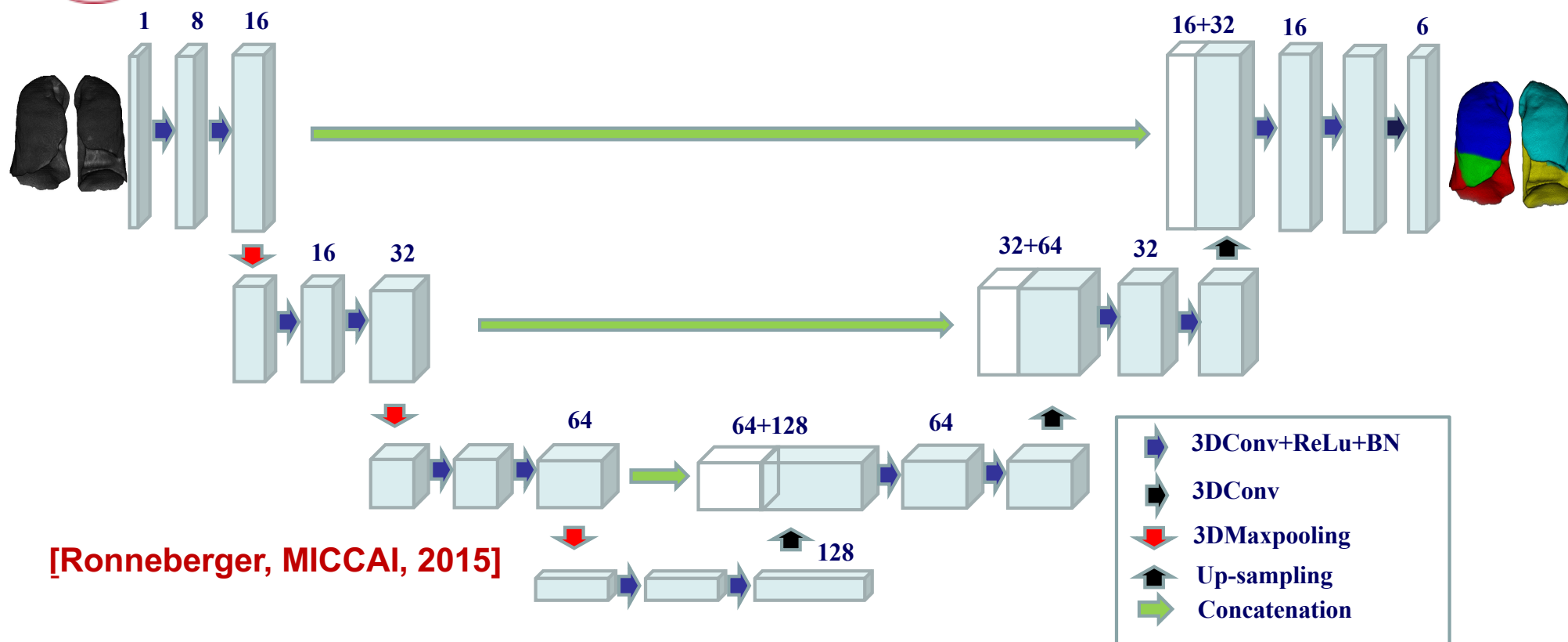


- 全卷积网络 -- 可适应任意尺寸输入
- 上采样 -- 增大图像尺寸，输出精细结果
- 忽略考虑像素之间的关系

[Long, CVPR, 2015]



