人类社会发展的重要标志是媒体与认知技术的创新。人类的发展与文明的进步，主要表现在人类不断提高对自身和客观世界的认识、不断创造新媒体的过程。

媒体：是**信息的载体**。分为**感觉媒体，表示媒体，显示媒体，存储媒体，传输媒体**。

数字媒体信息的生命周期：获取、存储、传输、分析、处理、显示、利用。

信息：**是物质相互作用中反映出的事物**（客观世界的物质现象/主观世界的意识现象）**状态与属性**（信息的定义）。信息是熵的减少。

信息熵：$H_c(X) = -\int_g p(x)\log p(x)\,dx$

互信息：$I(X,Y) = \sum_{xy} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} = H(X) - H(X|Y)$

多元正态分布的信息熵：$H_c(X) = \frac{d}{2}\log(2\pi e) + \frac{1}{2}\log|\Sigma|$

信息可以从一种表达形式转成另一种表达形式。

信息的传递：信源、信道、信宿。

电子信息科学技术服务于人的途径是电子媒体。

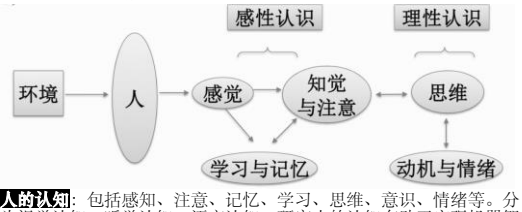
信号：信息的物理层次载体。

数据：反映客观事物运动状态的信号通过感觉器官或观测仪器感知，形成了文本、数字、事实或图像等形式的数据。数据是最原始的记录，对数据进行加工处理，使数据之间建立相互联系可以得到信息。

知识：人们在改造世界的实践中所获得的认识和经验的总和

人工智能流派：符号主义方法（认知基元是符号，智能行为通过符号操作来实现。如自动定理证明和专家系统）、连接主义方法（思维的基元是神经元，把智能理解为相互连接的神经元竞争与协作的结果。新进展：深度学习）、行为主义方法（反馈是控制论中的基石，没有反馈就没有智能。智能行为体现在系统与环境的交互之中）

媒体与认知相互作用：**认知系统利用媒体获取信息**（感官与媒体相适应；认知系统利用媒体实现主动式的信息获取，包括对客观世界和人的认知规律的认识）；**媒体对认知系统的拓展**（提升感知能力；提升认知能力一大数据；取代人的脑力劳动。人的自身认知能力是有限的，媒体技术可以为提高人类认知客观世界的能力提供有效的手段和工具）；**创造新的媒体**。



人的认知：包括感知、注意、记忆、学习、思维、意识、情绪等。分为视觉认知、听觉认知、语言认知。研究人的认知有助于实现机器智能、创造新媒体。

信息与认知的关系：信息必须经过加工、处理才能被认知。

对人的认知的研究层次：**分子层次**（脑中最基本的成份—各种分子的功能，如神经递质、蛋白质等）**细胞层次**（神经元是如何行使其功能）**系统层次**（大量神经元构成了复杂的环路完成某一功能，如视觉感知系统、运动系统）**认知层次**（神经系统是如何产生认知和协调的行为）

研究认知的技术手段：**脑电图**（是通过电极记录下来的脑细胞群的自发性、节律性电活动）、**脑成像技术**（正电子发射层描记术—使用对比剂；事件相关电位&脑磁图—高时间分辨率&低空间分辨率；磁共振成像—非侵入性、没有辐射，应用广泛）、**反应时间**

人的行为反应

刺激：围绕机体的一切外界因素，都可以看成是环境刺激因素，同时也可把刺激理解为信息。

行为：有机体对于所处情境的反应形式，心理学家将行为分解为刺激、生物体、反应三项因素研究。

心理实验：在严密控制的条件下，有组织地逐次变化条件，对相伴随的心理现象的变化进行观察，记录和测定，从而确定条件与心理现象之间关系的方法。

反应时间：从刺激的呈现到反应的开始之间的时间间隔（包含感觉器官、大脑加工、神经传入传出所需的时间以及肌肉效应器反应所需的时间，其中大脑加工所消耗的时间最多，STROOP 效应）。

感觉：人脑对事物的个别属性的认知。感觉提供了内外环境的信息，是人全部心理现象的基础。

知觉：将感觉信息组成有意义的对象，在已贮存的知觉知识经验的参与下，理解当前刺激的意义。对这种刺激意义的理解（获得）就是当前刺激和已贮存的知识经验相互作用的结果

视觉感知：光刺激作用于人眼所产生。生理机制包括折光机制、感觉机制、传导机制、中枢机制。

视觉的生理机制：视网膜是眼球的光敏感层，外层有锥体细胞（主要感受物体的细节和颜色）和棒体细胞（主要感受物体的明暗）、中间有双极细胞、内层有神经节细胞。

电信号从感受器产生以后，沿着视神经传至大脑。

视觉的传递机制由三级神经元实现：视网膜双极细胞（具有侧抑制作用）视神经节细胞（发出的神经纤维，经视交叉，传至丘脑外侧膝状体）第三级神经元（纤维从丘脑外侧膝状体发出，终止于大脑枕叶的纹状区）。

视觉系统的侧抑制：视网膜的双极细胞上，在神经节细胞的感受野里，在外侧膝状体以及视皮层细胞中都能产生侧抑制。侧抑制有利于视觉从背景中分出对象，尤其在看物体的边角和轮廓时会提高视敏度，使对比的差异增强。

