了解人脸识别

从人工智能说起……

人工智能：造出能够模拟人类智能的机器

从数据角度而言，由于要人工构造数据集——“先人工，后智能”

$\{x_1, y_1\}$……, $\{x_n, y_n\}$，x 为输入，y为类别标签

该领域的研究包括机器人、语音识别、**图像识别**、自然语言处理和专家系统等。

人工智能的背景-纪元前

神话，幻想和预言中的AI

希腊神话中已经出现了机械人和人造人,如赫淮斯托斯的黄金机器人和皮格马利翁的伽拉忒亚。

自动人偶，许多文明中都有创造自动人偶的杰出工匠

- 偃师（中国西周）
- 希罗（希腊）
- 加扎利
- Wolfgang von Kempelen 。

加扎利的可编程自动人偶(1206年)

已知最古老的“机器人”是古埃及和古希腊的圣像

人工智能的背景-孕育

形式推理

人工智能的基本假设是人类的思考过程可以机械化。中国，印度和希腊哲学家均已在公元前的第一个千年里提出了形式推理的结构化方法。

计算机科学

- 19世纪初，查尔斯·巴贝奇设计了一台可编程计算机（“分析机”）。
- 基于图灵和冯诺依曼提出的学说，第一批现代计算机是二战期间建造的大型译码机（包括Z3，ENIAC和Colossus等）。

莱布尼兹猜测人类的思想可以简化为机械计算

人工智能的背景（诞生:1943 - 1956 ）

图灵测试

1950年，图灵发表了一篇划时代的论文，他提出了著名的**图灵测试**:

如果一台机器能够与人类展开对话（通过电传设备）而不能被辨别出其机器身份，那么称这台机器具有智能。

符号推理与“逻辑理论家”程序

50年代中期，随着数字计算机的兴起，一些科学家直觉地感到可以进行数字操作的机器也应当可以进行符号操作，而符号操作可能是人类思维的本质。这是创造智能机器的一条新路。

人工智能的背景（诞生:1943 - 1956 ）

1956年达特茅斯会议：AI的诞生

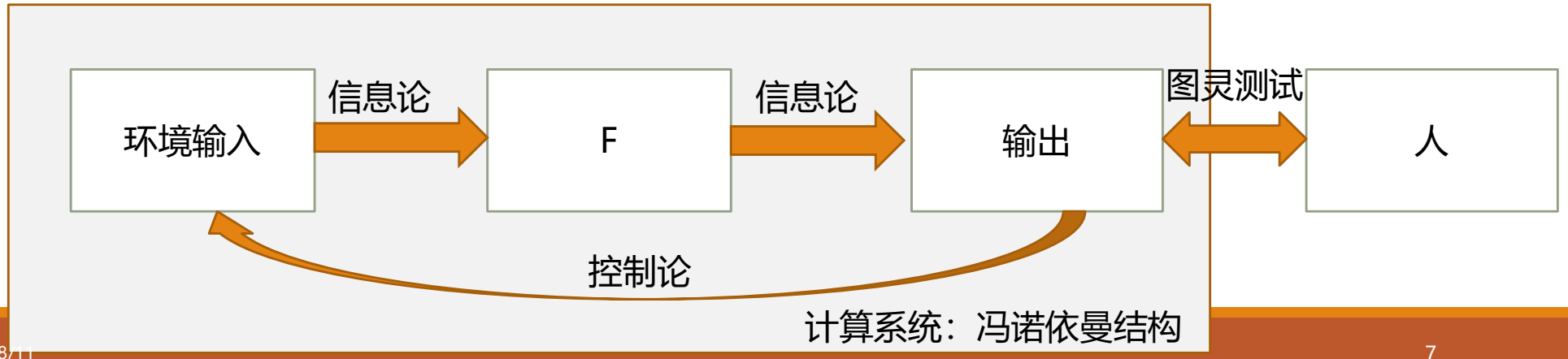
会议提出的断言之一是“学习或者智能的任何特性的每一个方面都应能被精确地加以描述，使得机器可以对其进行模拟。”

- 麦卡锡说服与会者接受“人工智能”一词作为本领域的名称。
- 1956年达特矛斯会议上确定了AI的名称和任务，因此这一事件被广泛承认为AI诞生的标志。

人工智能的背景（诞生）

人工智能(1956)建立在信息科学（before 1940s）的四大基础理论之上：

- 香农-信息论: 建立了信息的可计算性理论
- 维纳-控制论: 建立了机器系统与外界环境的交互机制
- 图灵-图灵机，图灵测试: 定义了人工智能的过程及目标
- 冯诺依曼-计算机体系结构: 建立了计算机基础架构