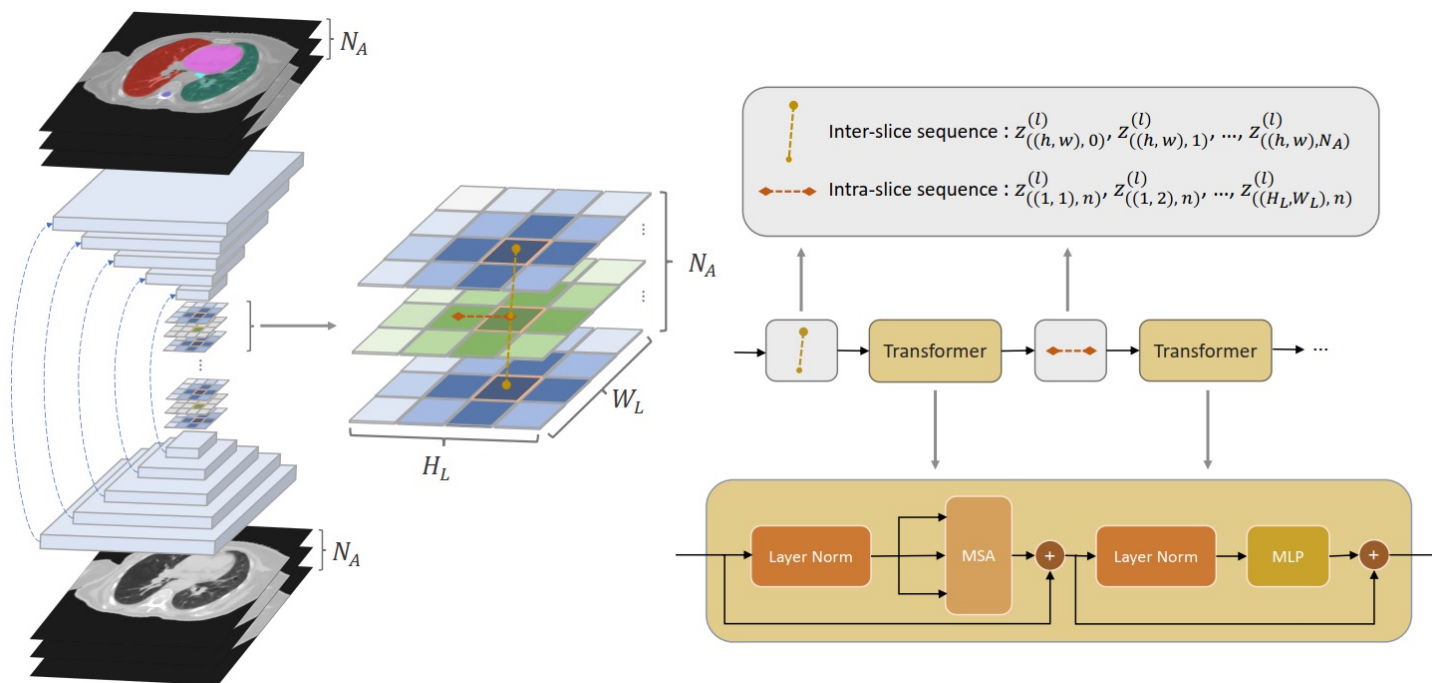
经典CNN模型-UNet分割网络



UNet: 下采样和上采样过程组成, 相当于对每个像素进行分类



经典CNN模型-UNet分割网络

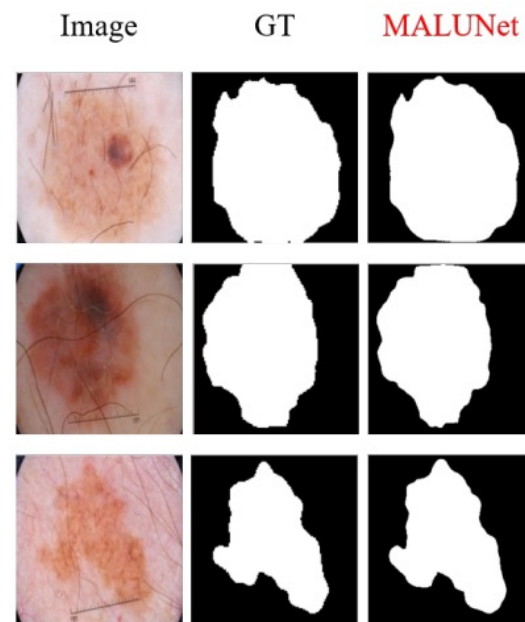
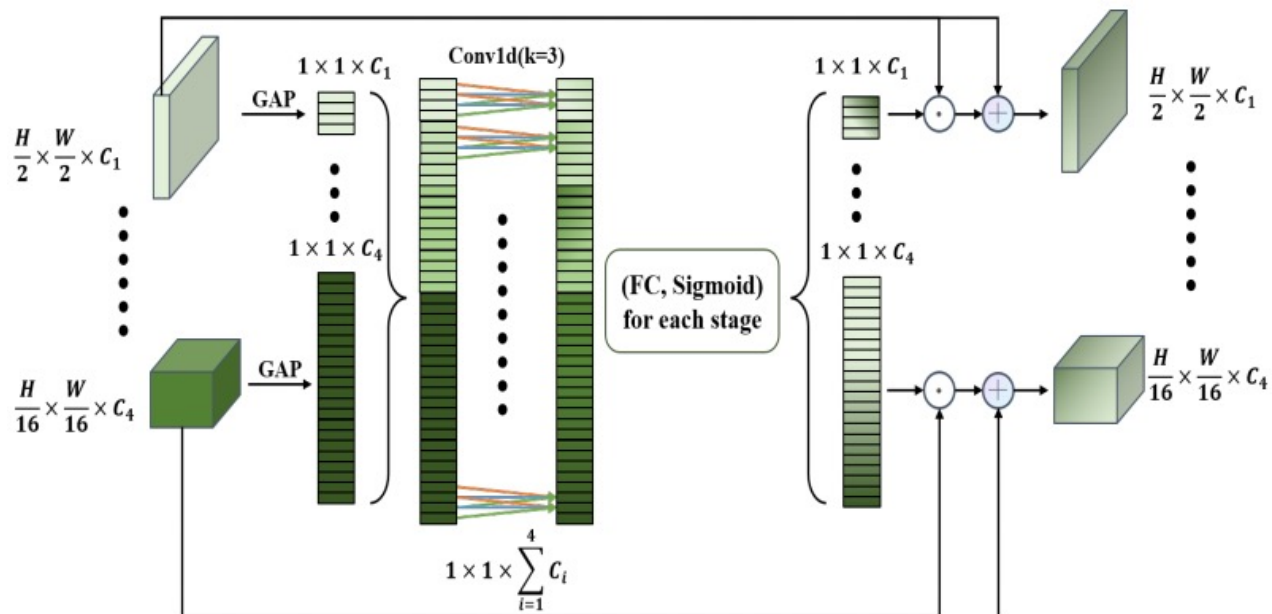


医学图像 – 多器官分割

[Yan, WACV, 2022]



轻量级CNN网络模型-MALUNet



基于深度可分离卷积捕捉全局和局部信息

[Ruan, BIBM, 2022]