视觉系统的中枢机制：人类的视觉皮层包括初级视皮层（V1，也称作纹状皮层）和纹外皮层（V2，V3，V4，V5 等）

感受野：视网膜上的一定区域受到刺激后会激活视觉系统中与这片区域有联系的各层神经细胞的活动。视网膜上的这个区域就是神经细胞的感受野。V1 视神经细胞主要有 3 种感受野：**同心圆感受野**（又称中心-周边感受野，分为 on-center 和 off-center）、**简单感受野**（对方向、位置和空间频率有明显选择性，为简单细胞，比较适合于检测具有明暗对比的边缘，且对边缘的位置和方位有严格的选择性。响应与 Gabor 滤波器相似）、**复杂感受野**（对于视觉刺激在视野中的位置没有选择性，对应于特定方向的条形刺激，具有位置不变性）。排成一条线的同心圆感受野聚合成一个简单感受野，从而对一定朝向的条形物敏感；若干个同一朝向的简单感受野，聚合到一个复杂感受野，从而使复杂感受野对任一地点的同一朝向的条形物敏感。**纹外皮层**具有更高级的视觉感知功能：在纹外皮层的第一个皮层区域，包含一些粗细条纹。对颜色有选择性的细胞集中在细条纹中、对运动方向有选择性的细胞则存在于粗条纹中、对形状敏感的细胞则在粗条纹和中间条纹中都有所分布。

知觉的信息加工过程：**自下而上加工**（数据驱动加工。是由外部刺激开始的加工，通常是先通过对较小的知觉单元进行分析，然后再转向较大的知觉单元，经过一系列连续阶段的加工，而达到对感觉刺激的解释）、**自上而下加工**（概念驱动加工。是由有关知觉对象的一般知识开始的加工。由此可以形成期望或对知觉对象的假设。这种期望或假设制约着加工的所有阶段或水平，从而调整特征觉察器，引导对细节的注意等）

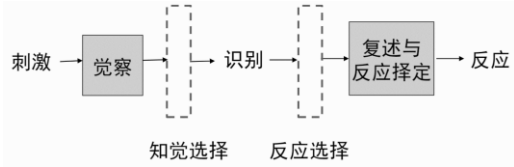
人已有的知识和知识结构，对当前的认知活动，具有决定性作用。

注意：**是人的心理活动对一定对象的指向和集中**。功能包括信号检测、选择性注意、分配性注意。特征包括：选择性、持续性、注意的转移。

注意的选择性：包括**指向性**（在每一瞬间，其心理活动或意识选择了某个对象，而忽略了另一些对象）和**集中性**（当心理活动或意识指向某个对象的时候，它们会在这个对象上集中起来）

注意选择的理论模型：**过滤器模型**（来自外界的信息是大量的，而人的神经系统高级中枢的加工能力是有限的，于是出现瓶颈。为避免系统超载，需要某种过滤器进行调节，选择其中较少的信息，使其进入高级分析阶段，这类信息受到进一步加工而被识别和存储，其他信息则不让通过。在感觉和察觉之间进行过滤。双耳分听实验。在嘈杂的环境中，也能听到别人喊自己的名字。）、**衰减模型**（又称**中期选择模型**：高级分析水平的容量有限，必须由过滤器加以调节，不过这种过滤器不只允许关注的通道的信息通过，也允许非关注的信息衰减通过，其中一些信息仍然可得到高级加工）、**反应选择模型**（几个输入通道的信息均可进入高级分析水平，得到全部的知觉加工。注意不在于选择知觉刺激，而在于选择对刺激的反应，即输出是按其重要性安排的，这种安排依赖于长期的倾向、上下文和指导语。双耳同时分听的追随靶子词实验）

过滤器模型与衰减模型的比较：不同一过滤器模型假设选择性注意的基础是对进来刺激物理属性的较粗略的分析；而衰减模型则认为，前注意分析更为复杂，甚至可能由语义加工组成。过滤器理论中的过滤器是“全或无”性质，什么都未选择的通道是完全关闭的；而衰减模型则认为未选择的通道不是完全关闭的，而只是关小或阻抑。相同—两模型的根本出发点相同：高级分析水平的容量有限或通道容量有限，必须过滤器予以调节。过滤器的位置在两模型中是相同的，都处于初级分析和高级的意义分析之间。过滤器的作业又都是选择一部分信息进入高级的知觉分析水平，使之得到识别。并且，注意选择具有知觉性质，因此，二者并称为**知觉选择模型**。

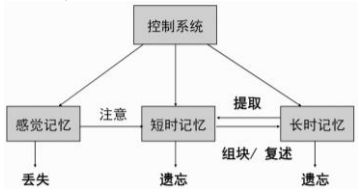


注意的认知资源分配：双加工理论：**控制性加工**（受到人的意识控制与认知资源的限制，需要注意的加工。其容量有限，可灵活用于变化着的环境。习得后加工过程较难改变）**自动加工**（不受人所控制的加工，也不受认知资源的限制，无需应用注意。没有一定的容量限制，而且一旦形成就很难改变）经过大量的练习后，可能转变为自动加工。

注意的生理机制：**朝向反射**（情景的新异性引起一种复杂而又特殊的反射。它是注意的最初生理机制）、**脑干网状结构**（脑干网状结构的激活作用使脑处于觉醒状态，是和边缘系统和大脑皮层相联系的）、**大脑皮层**（大脑皮层是产生注意的最高部位）