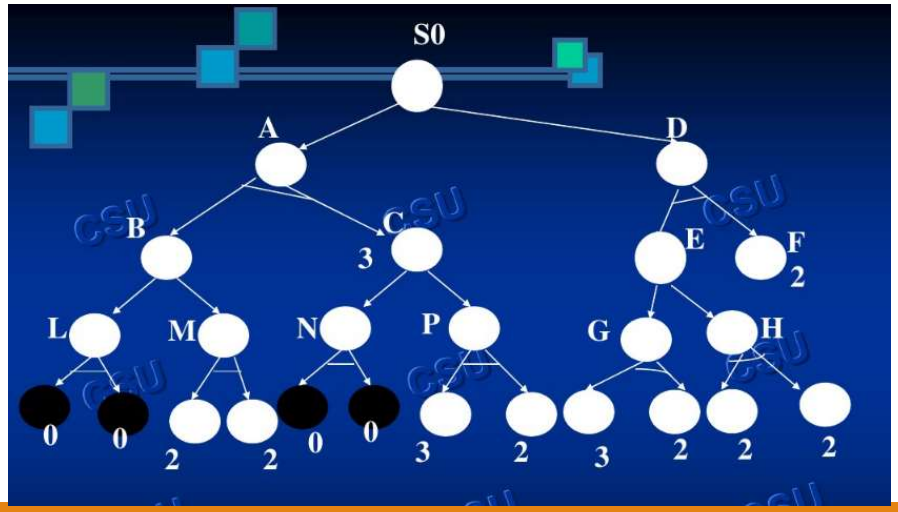


人工智能的背景（黄金年代:1943 - 1956 ）

涌现了大批成功的AI程序和新的研究方向。

搜索式推理

为实现目标（例如赢得游戏或证明定理），一步步地前进，类似于走迷宫，如果遇到了死胡同则进行回溯。这就是“搜索式推理”。



人工智能的背景（黄金年代:1943 - 1956 ）

涌现了大批成功的AI程序和新的研究方向。

自然语言

AI研究的一个重要目标是使计算机能够通过自然语言（例如英语）进行交流。早期的一个成功范例是Daniel Bobrow的程序STUDENT，它能够解决高中程度的代数应用题。

(WITH MANDATORY SUBSTITUTIONS THE PROBLEM IS)
(IF THE NUMBER OF CUSTOMERS TOM GETS IS 2 TIMES THE SQUARE 20 PERCENT OF THE NUMBER OF ADVERTISEMENTS HE RUNS, AND THE NUMBER OF ADVERTISEMENTS HE RUNS IS 45, WHAT IS THE NUMBER OF CUSTOMERS TOM GETS Q.)

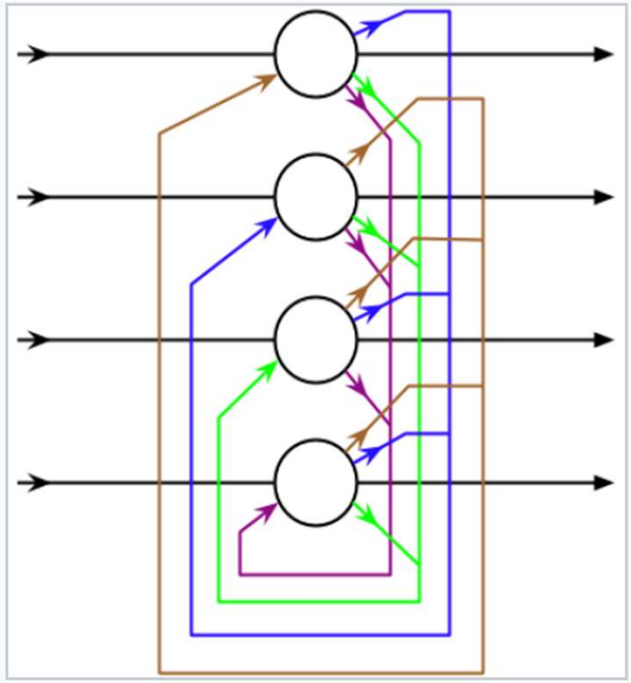
人工智能的背景（第一次AI低谷：1974 - 1980 ）

到了70年代，AI开始遭遇批评，随之而来的还有资金上的困难。AI研究者们对其课题的难度未能作出正确判断：此前的过于乐观使人们期望过高，当承诺无法兑现时，对AI的资助就缩减或取消了。 AI研究者们遭遇了无法克服的基础性障碍。

- 计算机的运算能力。
- 计算复杂性和指数爆炸。
- 常识与推理-莫拉维克悖论：人类所独有的高阶智慧能力只需要非常少的计算能力，例如推理，但是无意识的技能和直觉却需要极大的运算能力
- 框架：符号主义、连接主义、行为主义、仿生主义。。。

人工智能的背景（繁荣：1980 - 1987）

在80年代，一类名为“专家系统”的AI程序开始为全世界的公司所采纳，而“知识处理”成为了主流AI研究的焦点。日本政府在同一年代积极投资AI以促进其第五代计算机工程。80年代早期另一个令人振奋的事件是John Hopfield和David Rumelhart使联结主义重获新生。AI再一次获得了成功。



一个四节点的Hopfield网络.

人工智能的背景（第二次AI低谷：1987 - 1993 ）

“AI之冬”一词由经历过1974年经费削减的研究者们创造出来。他们注意到了对专家系统的狂热追捧，预计不久后人们将转向失望。事实被他们不幸言中：

从80年代末到90年代初，AI遭遇了一系列财政问题。

变天的最早征兆是1987年AI硬件市场需求的突然下跌。

到了80年代晚期，战略计算促进会大幅削减对AI的资助。

人工智能的背景（AI：1993 - 现在）

1997年5月11日，深蓝成为战胜国际象棋世界冠军卡斯帕罗夫的第一个计算机系统。
2005年，Stanford开发的一台机器人在一条沙漠小径上成功地自动行驶了131英里，赢得了DARPA挑战大赛头奖。
2009年，蓝脑计划声称已经成功地模拟了部分鼠脑。
2011年，IBM 沃森参加《危险边缘》节目，在最后一集打败了人类选手。
2016年3月，AlphaGo击败李世乭，成为第一个不让子而击败职业围棋棋士的电脑围棋程式。
2017年5月，AlphaGo在中国乌镇围棋峰会的三局比赛中击败当时世界排名第一的中国棋手柯洁。