目录

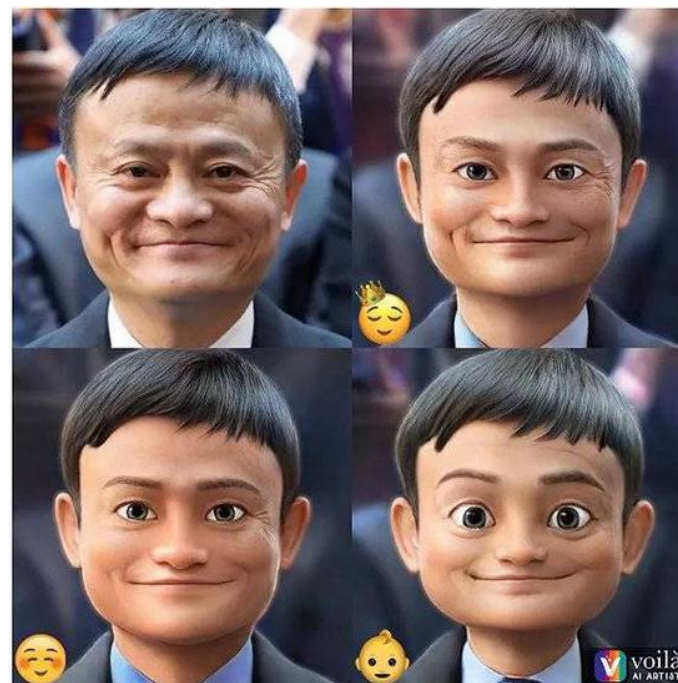
- 对抗生成网络
- Transformer模型
- 可解释性/热力图



对抗生成网络的应用



现实风格→油画风格



现实风格→3D卡通风格



对抗生成网络的应用

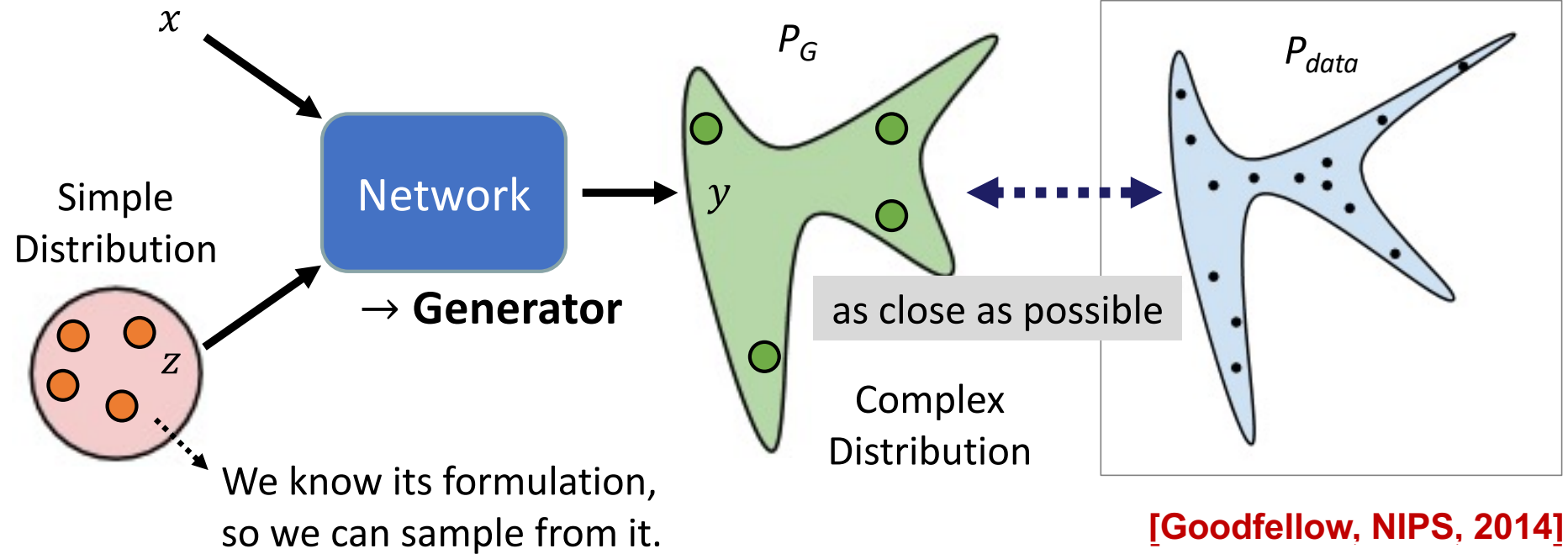


海滩风景生成

<https://nvlabs.github.io/stylegan3/>



对抗生成网络模型-Generative Adversarial Network



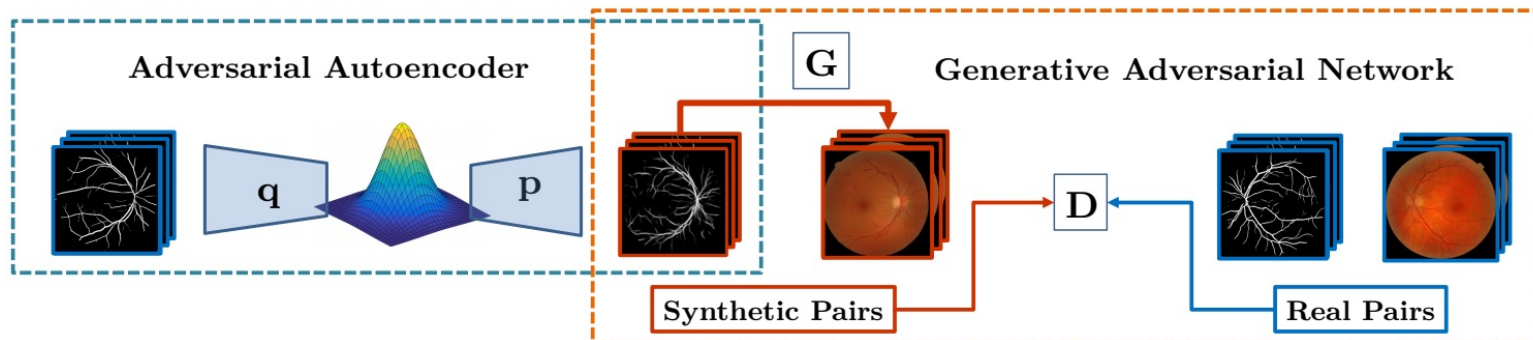
[Goodfellow, NIPS, 2014]

$$G^* = \arg \min_G \underline{Div}(P_G, P_{data})$$



对抗生成网络

TRAINING STAGE



TRAINED SYSTEM



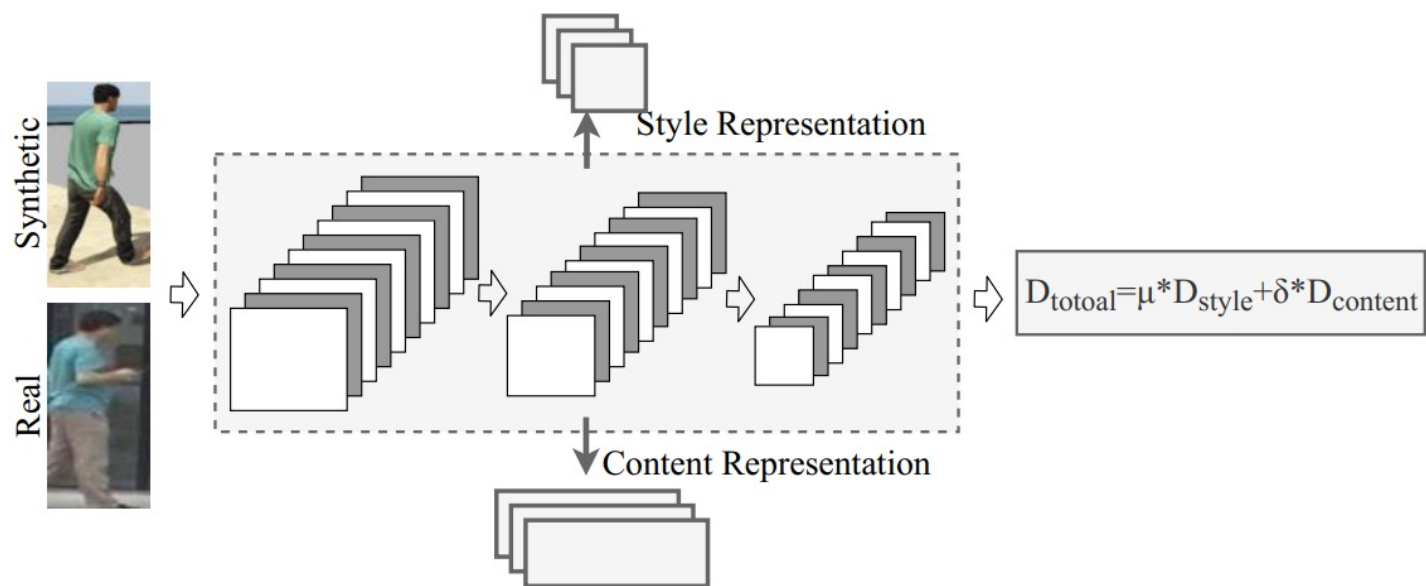
基于GAN进行视网膜眼底图像合成

视网膜血管网络 / 视网膜眼底图像 == Paired images

[Costa, TMI, 2017]