计算机视觉中的注意力机制：注意力机制模仿了生物观察行为的内部过程，即将内部经验和外部感觉对齐从而增加部分区域的观察精细度的机制。当前时刻输出 $h_t = \sum_{t'=1}^{T} \alpha(x_t, x_{t'}) f(x_t, x_{t'})$（注意力*特征）

记忆：**是在头脑中积累和保存个体经验的心理过程**。在信息加工的术语中，记忆是人脑对外界输入的信息进行编码、存储和提取的过程。记忆是一个系统，具有自身结构，由三个子系统构成：感觉记忆、短时记忆、长时记忆。



记忆的生理基础：皮层运动区—程序性记忆；额叶—语义与情节记忆；前额叶—短时记忆；颞叶—颞叶参与长时语义和情节记忆的整合与存储，对短时记忆中的新材料加工也起作用；杏仁核—新情绪记忆信息的整合；海马—整合新的长时语义和情节记忆；小脑—程序性记忆。

感觉记忆：又可称为**瞬时记忆**，是一种信息存储时间以毫秒或秒计的记忆，容量≥9。心理学家假设每一种感觉通道都有一种感觉记忆，每一种感觉记忆都能将感觉刺激的物理特征的精确表征保持几秒钟或更短的时间。感觉记忆是记忆系统的开始阶段，它是一种原始的感觉形式，是记忆系统在对外界信息进行进一步加工之前的暂时登记。特点：存储时间非常短（图像记忆在几百毫秒钟内，声像记忆可达 4 秒；信息加工只是初步的，但也可以进行信息整合；基本是按照刺激的物理特点进行编码，是外界刺激的真实复本）、记忆容量非常大（图像记忆在 9-20 个项目内，声像记忆容量小于图像记忆；但只有一部分信息会进入到高一级的短时记忆中）、记忆过程是无意识的自动化的，人无法控制。

短时记忆：一种信息存储时间为 1 分钟以内（约 15-30 秒）的记忆，又可以被称为**电话号码式记忆**，容量为 7±2。是个体对刺激信息进行加工、编码、短暂保持和容量有限的记忆。在短时记忆阶段，人脑同时能容纳 5-9 组内容。从短时记忆向长时记忆存入一项需要 5～10 秒钟（西蒙认为可能是 8 秒钟）。短时记忆信息的编码：在记忆系统中对信息进行转换，使之获得适合于记忆系统的形式，经过编码所产生的具体的信息形式为代码。影响因素及遗忘原因：干扰作业难度大小，记忆材料的熟悉程度；痕迹衰退说：记忆痕迹将随时间而消退；干扰说：已有信息的干扰。

长时记忆：指信息保持存储时间在一分钟以上的记忆，可以是数年甚至终生难忘，容量巨大。量的变化：存储信息的数量随时间的推移而逐渐下降。质的变化：受知识和经验差异的影响，人们存储的经验可能会发生不同程度的变化：会发生记忆的扭曲、记忆的错觉。

思维：**借助语言、表象或动作实现的、对客观事物的概括和间接的认识，是认识的高级形式**。

思维的特征：**概括性**（在大量感性材料的基础上，把一类事物共同的特征和规律抽取出来，加以概括。例如：笔是能书写的工具）**间接性**（人们借助于一定的媒介和知识经验对客观事物进行间接的认识，例如医生诊断疾病）对经验的改组（思维是探索和发现新事物的过程。它需要人们对头脑中已有的知识经验不断进行更新和改组）

思维的过程：分析和综合、比较、抽象和概括。

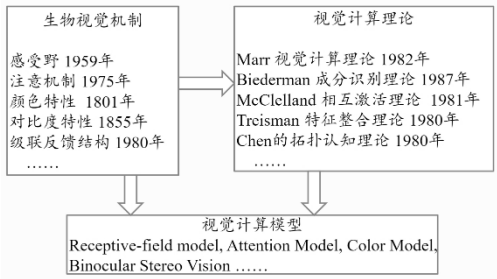
思维的种类：直观动作思维&形象思维&逻辑思维、经验思维和理论思维、直觉思维和分析思维、辐合思维和发散思维、常规思维与创造思维。

人的思维方式：心理学家认为，我们大脑中存在着两套思维系统，一套无意识运作（快速；情绪、本能反应）一套受控制运行（通常与行为、选择和专注等主观体验相关；理智思考）

批判性思维：是面对认识的对象，做出肯定什么，否定什么，或要有些什么新见解、新举措的一个系列的思考过程。要得出合理的结论必须有正确的思考方法或途径。

媒体与认知相互作用：

类脑认知计算



像素级图像校正方法：为输入图像的每个像素预测一个采样的偏移量和幅度值、基于预测的偏移量从该像素的临近位置进行重新采样（基于双线性插值）并乘以幅度值，从而得到校正图像。校正网络和识别网络端到端联合训练。

信息的获取与利用：**被动式获取**（视觉、光场相机）、**主动式获取**（超声探测、结构光成像）。**深度信息获取**：**双目测距**（利用双目摄像机拍摄物体，再通过视差估计以及成像几何关系计算物体距离）、**TOF 飞行时间**（通过专有传感器，采集近红外光从发射到接收的飞行时间，计算物体距离）**结构光**（结构光投射特定的光信息到物体表面后，由摄像机采集反射信号。根据物体表面变化造成的光信号的变化来计算物体的位置和深度等信息）