人工智能的应用前景（弱人工智能）

个人助理（智能手机上的语音助理、语音输入、家庭管家和陪护机器人） 产品举例：微软小冰、百度度秘、科大讯飞等、Amazon Echo、Google Home等

安防（智能监控、安保机器人） 产品举例：商汤科技、格灵深瞳、神州云海

自驾领域（智能汽车、公共交通、快递用车、工业应用） 产品举例：Google、Uber、特斯拉、亚马逊、奔驰、京东等

医疗健康（医疗健康的监测诊断、智能医疗设备） 产品举例： Enlitic、Intuitive Sirgical、碳云智能、Promontory等

电商零售（仓储物流、智能导购和客服） 产品举例：阿里、京东、亚马逊

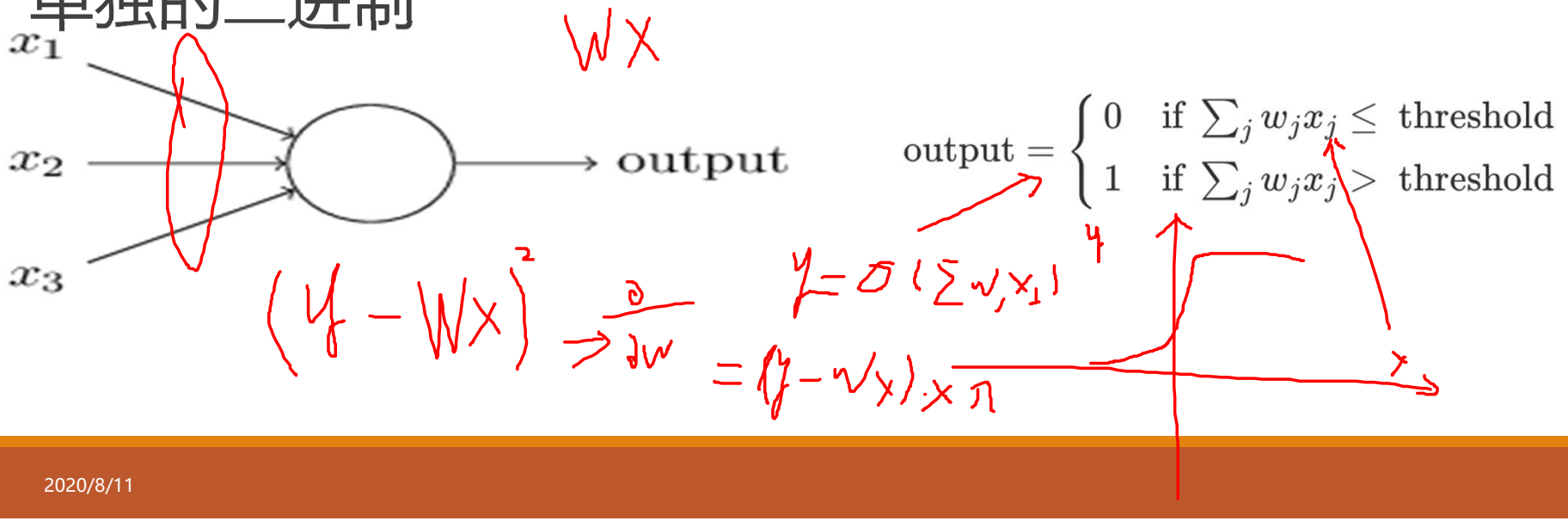
金融（智能投顾、智能客服、安防监控、金融监管） 产品举例：蚂蚁金服、交通银行、大华股份、kensho

教育（智能评测、个性化辅导、儿童陪伴） 产品举例：学吧课堂、科大讯飞、云知声

人工智能主要技术思想

(50s-80s)

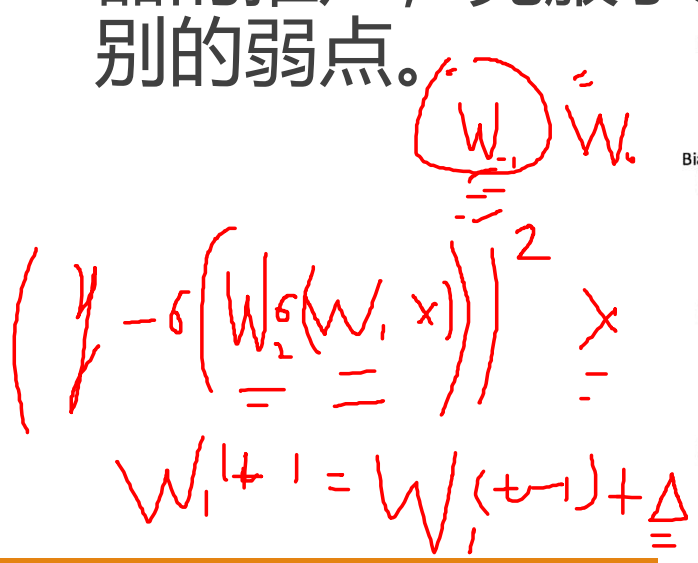
感知机由科学家Frank Rosenblatt发明于1950至1960年代，感知机的输入是几个二进制， $x_1, x_2, \ldots$， 输出是一位单独的二进制

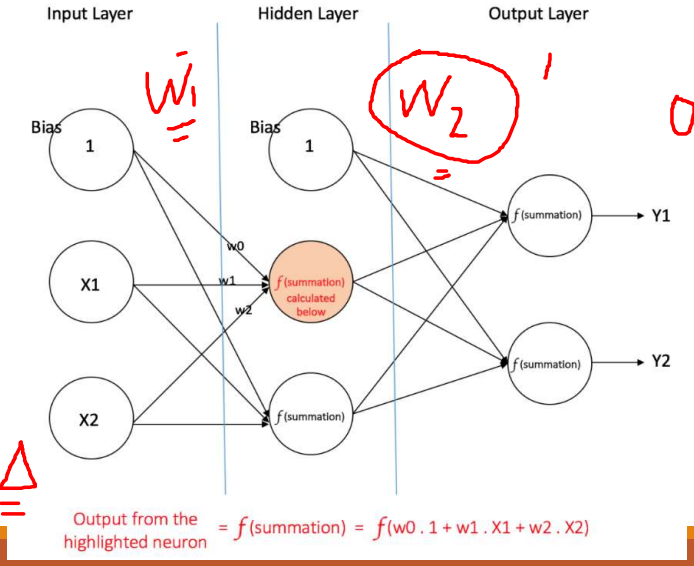


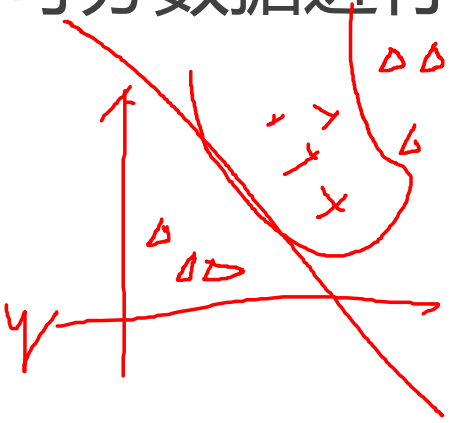
人工智能主要技术思想

(50s-80s)