对抗生成网络

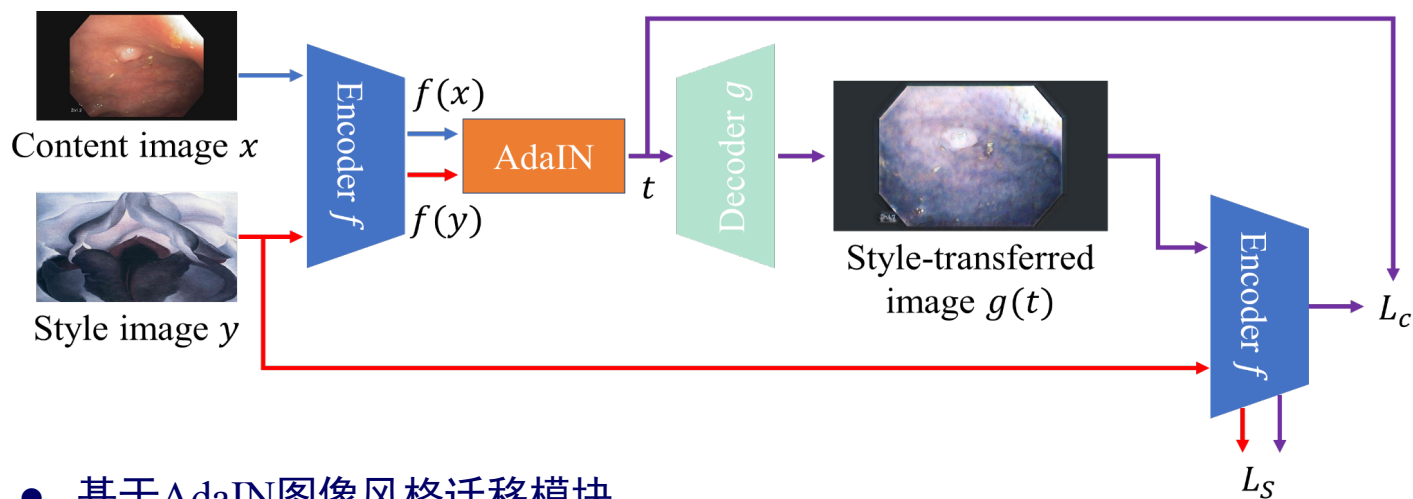


基于风格和内容表征的图像距离度量

[Xiang, ACM TOMM, 2023]



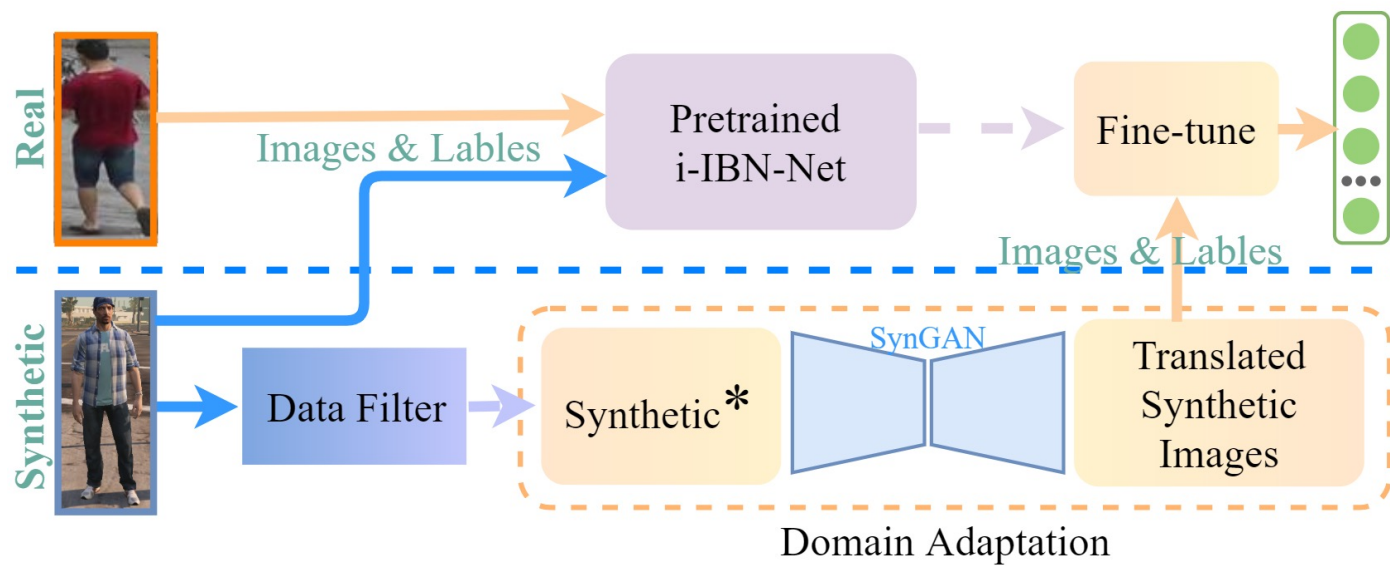
对抗生成网络



[Xu, Biomedical Signal Processing and Control, 2021]



对抗生成网络



基于GAN实现图像领域的域适应任务

[Xiang, ICME, 2020]



注意力机制-Transformer



This picture is from:

<https://www.xiaohongshu.com/explore/62dfc3ae000000001b022ab4>

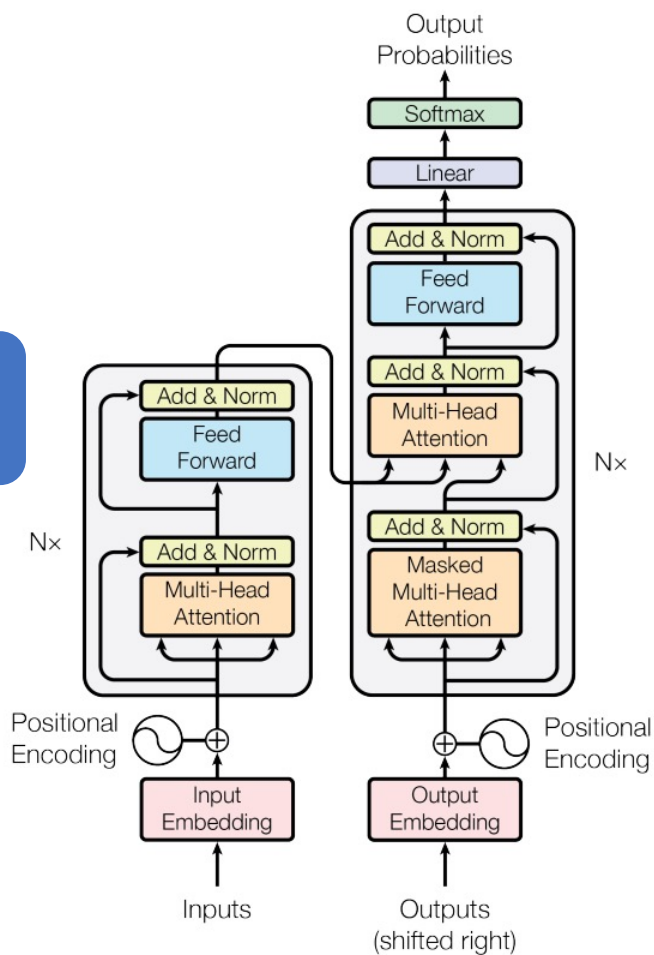


注意力机制-Transformer



Attention is all you need!