创造新媒体：视觉认知与媒体技术：**视觉暂留与电影**（视觉暂留现象视神经对物体的印象会保持 0.1-0.4 秒的时间）、**立体视觉与立体显示**（人脑将二维图像转化为三维图像：现代心理学认为这个复杂的处理过程分为生理学和心理学两个层面，具体又分为 10 种深度暗示，人们就是通过这些深度暗示来感知三维物体。线性透视、像的大小、重叠光照及阴影、结构梯度、面积透视；调节睫状体平滑肌、汇聚两眼的视轴、双目视差、移动视差。目前三维显示是基于双目视差原理，如果能同时提供人眼在生理学上的 4 种深度暗示，那么该技术可以称为“真”三维显示技术。）、**虚拟现实与增强现实**（VR：一种能够创建和体验虚拟环境的，由计算机生成的，提供多种感官刺激的自然人机交互系统。AR：在虚拟现实的基础上发展起来的新技术，也被称之为混合现实）

似动现象：两个相距不远、相继出现的视觉刺激物，呈现的时间间隔如果在 1/10 秒到 1/30 秒之间，那么我们看到的不是两个物体，而是一个物体在移动。

媒体信息处理中的注意力机制：提高模型描述能力，解决信息超载问题。

注意力系数：计算向量相似度 $s(x_i, q)$。有以下四种模型：

加性模型：$u^T \tanh(W_x x_i + W_q q)$

点积模型：$x_i^T q$

缩放点积模型：$x_i^T q/\sqrt{d}$

双线性模型：$x_i^T W q$

在特征表示中引入加权系数：卷积神经网络对多通道特征图进行加权循环神经网络对特征序列进行加权。基于自注意力机制的 Transformer 对特征序列进行加权。

卷积神经网络中的注意力机制：特征图反映了基于空间位置和通道的注意力机制。

基于空间的注意力机制：Hard Attention—选取图像中一个子区域进行处理、Soft Attention—利用空间位置掩模作为注意力加权系数。

基于通道的注意力机制：Squeeze-and-excitation。Squeeze 进行权重预测分支对该输入特征图进行按通道的全局平均池化得到维度尺寸为 1×1×1000 的向量；Excitation 通过两层全连接网络以及中间的 relu 激活层最后通过 sigmoid 函数输出维度尺寸为 1×1×1000 的权重向量，分别乘以对应的 1000 个通道的特征图作为输出。

自注意力机制：对于时序长度可变的特征向量序列，建立非局部依赖关系。全连接无法处理变长问题、卷积网络依赖输入数据的局部时序循环神经网络不易学习长距离时序依赖关系，故使用连接权重 α_{ij} 由注意力机制动态生成的“全连接”：对于输入序列 X，生成 Query、Key Value，计算网络中的隐含表示 H：

$h_i = att((K,V), q_i) = \sum_{j=1}^{N} \alpha_{ij} v_j = \sum_{j=1}^{N} softmax(s(k_j, q_i)) v_j$。

多头（multi-head）自注意力：特征矩阵中的通道拆分采用多头分支并行计算方式。将各分支计算结果并接得到最终输出。

Transformer Encoder 编码器中其它技术：位置编码、层归一化、残差连接。

机器学习：如何通过计算的手段，利用经验（数据）来改善系统自身的性能。从数据中产生“模型”，即“学习算法”。

机器学习建模任务：**监督学习**（训练样本有真值标签，如**回归、分类**）、**非监督学习**（训练样本无标签，如**聚类**）

模式是人们根据需要及特定的环境条件，对自然事物形成的抽象分类概念。**样本**是自然界的具体事物，具有一定的类别特性，是抽象模式的具体体现。**特征**是某一事物表现出来的特点与表征，是区别于其他事物的关键。

模式识别/模式识别：寻找事物和现象的相同与不同之处，根据使用目的进行分类、聚类和判断。具有多样性和多元化，可以在不同的概念粒度上进行。主要方法有早期的模板匹配、结构模式识别、统计模式识别、人工神经网络（深度学习）

特征提取：对输入的样本观测数据进行处理或变换，得到有利于分类的特征向量

模式识别系统：**数据获取**-信号空间；**预处理**；**特征提取与变换**-特征空间；**分类决策**-模式空间。

设计模式识别系统的阶段：训练、测试。训练使用训练集&验证集数据测试使用测试集数据。

生成式模型：对特征向量和类别的联合概率分布建模。如最小距离分类器

判别式模型：直接用函数对分类决策面进行建模，如神经网络。

线性回归：拟合线性变化的数据。模型为 $y = w^T x + b$。目标函数是均方误差。迭代求解方法：参数初始化、迭代（计算梯度、更新参数）

感知机：用于二分类任务，模型为 $y = \text{sign}(w^T x + b)$，类标签为 ± 1，目标函数为误分类点到超平面的总距离。属于鉴别式模型。

感知机的损失函数：$\Sigma_{x \in D} \max(0, -y(w^T x + b))$

逻辑回归：用于二分类任务，模型为 $y = \sigma(w^T x + b)$，类标签为 0&1 目标函数为交叉熵。

交叉熵目标函数：$H(P,Q) = -\int_{\mathbb{R}} p(x)\log q(x)\,dx$

异或问题：线性不可分情形。单个人工神经元（感知机、逻辑回归模型）为线性分类模型，不能解决异或问题。

批量梯度下降法：利用所有 N 个训练样本计算平均梯度

随机梯度下降法（SGD）：随机选取单个样本计算梯度

小批量随机梯度下降法：每次选 Batch Size 个样品计算平均梯度。N 个样本训练 1Epoch 的迭代次数为（N+M-1)/M

模型的评估：**容量**（模型复杂度，模型参数量大小等）、**误差**（训练误差/经验误差、测试误差/泛化误差）、**模型的选择**（拟合能力强的复杂模型容易过拟合、限制复杂度降低拟合能力可能欠拟合）