多层感知器（Multilayer Perceptron,缩写MLP）是感知器的推广，克服了感知器不能对线性不可分数据进行识别的弱点。







人工智能主要技术思想

（50s-80s）

多层感知机的难题-随着网络层数增加，“梯度消失”（或者说是梯度发散）或梯度爆炸现象更加严重。

具体来说，我们常常使用sigmoid作为神经元的输入输出函数。对于幅度为1的信号，在BP反向传播梯度时，每传递一层，梯度衰减为原来的0.25。

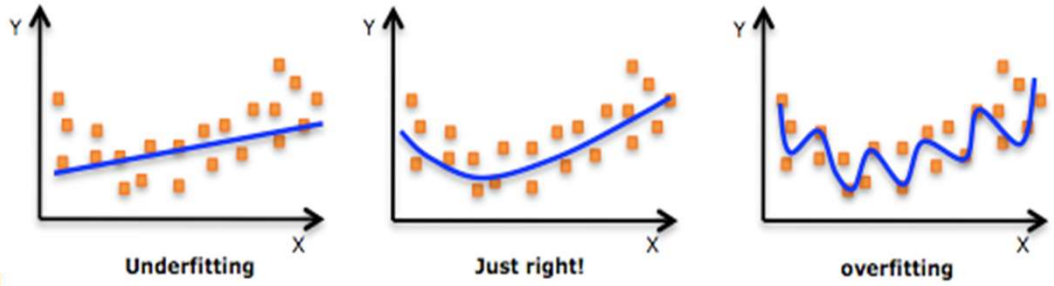
层数一多，梯度指数衰减后低层基本上接受不到有效的训练信号。而如果梯度大于1，层数一多则会导致梯度爆炸。

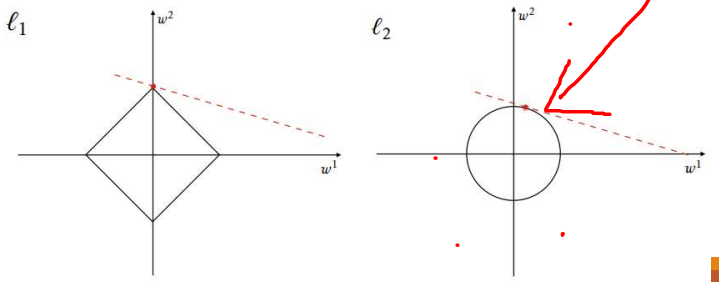
针对感知机的梯度消失和梯度爆炸问题的探索，促成了后来的多种深度学习方法的诞生：
多层RBM[Hinton 06]：逐层预训练
CNN[Lecun'97，Glorot'11]：通过卷积层设计减少Sigmoid神经元的数量和深度，替换Sigmoid 为ReLU
LSTM[Schmidhuber'97]：通过Gate和Memory机制来避免梯度爆炸和梯度
ResNet[He'15]: 通过建立短路结构Identity Mapping来实现极深网络
Densely Connected Network [CVPR'17]: 通过多层级联来增加端到端的信息路径

人工智能主要技术思想

(80s-00s)

正则化：一个假设在训练数据上能够获得比其他假设更好的拟合，但是在训练数据外的数据集上却不能很好地拟合数据，此时认为这个假设出现了过拟合的现象。正则化是针对过拟合问题的一种解决方法，即在进行目标函数或代价函数优化时，在目标函数或代价函数后面加上一个正则项，一般有L1正则与L2正则等。





人工智能主要技术思想

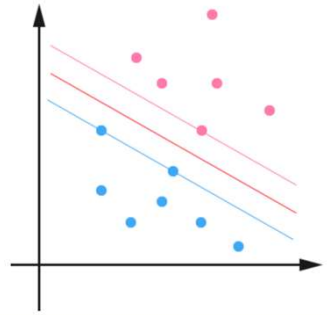
（80s-00s）

支持向量机：支持向量机（SVM, Vapnik 95）是90年代中期发展起来的基于统计学习理论的一种机器学习方法，通过寻求结构化风险最小来提高学习机泛化能力，实现经验风险和置信范围的最小化，从而达到在统计样本量较少的情况下，亦能获得良好统计规律的目的。通俗来讲，它是一种二类分类模型，其基本模型定义为特征空间上的间隔最大的线性分类器，即支持向量机的学习策略便是间隔最大化，最终可转化为一个凸二次规划问题的求解。