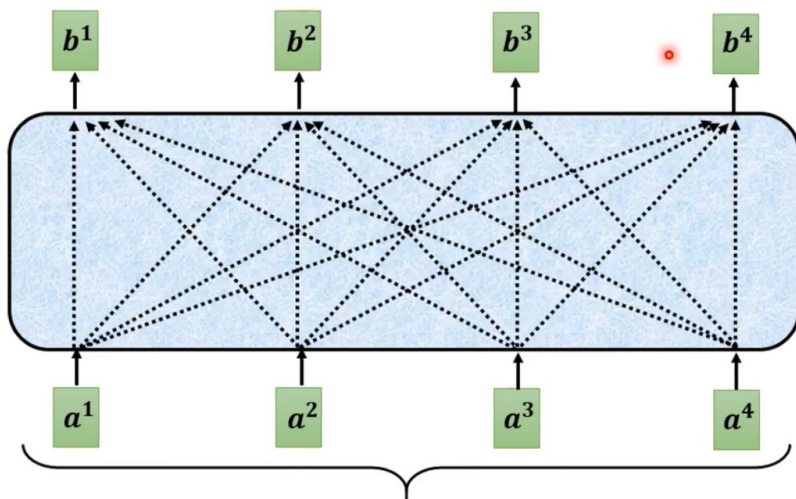
编码器-解码器的结构



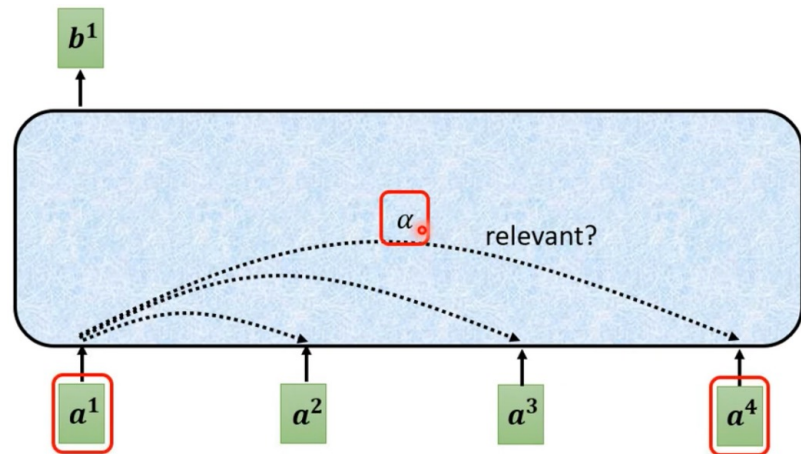
[Vaswani, NIPS, 2017]



注意力机制-Transformer



Self attention: 专注于整个序列信息



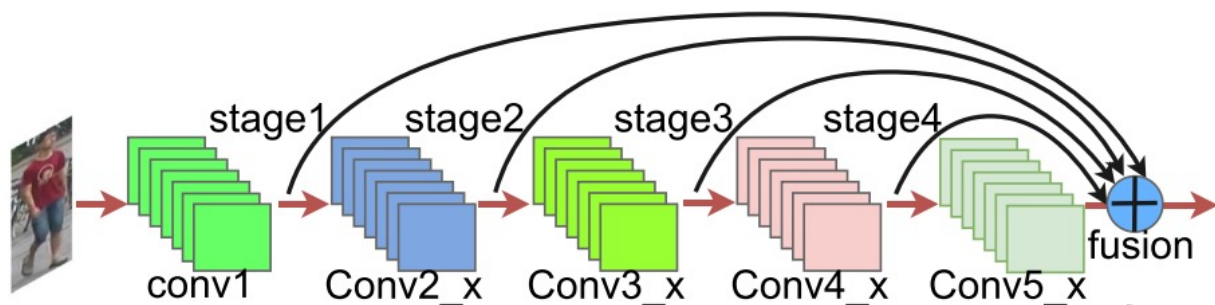
各个元素之间的相关性计算

This picture is from:

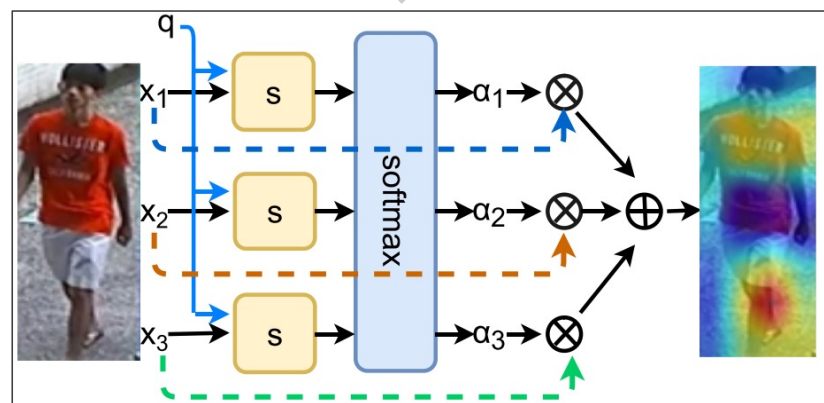
<https://blog.csdn.net/jy0608/article/details/124216732>



图像检索中的注意力机制



基于注意力机制的多尺度特征融合

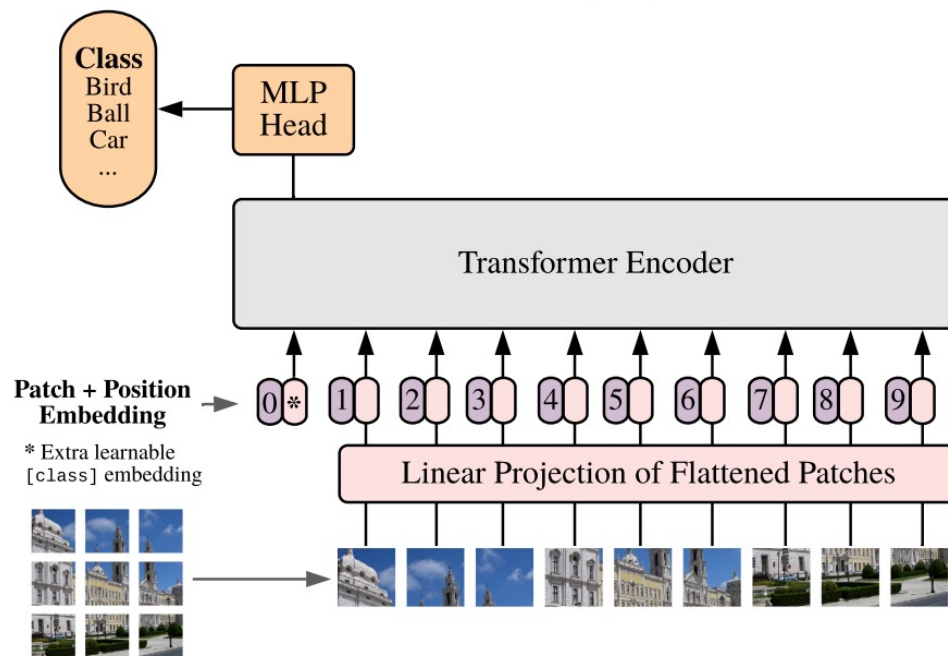


[Xiang, MTA, 2020]