过拟合：模型在训练集上过度学习，把训练样本部分自身性质当作所有潜在样本的一般性质，缺乏对问题本质的理解，表现为训练误差小，但测试误差大，出现泛化性能下降。

欠拟合：训练集样本的一般性质没学好，表现为训练误差大。

模型的泛化能力：由模型复杂度、数据的分布以及学习任务的难度共同决定。

错误率：分类错误的样本占样本总数的比率。

识别率：1-错误率

两类模式分类的**混淆矩阵**：[[TP, FN], [FP, TN]]

召回率（真阳性率）：TP/(TP+FN)

准确率：TP/(TP+FP)

F1 分数：2/(1/召回率 + 1/准确率)

假阳性率：FP/(TN+FP)

交叉验证方法：K-折交叉验证、留一法。

前馈神经网络(FNN)：MLP、CNN。

人工神经网络(ANN)：一种机器学习方法。

多层感知机 (MLP)：在输入层和输出层之间增加一个或多个隐含层，每层有多个神经元节点。

激活函数：$\sigma'(x)=y(1-y), \tanh' x=1-y^2, \text{relu}' x=I_{x>0}$

目标函数：均方误差（MSE，误差平方的均值/2）、交叉熵（CEL，当多分类的真值为 i 时，为$-\ln q_i$）

softmax 的导数（真值是 i）：$\frac{\partial q_i}{\partial z_j}=q_i(\delta_{ij}-q_j)$

深度学习（DL）：基于人工神经网络的机器学习技术，通过构建具有多个隐含层的深层神经网络，来实现数据驱动的模型参数学习。

基本深度学习网络架构：**多层感知机**（MLP）、**卷积神经网络**（CNN）、**循环神经网络**（RNN）

初始化方法：**预训练初始化**、**随机初始化**（Gaussian 分布初始化、均匀分布初始化）、**固定值初始化**（偏置通常用 0 来初始化）

超参数：层数、每层神经元个数、学习率（以及动态调整算法）、正则化系数、mini-batch 大小

数据预处理：**图像大小归一化**、数值动态范围归一化（**最小最大值归一化**、**标准化**、**PCA**）

优化方法：调整学习率（分段衰减、指数衰减）、梯度估计调整动量法）、梯度估计调整+自适应学习率（Adam）、SGD。**有助于优化的方法**：使用线性函数、增加跳跃链接、增加隐含层辅助代价函数。

正则化方法：**防止过拟合**。可增加优化约束（**L1 或 L2 约束**、**数据增强**）或干扰优化过程（**权值衰减**、**Dropout**、**提前停止**）

Lp 正则化：在目标函数上添加 $\lambda L_p(\theta)$ 惩罚项，以参数的 p 范数作为惩罚，λ为正则化系数。$L_{\infty}(\boldsymbol{x})=\max \boldsymbol{x}, L_0(\boldsymbol{x})=\#(\boldsymbol{x})$。

权值衰减：参数更新时引入衰减系数。$\theta_t=(1-\lambda)\theta_{t-1}-\eta g_t$。在标准的随机梯度下降中和 L2 正则化效果一样。

丢弃法 Dropout：在训练过程中每次迭代时，将各层的输出节点以一定概率随机置为 0。

提前停止：如果在验证集上的错误率不再下降，就停止迭代。

数据增强：对图像进行转变，引入噪声等方法来增加数据的多样性。

卷积层：每个神经元的输入为前层输出的局部区域，通过卷积计算提取该局部区域的特征。**局部感受野，权值共享，稀疏连接**。

卷积核的参数：长、宽、输入通道数；移动步长 Stride、边界延拓 padding、空洞卷积的膨胀率 dilation（默认为 1）、分组数目。

卷积层的尺寸：$C_{out}\times K_H\times K_W$。

卷积层输出特征图大小：$W_{out}=(W_{in}-K_W+2P)/S+1$

卷积层参数量：$(K_H\times K_W\times C_{in}+1)\times C_{out}$（含偏置量）

池化层（亚采样层、汇聚层）：做尺度变化、参数降维。对卷积层输出的特征图进行亚采样，在保留有用信息的基础上减少数据处理量，实现信息的汇聚。有**最大值池化&平均值池化**。

分组卷积：在深度上进行划分，即某几个通道编为一组，相应的，卷积核深度等比例缩小而大小不变；利用每组的卷积核同它们对应组内的输入数据卷积，得到了输出数据以后，再连接组合；分组后并行计算。即将一个卷积核沿深度方向切成两个卷积核。

深度可分离卷积：每一个通道用一个 filter 卷积之后得到对应一个通道的输出，然后再进行信息的融合。深度可分离卷积比普通卷积减少了所需要的参数。重要的是深度可分离卷积将以往普通卷积操作同时考虑通道和区域改变成，卷积先只考虑区域，然后再考虑通道。实现了通道和区域的分离。

可变形卷积：引入卷积核偏置量的学习。

反向传播：

$Z_m=\sum_{j=0}^{K-1}W_jx_{m+j}\Rightarrow\frac{\partial z_m}{\partial W_j}=x_{m+j},\frac{\partial z_m}{\partial x_i}=W_{i-m}(0\le i-m\le K-1)$。