同期Boosting和random forest等方法也被提出

SVM的出现使得多年来深受神经网络局部解困扰的科学家们振奋不已：

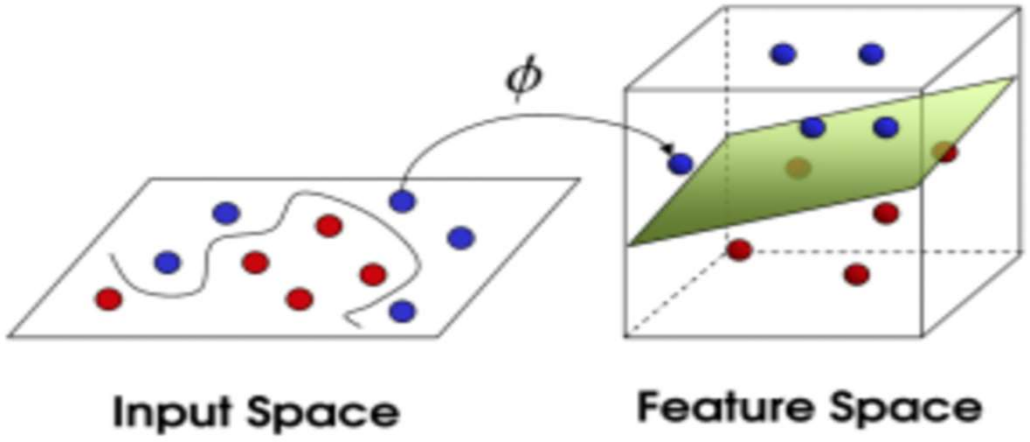
- 有唯一最优解
- 结构风险最小化是机器学习发展史上几种较直观的估计泛化能力的理论依据
- 可扩展至非线性表示空间



人工智能主要技术思想

(80s-00s)

核函数：在线性不可分的情况下，支持向量机通过某种事先选择的非线性映射(核函数)将输入变量映射到一个高维特征空间，在这个空间中构造最优分类超平面。



人工智能主要技术思想

（90s-now）

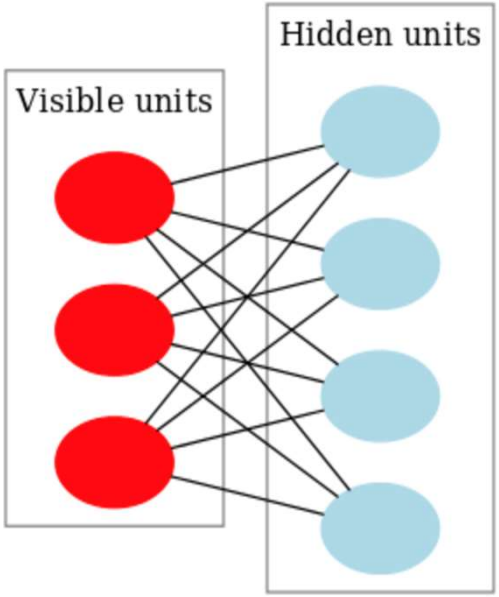
深度学习

深度学习是机器学习中一种基于对数据进行表征学习的方法。观测值（例如一幅图像）可以使用多种方式来表示，如每个像素强度值的向量，或者更抽象地表示成一系列边、特定形状的区域等。而使用某些特定的表示方法更容易从实例中学习任务（例如，人脸识别或面部表情识别）。深度学习的好处是用非监督式或半监督式的特征学习和分层特征提取高效算法来替代手工获取特征。

人工智能主要技术思想

(90s-now)

RBM：受限玻尔兹曼机（英语：restricted Boltzmann machine, RBM）是一种可通过输入数据集学习概率分布的随机生成神经网络。RBM最初由发明者保罗·斯模棱斯基于1986年命名为簧风琴（Harmonium），但直到杰弗里·辛顿及其合作者在2000年代中叶发明快速学习算法后，受限玻兹曼机才变得知名。受限玻兹曼机在降维、分类、协同过滤、特征学习和主题建模中得到了应用。根据任务的不同，受限玻兹曼机可以使用监督学习或无监督学习的方法进行训练。



人工智能主要技术思想

（90s-now）

卷积神经网络（Convolutional Neural Network, CNN）是一种前馈神经网络，它的人工神经元可以响应一部分覆盖范围内的周围单元，对于大型图像处理有出色表现。

卷积神经网络包括：

- 多个卷积层和顶端的全连通层（对应经典的神经网络）
- 关联权重和池化层（pooling layer）
- <=3层全连接层

这一结构使得卷积神经网络能够利用输入数据的二维结构。与其他深度学习结构相比，卷积神经网络在图像和语音识别方面能够给出更优的结果。这一模型也可以使用反向传播算法进行训练。相比较其他深度、前馈神经网络，卷积神经网络需要估计的参数更少，使之成为一种颇具吸引力的深度学习结构。

人工智能主要技术思想

（90s-now）

长短期记忆（Long Short-Term Memory, LSTM）是一种时间递归神经网络(RNN)，论文首次发表于1997年。由于独特的设计结构，LSTM适合于处理和预测时间序列中间隔和延迟较长的重要事件。

LSTM的表现通常比其他RNN及隐马尔科夫模型（HMM）更好：

- 2009年，在不分段连续手写识别上，基于LSTM的人工神经网络赢得ICDAR手写识别比赛冠军。
- 2013年LSTM在TIMIT自然演讲数据库上达成17.7%错误率的纪录。

作为非线性模型，LSTM可作为复杂的非线性单元用于构造更大型深度神经网络。

S. Hochreiter and J. Schmidhuber. Long short-term memory. Neural Computation, 9(8):1735–1780, 19

概念界定

深度学习/机器学习/人工智能

人工智能是最早出现的，也是最大、最外侧的同心圆；其次是机器学习，稍晚一点；最内侧，是深度学习，当今人工智能大爆炸的核心驱动。

Q & A

图像识别和模式识别、机器学习、深度学习的关系

给定一些标签，比如：汽车、狗、猫、电脑、书包

What's this?

狗!

图像识别为什么难?