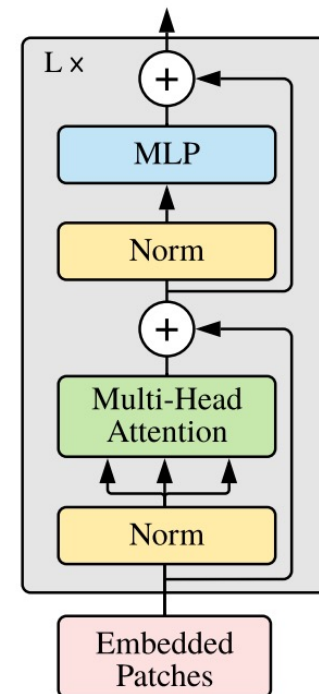


Vision Transformer

Vision Transformer (ViT)



Transformer Encoder

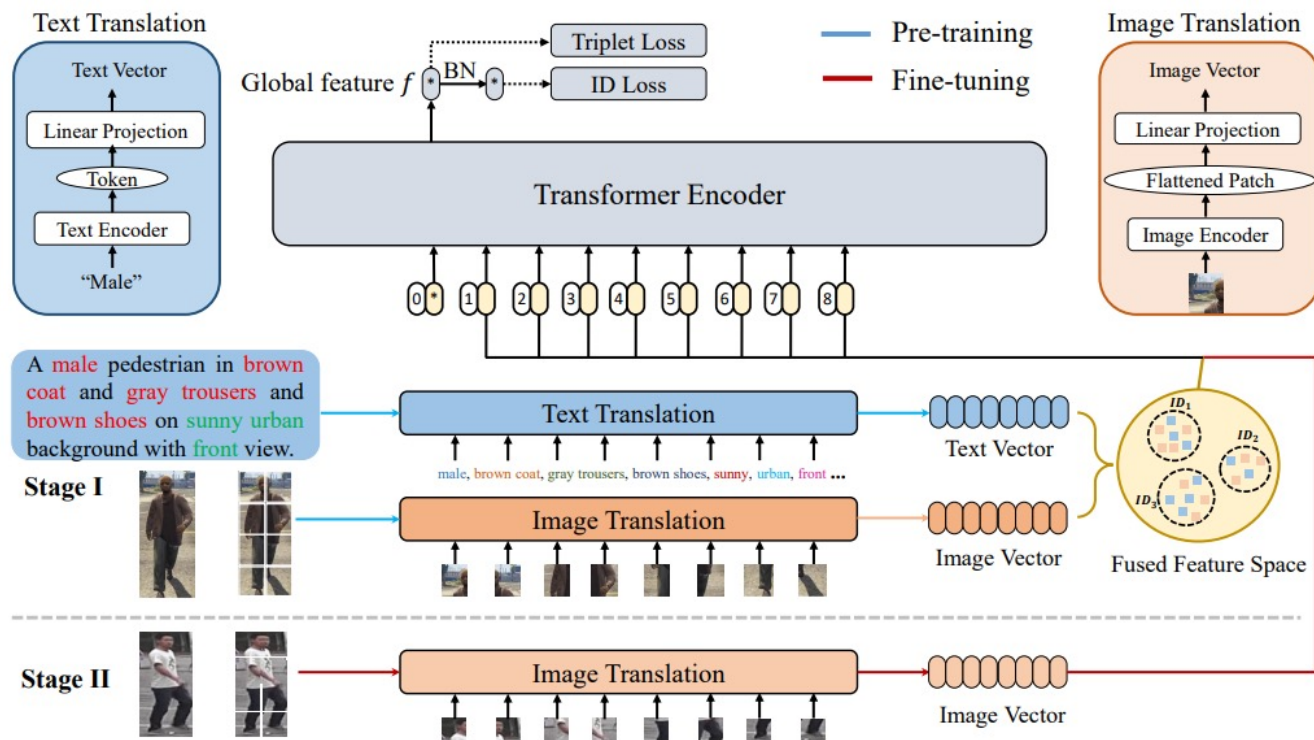


致力于打造视觉-文本的大一统模型

[Dosovitskiy, ICLR, 2021]



Vision & Text with Transformer

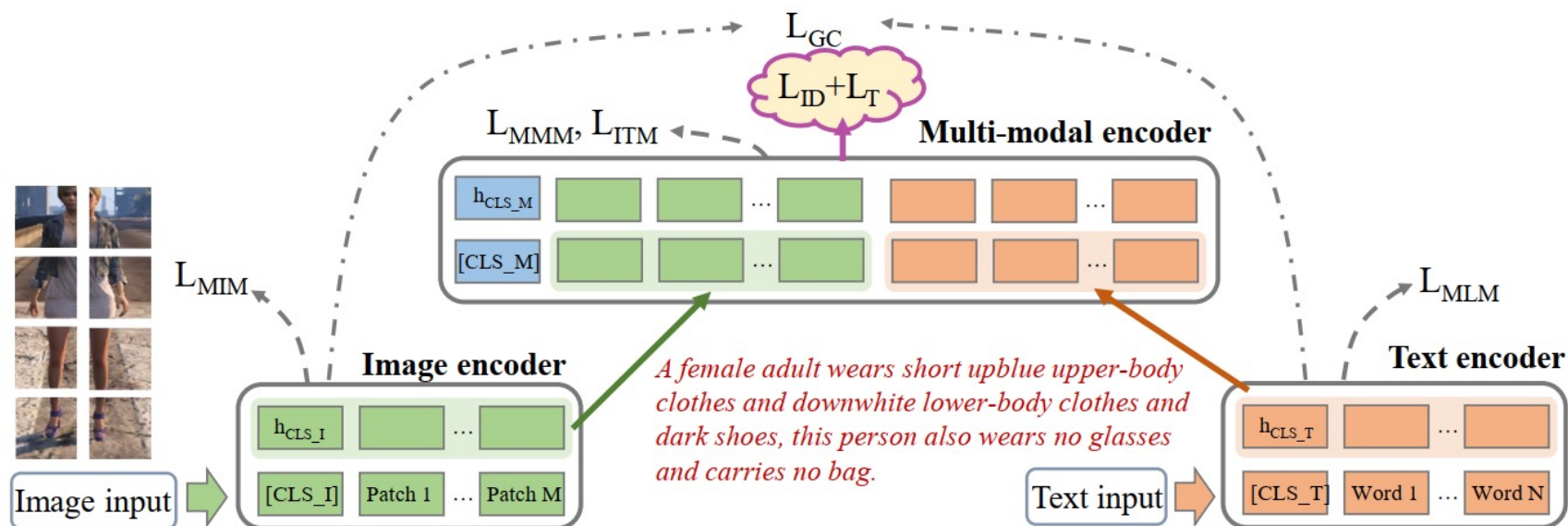


[Xiang, arXiv, 2022]