批量归一化：逐层对数据进行尺度归一化，用于**缓解梯度消失问题**。对特征图按每一批次样本计算均值和方差，对各维度归一化处理得到标准正态分布数据。$y=\beta+(x-\mu)*\gamma/\sqrt{\sigma^2+\varepsilon}$，$\beta,\gamma$需要学习。训练完成时计算整个数据集上的均值$\boldsymbol{\mu}$和方差$\boldsymbol{\sigma}$用于测试过程。测试过程中每个批次计算所用的均值和方差采用训练完成时得到的统计量。批量归一化操作可以看作一个特殊的网络层，加在每一层非线性激活函数之前。可以加速训练过程的收敛速度，防止出现梯度消失问题，也可看作一种正则化方法。

常见 CNN 结构：**LeNet**（手写数字识别，大数据+深度卷积神经网络）**AlexNet**（图像分类，GPU 训练，百万量级数据）**VGGNet**（很深，使用 3*3 卷积核）**GoogLeNet**（Inception 结构，1*1 卷积用于特征降维。利用池化后的特征图进行深度上的合并，拓宽网络的广度，增加网络的尺度适应性）、Inception-Net（利用辅助损失层以避免梯度消失）、**ResNet**（深度残差网络。针对信息传递受阻问题，在普通浅层网络中引入短接分支，转变成残差网络实现）、**ShuffleNet**（利用分组卷积，又利用 channel shuffle 解决全局信息流通不畅，网络表达能力不足的问题）、**DenseNet**（前层网络输出信息共享）、**MobileNet**（首先是采用 depthwise convolution 对不同输入通道分别进行卷积，然后采用 pointwise convolution 将上面的输出再进行结合）

图像分类：没有空间位置信息（AlexNet/VGG/GoogLeNet/ResNet）

图像分割：没有物体类别信息，只有像素分类结果。

迁移学习：在标记样本较为充足的源域任务上预训练模型，在标记样本较少的目标域任务上进行迁移学习。预训练模型、模型参数细调（根据目标域任务重设最后一个全连接层的输出节点数）

图像语义分割：通过全卷积网络实现。使用 Unpooling 或转置卷积实现升采样。

目标检测：输入图像 I，输出每个物体外接框与物体类别号（背景类别号为 0。多任务学习：分类、外接框位置回归等。

传统目标检测流程：**区域选择**（如穷举策略，采用不同的大小、不同长宽比的滑动窗口对图像进行遍历，时间复杂度高）、**特征提取**（如 HOG (histogram of oriented gradient)）、**分类器分类**（如 SVM (support vector machine) ）

基于深度学习的目标检测方法：**单阶段法**（图像经过网络一次可以预测出所有的边界框，速度相对较快，适合移动平台计算。如 SSD, YOLO 等）、**两阶段法**（先对输入图像生成候选框，即可能包含物体的区域，然后再对每个候选框进行分类（也会修正位置）。如 Faster R-CNN 等）

单发多框检测（SSD）：比 RastRNN 快，使用了不同尺度的特征图。

YOLO（UonlyLookOnce）：速度更快，只使用了最高层特征图。

目标检测思路：可能的检测框过多，则在有限集合上进行优选；分类方法未定义，则结合分类任务进行训练；冗余的输出，则基于物体实例的输出；前景背景样本数量不平衡，则合理的采样、设计目标函数、多尺度处理等。

交并比 $J(A,B)=|A\cap B|/|A\cup B|$

目标检测算法的评价：若预测框与真值框的 IoU 大于阈值，且类别相同，且该真值框未被其它预测框成功预测，则为 True Positive。否则 FP。

序列建模方法：隐含马尔可夫模型（HMM）、CNN、RNN

传统 RNN、**Elman 网络**（将上一时刻隐含层状态作为隐含层的反馈输入）**Jordan 网络**（将上一时刻网络输出作为隐含层的反馈输入）

权值按时间共享 $y=\text{softmax}(Vh_t), h_t=\phi(Wx_t+Uh_{t-1})$

目标函数：各时刻的目标函数之和。样本类别真值序号为 $g_t=i,\quad i\in\{1,2,\dots,C\}$，时刻 t 采用交叉熵作为目标函数 $L_t=-\log(y_t)_i$

误差随时间反向传播算法 BPTT：与 BP 原理相同，需要考虑网络按时间展开权值共享的因素。

$\frac{\partial L}{\partial U}=\sum_{t=1}^{T}\frac{\partial L}{\partial U_t}=\sum_{t=1}^{T}\sum_{s=t}^{T}\frac{\partial L_S}{\partial U_t},\ \frac{\partial L}{\partial x_t}=\sum_{s=t}^{T}\frac{\partial L_s}{\partial x_t}=\sum_{s=t}^{T}\frac{\partial L_s}{\partial h_t}\frac{\partial h_t}{\partial x_t}$

理论上需要把误差反传到最初时刻，但实际上我们采用时间截断的 BPTT，误差仅在有限时间内反传。

传统循环神经网络的缺陷：梯度消失/爆炸问题。在前馈神经网络中梯度消失问题存在，特别是激活函数采用 Sigmoid；梯度爆炸问题不存在，因为各层权值矩阵不同。

引入门控机制以解决梯度问题：gate $\sigma(\mathbf{w}_x^T\mathbf{x}_t+\mathbf{w}_h^Th_{t-1}+b)$。

长短时记忆单元 (LSTM)：输入门、输出门、遗忘门、输入节点。

Input Node: $g_t=\phi(W_{xg}x_t+W_{hg}h_{t-1}+b_g)$

Input Gate: $i_t=\sigma(W_{xi}x_t+W_{hi}h_{t-1}+b_i)$

Forget Gate: $f_t=\sigma(W_{xf}x_t+W_{hf}h_{t-1}+b_f)$

Output Gate: $o_t=\sigma(W_{xo}x_t+W_{ho}h_{t-1}+b_o)$

Cell: $c_t=i_t\odot g_t+f_t\odot c_{t-1}$

Output: $h_t=o_t\odot\phi(c_t)$

门控循环单元 GRU：更新门、重置门。

比较 LSTM 和 GRU：大部分情况下是 LSTM 效果好，但 GRU 参数少。

深层循环神经网络常用网络架构：双向长短时记忆网络及 CTC 解码（Bi-LSTM + CTC）、编码器 - 解码器（encoder - decoder）。

Hidden Unit Candidate: $\widetilde{h_t}=\phi(W_{xh}x_t+W_{hh}(r_t\odot h_{t-1})+b_h$