计算机：数字→理解内容

图片在计算机中保存成3维数组，每个数字0-255之间

- n×m×3，n×m为像素，3为RGB三种颜色通道

所以计算机看到的是一堆数字

我眼中的图片

电脑中的图片

难点：不同形态



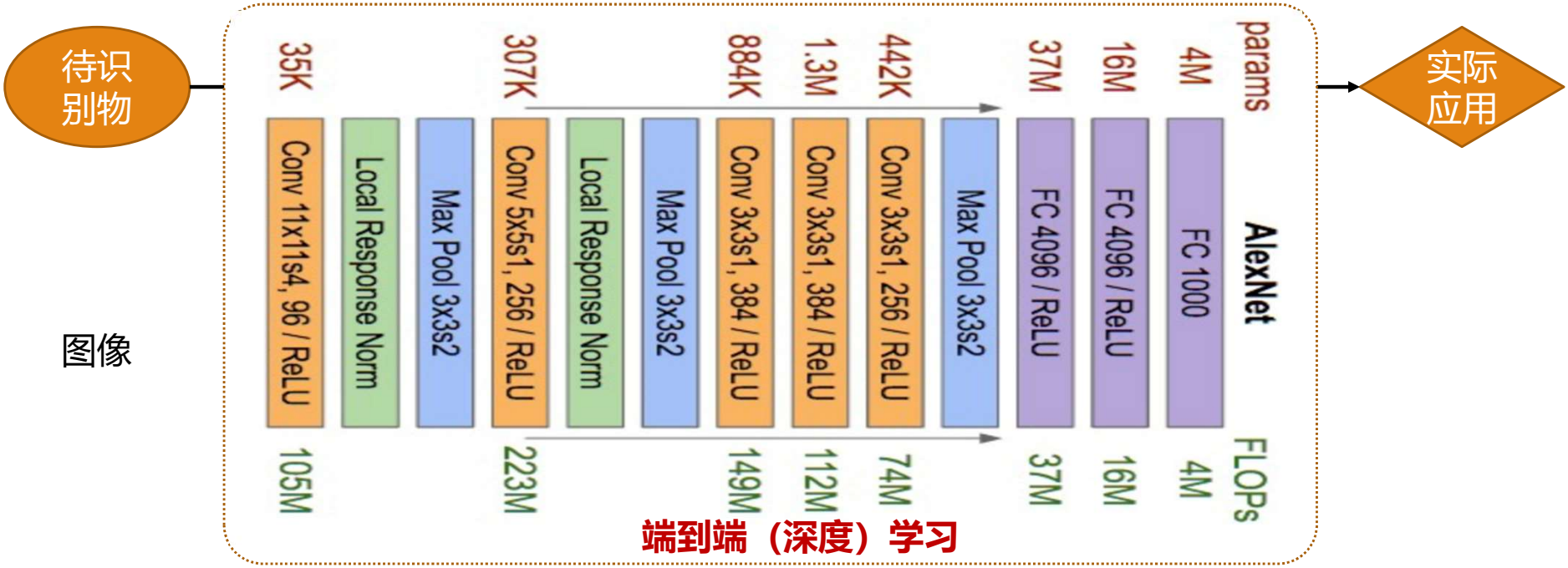

难点：拍摄角度




难点：局部遮挡

难点：光照强弱

图像识别的基本步骤


待识别物
图像
params
AlexNet
FLOPs
4M
FC 1000
4M
16M
FC 4096 / ReLU
16M
37M
FC 4096 / ReLU
37M
Max Pool 3x3s2
442K
Conv 3x3s1, 256 / ReLU
74M
1.3M
Conv 3x3s1, 384 / ReLU
112M
884K
Conv 3x3s1, 384 / ReLU
149M
Max Pool 3x3s2
307K
Local Response Norm
Conv 5x5s1, 256 / ReLU
223M
Max Pool 3x3s2
35K
Local Response Norm
Conv 11x11s4, 96 / ReLU
105M
端到端（深度）学习
实际应用

为什么人脸识别很重要?

人类最先掌握的图像识别技术

人脸识别是最先开始的图像识别研究之一

无处不在的身份验证

假设你是美国公司老板，你要在中国的分公司引入一种门禁系统

- 禁止非本公司的人进入
- 员工上班打卡、签到

传统方法

- 门禁卡、身份证
- What you know
  - Password

传统方法有什么问题?

卡丢失；卡可代刷

密码危机

- 密码遗忘
  - 纽约每天1000人以上忘记密码 ☹
- 密码被猜中
  - 生日、电话号码、车号、宿舍...
- Heavy web users have an **average of 21 passwords**; 81% of users select a common password and 30% write their passwords down or store them in a file. *(2002 NTA Monitor Password Survey)*

损失

- 2002年，仅美国330万人次的身份盗用；670万信用卡诈骗案

一种新的技术手段

生物特征识别技术Biometrics

什么是Biometrics?

- Bio——生物
- Metrics——测量

事实含义

- 通过人体自身的生理特征(what you are)或行为特征(how you do)进行身份验证的技术

哪些Biometrics?

- 指纹（感觉像罪犯，一般不让轻易采集）
- 虹膜、视网膜（不容易采集和扫描）
- 声纹（容易被外界的噪音干扰）
- ……
- 人脸！！（有个摄像头就行）

1960-1970年代

半自动的人脸识别系统

手工标记眼睛、嘴巴和鼻子等位置

程序通过位置特征与数据库中的位置特征进行匹配，从而识别人脸

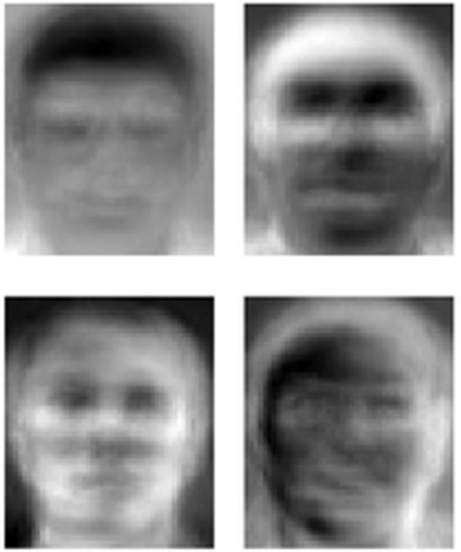
准确率不高，适用于数据库中样本量相对较少的情况

Given an image the face is matched to the face bunch graph to find the fiducial points