连接图像与文本的多模态融合模型



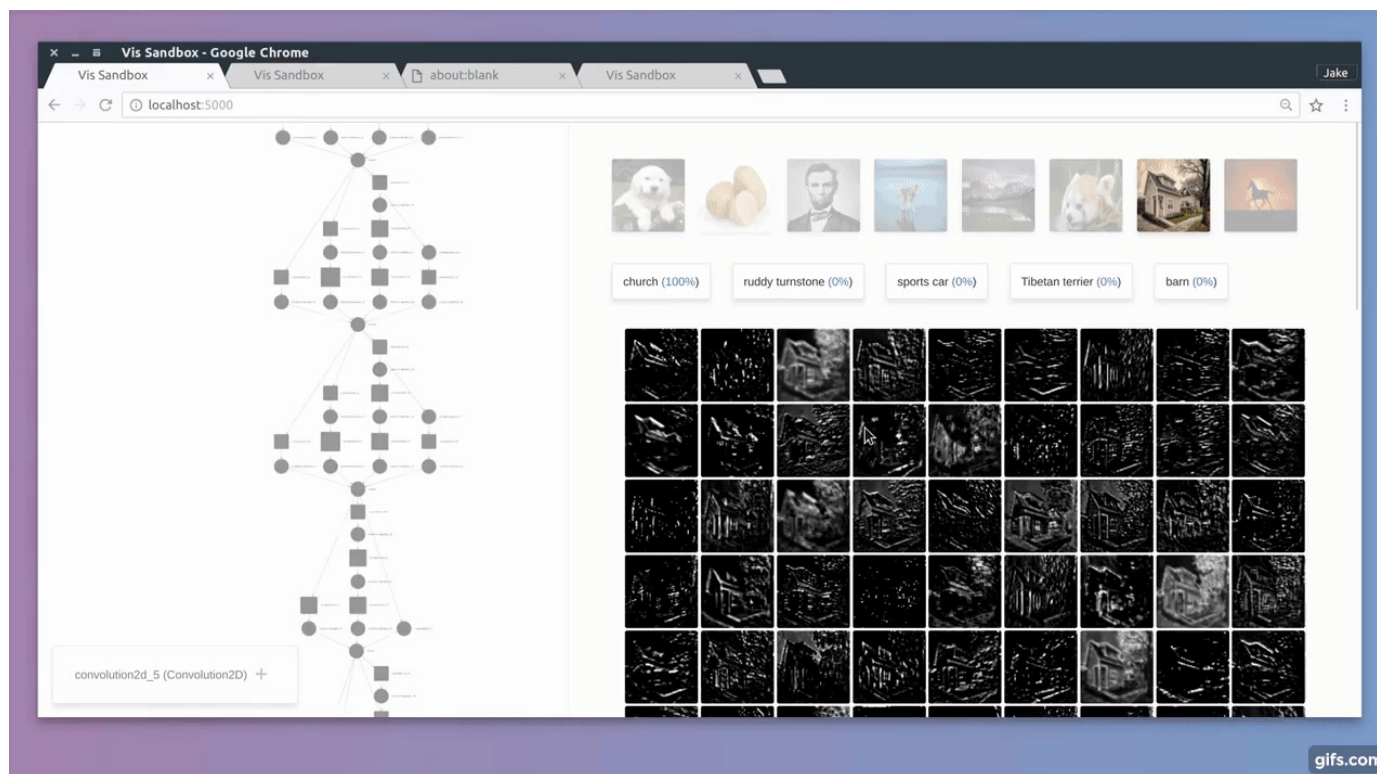
Vision & Text with Transformer



学习通用视觉表征的多模态融合模型



模型可视化

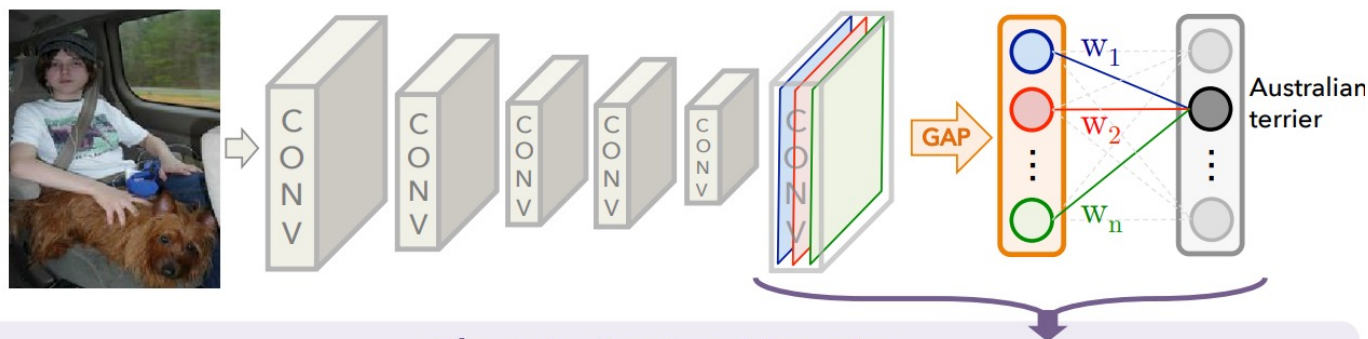


This project is from:
<https://github.com/keplr-io/quiver>

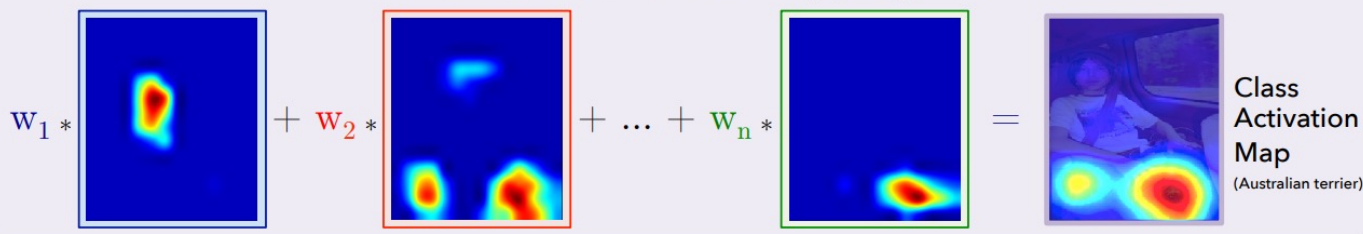


CNN可解释性—热力图

Gradient-weighted Class Activation Mapping



Class Activation Mapping



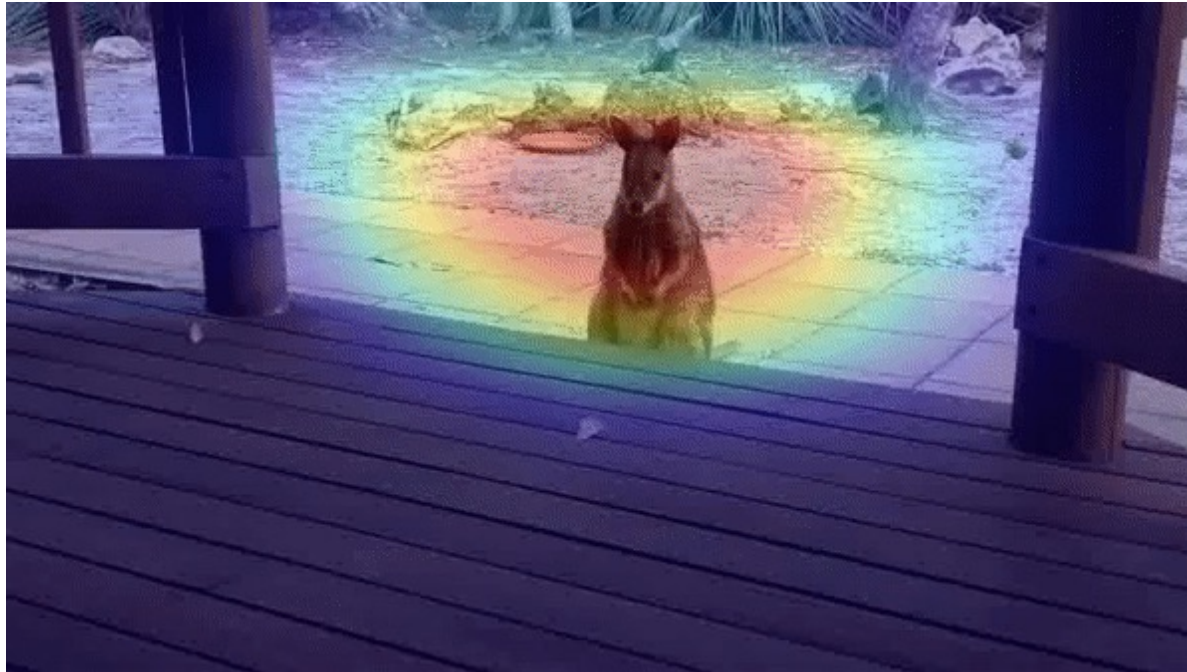
□ 优点：一定程度上对模型内部工作机制进行了可视化

[Zhou, CVPR, 2016]

□ 缺点：在医学领域作用有限，只能做主观解释



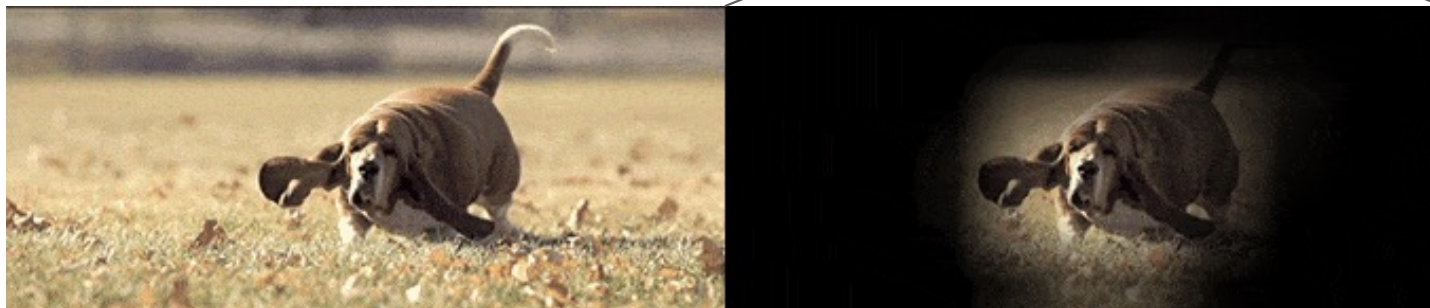
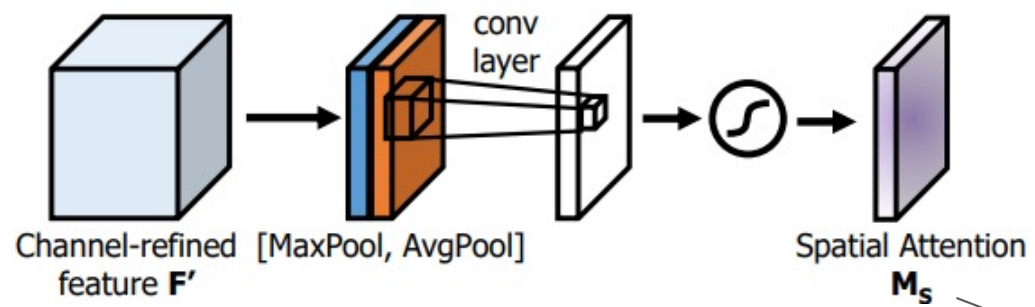
模型可视化



Activation maps created by Layer-CAM with a pretrained ResNet-18