Hidden Unit: $h_t=z_t\odot h_{t-1}+(1-z_t)\odot\widetilde{h_t}$

Update Gate: $z_t=\sigma(W_{xz}x_t+W_{hz}h_{t-1}+b_z)$

Reset Gate: $r_t=\sigma(W_{xr}x_t+W_{hr}h_{t-1}+b_r)$

循环神经网络：循环神经单元（普通 RNN 单元，LSTM 或 GRU）可以取代神经元，一般由同一类型的多个循环神经单元组成一层，多层堆叠形成网络

双向循环神经网络 Bi-directional RNN：输入序列可按时间正向和反向分别送入两个层，两层输出按对应时刻并接得到输出序列，送入后续网络

连接时序分类 CTC：RNN 中的解码层。

支持向量机 SVM：一种基于统计学习理论的机器学习方法。小样本条件下的统计学习方法；具备严格的数学理论基础和直观几何解释；处理不均匀、离散、稀疏的数据有明显优势；适用于样本有限情况下的优化求解。

线性分类器 $g(x)=g_1(x)-g_2(x)=w^Tx+b$，样本到分界面的有向距离为 $(w^Tx+b)/\|w\|$。最大间隔线性分类器：离分界面最近的支持向量满足 $w^Tx+b=\pm1$，间隔宽度为 $2/\|w\|$。

非线性分类问题：使用核函数实现高维空间中的内积运算。以 $K(\boldsymbol{x}_i,\boldsymbol{x}_j)=\boldsymbol{\phi}(\boldsymbol{x}_i)^T\boldsymbol{\phi}(\boldsymbol{x}_j)$ 代替内积 $\boldsymbol{x}_i^T\boldsymbol{x}_j$。$\boldsymbol{\phi}$为**基函数**，可显式写出。

线性核函数：$x_i^Tx_j$

多项式核函数：$(1+x_i^Tx_j)^p$

高斯核函数：$\exp\left(-\|x_i-x_j\|^2/2\sigma^2\right)$

Sigmoid 型核函数：$\tanh(\beta_0x_i^Tx_j+\beta_1)$

贝叶斯决策：贝叶斯决策具有最小错误率。

正态分布

$$=\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma_i|^{\frac{1}{2}}}\exp\left\{-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)\right\}$$

正态分布参数估计方法：

最大似然估计 $\hat{\boldsymbol{\mu}}=\sum x_k/n, \hat{\boldsymbol{\Sigma}}=\sum(x_k-\hat{\boldsymbol{\mu}})(x_k-\hat{\boldsymbol{\mu}})^T/n$

无偏估计 $\hat{\boldsymbol{\Sigma}}=\sum(x_k-\hat{\boldsymbol{\mu}})(x_k-\hat{\boldsymbol{\mu}})^T/(n-1)$

正态分布协方差矩阵的不同情形：

最小欧氏距离分类器：$\Sigma_i=\sigma^2I$

最小马氏距离分类器：$\Sigma_i=\Sigma$（马氏距离为 $\sqrt{(x-\mu)^T\Sigma^{-1}(x-\mu)}$）

二次判别函数：$\Sigma_i\neq\Sigma_j$。忽略先验概率和常数项，得到

$$g_{QDF}(x)=-\frac{1}{2}(x-\mu_i)^T\Sigma_i^{-1}(x-\mu_i)-\frac{1}{2}\ln|\Sigma_i|$$

$$g'_{QDF}(x):=-2g_{QDF}(x)=\sum_{k=1}^{n}\frac{1}{\lambda_k^{(i)}}\left(\phi^{(i)T}_k(x-\mu^{(i)})\right)^2+\sum_{k=1}^{n}\ln\lambda_i$$

线性判别函数：各类协方差矩阵都相等，去除判别函数公式中对各类别相同的项，并忽略先验概率，得到线性判别函数。

非监督学习任务：聚类（K-均值聚类）、混合高斯模型概率密度估计（EM 算法）

基于误差平方和准则的聚类：将样本分为 k 个子集，分别计算各子集样本均值 m_i，则 $J_e=\sum_i\sum\|x-m_i\|^2$

K-means 聚类方法：输入聚类中心数目k和n个样本。为每个聚类产生一个初始聚类中心，将样本按照最小距离原则到最近邻聚类，使用每个聚类中的样本均值作为新的聚类中心，不断重复直到收敛。

K-means 性能分析：**优点**：是解决聚类问题的一种经典算法，简单、快速；当数据聚类是紧致的，而类与类之间区别明显时，效果较好。**缺点**：在类的平均值被定义的情况下才能使用，这对于处理符号属性的数据不适用；必须事先给出 k；对于不同的初始值，可能会导致不同结果；对于“噪声”和孤立点数据较为敏感。**改进**：不同初始值导致不同结果，可多设置一些不同的初始值（但比较耗时和浪费资源）；分类数目 K 不确定，可使用 ISODATA 算法，让类自动合并和分裂。