An image graph is created using elastic graph matching and compared to databse of faces for recognition

1988年

Kirby and Sirovich首次使用PCA（主成分分析法）进行人脸特征的提取和处理，得到特征脸(eigenface)



用少量数值就能描述复杂人脸的主要部分

1990年代后

受惠于机器学习算法（深度学习）的发展，在特定环境下，人脸识别的准确率已经可以达到99%以上，也实现了商业化，代表性方法包括：

- 基于局部纹理特征(LBP、Gabor小波)和boosting/SVM的人脸检测/识别技术
- 基于矩阵因子化（稀疏编码）的人脸特征学习技术
- 基于非线性度量学习的人脸表示学习技术
- 基于深度卷积神经网络的人脸检测与识别方法

但是并非任意环境的人脸识别都能取得很高的准确率

Why Face Recognition is Hard!

From the same person? Yes? No? …

"The variations between the images of the same face due to illumination and viewing direction are almost always larger than image variations due to change in face identity."

-- Moses, Adini, Ullman, ECCV '94

技术挑战

影响人脸图像表观的因素

人与摄像设备的位置关系（距离角度等）

光照环境条件

饰物（眼镜帽子等）

摄像设备

化妆、整容

图像存储质量

精神状态

年龄变化

健康状况

意外损伤

面部毛发(头发，胡须)


http://www.urear.cn

人脸的全局特征和局部特征

全局特征主要包括：

- 人脸的肤色特征（比如白皙、黝黑）
- 总体轮廓（比如圆脸、鸭蛋脸、方脸、长脸等）
- 面部五官的分布特征（比如，在绘画界就有“国田由用，目甲风申”8种脸形之说）

中医也将人脸按照总体结构特征划分为“金木水火土”五行（侧重人脸3D结构和肌肉凸凹情况）

局部特征则主要指面部五官的特点，比如浓眉毛、丹凤眼、鹰勾鼻、大豁嘴、八字胡须、尖下巴等，以及面部的一些奇异特征，比如黑痣、伤痕、酒窝等等

局部特征 vs 全局特征

明星漫画：夸大了独特之处

- 问题：How to find these salient features automatically?

局部特征 vs 全局特征

Thatcher Illu

人类视觉识别系统特性简介及其借鉴意义

面部特征对识别的重要性分析

- 不同的面部区域对人脸识别的重要性是不同的，一般认为面部轮廓、眼睛和嘴巴等特征对人脸识别是更重要的，人脸的上半区域对识别的意义明显比下半区域重要；鼻子在侧面人脸识别中的重要性要高于其他特征

异族人脸识别困难现象

- 这涉及到识别算法的适应性和泛化能力问题，一方面可能需要尽可能大的学习集，另一方面也需要学习集必须具有较大的覆盖能力

性别和年龄阶段对于识别性能的影响

- 女性要比男性更难识别（为什么？请思考）
- 年轻人比老年人更难识别 （为什么？请思考）

图片是怎么形成的?

通常我们所说的人脸识别是基于可见光人脸图像的身份识别与验证的简称

光学人脸图像（以下简称人脸图像）是外界光源（包括太阳光、室内人造光源和其他物体表面反射而来的光线）发出的光线照射在人脸上，经人脸表面反射后传输到摄像机传感器的光线强度的度量。