模型可视化

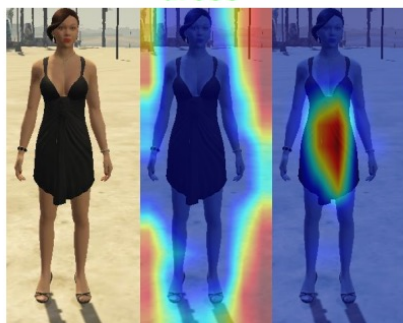


[Mohammed, IJCNN, 2020]



模型可视化

A young feamale with long hair who wears **black short dress**



(a) (b) (c)

A young male who wears **red short sleeve and a pair of black shoes**



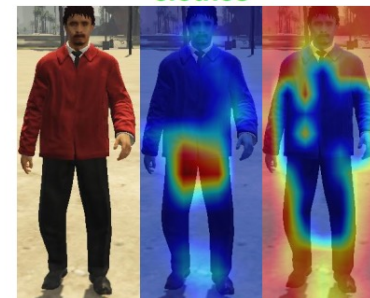
(a) (b) (c)

A young male who wears **white short sleeve and white lower-body clothes**



(a) (b) (c)

A young male who wears **red long top clothes and black lower-body clothes**



(a) (b) (c)

(a) Original Image

(b) CNN-based Method

(c) Semantic-based Method

[Xiang, arXiv, 2022]