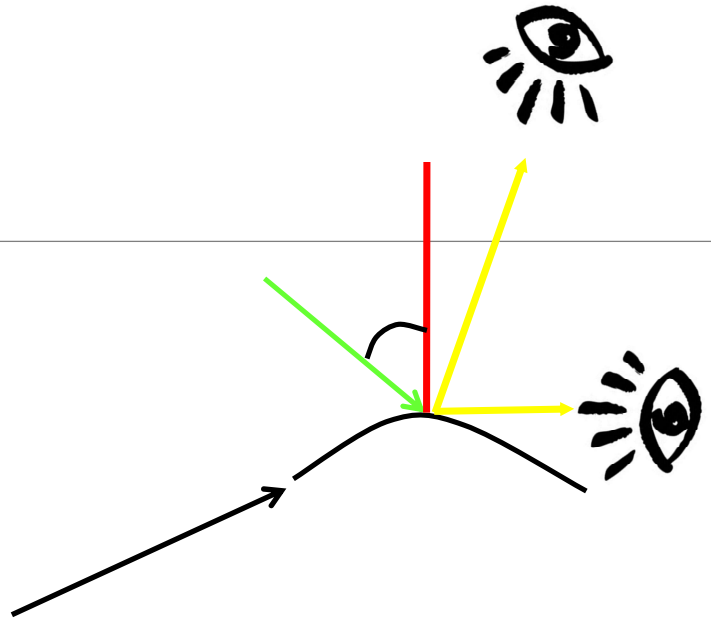
$$I = f(F; L; C)$$

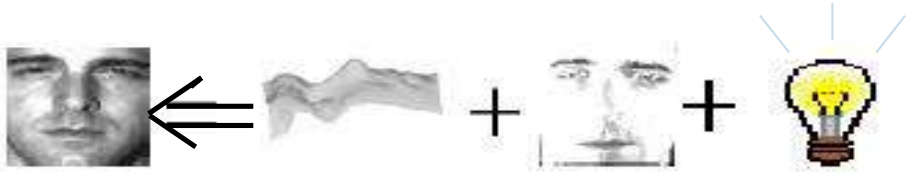
简化的人脸成像模型

Lambert反射模型（漫反射模型）

$$I = k\rho\cos(\theta) = k\rho\vec{n}\cdot\vec{s}$$

- I为相机接收到的可见光，k为光照强度，ρ, n, s分别为物体表面点漫反射系数，法向量方向，光源的方向，θ二者夹角
- 与视点无关（区别镜面反射）





Face Image = 3D Model + Texture + Illumination

人脸图像的生成要素

人脸图像实际上是三大类关键要素共同作用的结果

- 人脸内部属性F
  - **人脸3D形状（表面法向量方向）**
  - **包括人脸表面的反射属性（包括反射系数等，通常简称为纹理）**
  - 人脸表情、胡须等属性的变化；
- 外部成像条件I
  - 包括光源（位置和强度等）
  - 其他物体或者人体其他部件对人脸的遮挡（比如眼镜、帽子、头发）等。
- 摄像机成像参数C
  - 包括摄像机位置（视点）、摄像机的焦距、光圈、快门速度等内外部参数

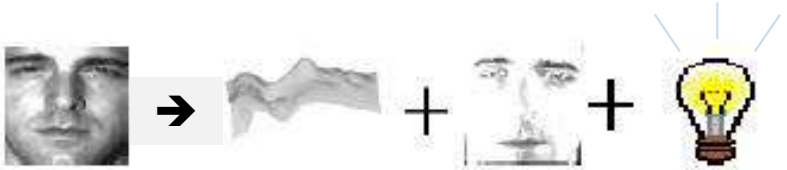
~~对识别人脸无益的干扰因素！~~

理想的识别模型

从人脸图像中剥离出

- 人脸稳定不变的本质属性（3D形状与表面反射率） **遗憾 ☹ 一个病态问题！**
- 外界条件及其摄像参数变化导致的图像变化

然后，从3D形状与表面反射率属性中提取不同人脸的差异信息，馈入到后端的判别分类器中进行识别



Face Image = 3D Model + Texture + Illumination

实际解决方案

目前的多数系统采用的人脸建模方法仍然停留在图像层面上，并没有显式地分离出3D形状和纹理的步骤

而是直接通过从“图像”中提取人脸表示特征并进行分类来完成识别

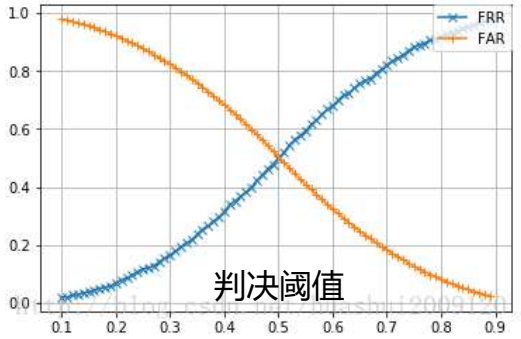
- 2D 结构信息——基于几何结构的人脸特征
- 利用**少量**3D信息进行识别的方法：Morphable models
- 2D 图像灰度数据统计特征—如模板匹配, Eigenface, Fisherface

领域技术发展现状

FAR = nontarget_is_target / (target_is_target + nontarget_is_target)
FRR = target_is_nontarget / (target_is_nontarget + nontarget_is_nontarget)

在比较良好的环境条件情况下，对基本正面人脸进行识别的性能：
- 首选识别率(Top 1 accuracy)：99%以上
- 等错误率：1%以下

在环境比较糟糕的情况下，对基本正面人脸进行识别的性能：
- 首选识别率：85%以下
- 等错误率：10%以上



我国人脸识别技术的发展状态在国际上处于领先地位（BAT等大公司都有专门团队）：
- 旷视科技（Face++）:国内最早提出并实现云识别技术
- 腾讯优图：近年来在公开评测上表现优异，其API性能也非常不错（与旷视相当甚至略优）
- 微软：云识别API，与微软系统兼容，但收费高，性能一般
- 商汤科技：不提供云识别API，一般将人脸识别集成到系统解决方案当中
- 虹软人脸：开源，免费，但性能一般，主要用在照相机当中
- 中科视拓：计算所创办公司，部分开源免费，服务中小企业和华为，在极端情况（视频）下性能有待提高

人脸图像数据库

人脸库

- FERET人脸库，1196人
- CMU-PIE姿态光照表情人脸库，68人
- CAS-PEAL人脸库，1040人，姿态表情饰物光照
- AR人脸库，126人
- XM2VTS多模态人脸库，200多人
- Yale Face Database B 光照人脸库，10人
- ORL，40人，
- Yale，15人
- VGGFace，
- **LFW**，马萨诸塞大学发布的非受限条件下人脸公开评测数据，13000人脸图片，共有5000多名人
- **MegaFace**, 美国华盛顿大学发布并维护的一套公开人脸数据集，包含一百万张图片，690000个人

http://www.easen-electron.com/news/gsxw/2017lfw.html

课程安排

第一节课：介绍，综述；作业：安装Matlab，收集10张同一个人的头像照片（照片略有变化），下载orl人脸数据库；文献阅读

第二节课：人脸检测等相关算法；作业：完成读入人脸图像模块；文献阅读

第三节课：降维技术的基本概念，PCA、LDA、流形降维；作业：完成PCA模块；文献阅读

第四节课：基于统计的机器学习，SVM；作业：完成SVM模块，调优；文献阅读

第五节课：检查完整的程序，检查必要的文档，答疑；同学讲解文献当中的部分内容（20分钟）

作业

准备10张同一个人的图片，图片之间略有变化（注意人脸框的区域）

- 自己的或者别人的

图片处理：另存为112x92像素的pgm格式的灰度图

照片放到**orl_face**/s41目录下

- orl_face一共有40个人，每个人10张照片，课后发给你们

文献阅读：

Zhao W, Chellappa R, Phillips P J, et al. Face recognition: A literature survey. ACM computing surveys (CSUR), 2003, 35(4): 399-458.