混合高斯模型：$P(x|\omega)=\sum_k\pi_kN(x|\mu_k,\Sigma_k),\ \sum_k\pi_k=1$，比高斯模型具有更强的描述能力，但其需要的参数也成倍增加，实际中通常对节点方差矩阵结构进行约束。如对角方差、对角共享方差。

期望-最大化 EM 算法：非监督参数估计方法。初始化 K 个高斯分布的参数 μ_k,Σ_k,π_k，保证 $\sum\pi_k=1$；求 x_n 属于第 k 类的概率 $\gamma(z_{nk})$；更新高斯分布参数

$\mu_k^{new}=\frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})x_n$ ，$\pi_k^{new}=\frac{1}{N}\sum_{n=1}^{N}\gamma(z_{nk})$

$\Sigma_k^{new}=\frac{1}{N_k}\sum_{n=1}^{N}\gamma(z_{nk})(x_n-\mu_k^{new})(x_n-\mu_k^{new})^T$

EM 算法对类别标号赋有概率度量，而 K-Means 算法是硬判决。EM 算法迭代使似然函数最大就相当于 K-means 迭代使失真函数达到最小。

马尔可夫过程：无后效性的随机过程。

马尔可夫链：参数集和状态空间都是离散的马尔可夫过程。

序列建模问题：X 为特征序列，W 为类别标记序列，$p(x|W_i)$ 称为声学模型，$p(W_i)$ 称为语言模型。

隐含马尔可夫模型（HMM）：一个双重随机过程。模型为 $\lambda=(\pi,A,B)$ A 为与时间无关的状态转移概率矩阵，B 为给定状态下观察值概率分布 π为初始状态空间的概率分布。可以研究评估问题（计算给定观测序列出现概率）、解码问题（根据给定观测序列计算最优状态序列）、学习问题（根据观测序列估计模型参数）。实现中的问题：初始模型选取、数据下溢问题（概率递推计算中数值趋于 0，可以增加比例因子或取对数）

前向算法：$\alpha_t(i)\coloneqq P(O_1,\dots,O_t,q_t=S_i|\lambda),1\le t\le T,1\le i\le N$

初始化 $\alpha_1(i)=\pi_ib_i(O_1)$、递归 $\alpha_{t+1}(j)=[\alpha_tA]_jb_j(O_{t+1})$，最终 $P(O_1,\dots,O_t|\lambda)=\sum_i\alpha_T(i)$

后向算法：$\beta_t(i)\coloneqq P(O_{t+1},\dots O_T|q_t=S_i,\lambda),1\le t\le T,1\le i\le N$

初始化 $\beta_T(i)=1$、递归 $\beta_t(i)=\left[A\left(b_j(O_{t+1})*\beta_{t+1}(j)\right)\right]_i$、最终 $P(O|\lambda)=\sum_i\beta_1(i)\pi_ib_i(O_1)$

前向-后向算法：$P(O,q_t=S_i|\lambda)=P(O_1,\dots,O_t,q_t=S_i|\lambda)P(O_{t+1},\dots,O_T|q_t=S_i,\lambda)=\alpha_t(i)\beta_t(i)$，$P(O|\lambda)=\sum_i\alpha_t(i)*\beta_t(i)$。

Viterbi 算法：处理解码问题。用动态规划求最优路径概率，使每一时刻状态序列出现相应观测值的可能达到最大。与前向算法的区别在于前向算法累计所有路径的概率，而 Viterbi 只计算最优路径的概率。

定义 $\delta_t(i)=\max P(q_1,\dots,q_{t-1},q_t=i,O_1,\dots,O_t|\lambda)$

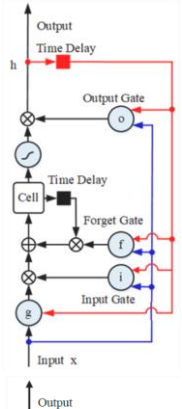
初始化：$\delta_1(i)=\pi_ib_i(O_1),1\le i\le N,\varphi_1(i)=0,1\le i\le N$。

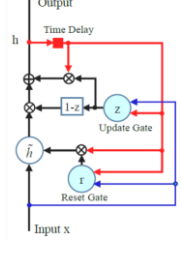
递归：$\delta_t(j)=\max[\delta_{t-1}A]\,b_j(O_t),\varphi_t(j)=\arg\max[\delta_{t-1}A]$

最终：$P^*=\max[\delta_T(i)],q_T^*=\arg\max[\delta_T(i)],q_t^*=\varphi_{t+1}(q_{t+1}^*)$

Baum-Welch 算法：求解学习问题，是一种 EM 算法，即从不完全数据（样本特征序列与隐含状态序列的对齐关系未知）中求解模型参数的最大似然估计方法：选择 HMM 初始参数 λ_0；求期望，即利用给定的 HMM 参数求样本特征序列的状态对齐结果；最大化：根据上一步的状态对齐结果，利用最大似然估计更新 HMM 参数 λ；直到收敛：$\log P(O|\lambda)-\log P(O|\lambda)<d$。

$\xi_t(i,j)\coloneqq P(s_t=i,s_{t+1}=j|O,\lambda)=\alpha_t(i)a_{ij}b_j(O_{t+1})\beta_{t+1}(j)/\sum_i\sum_j$ 分子





语音识别：语音识别是机器通过识别和理解过程把人类的语音信号转变为相应的文本或命令的技术。

语音识别的困难：连续语音流中各语音单位之间不存在明显的界线、语音特征变化差异大（音素的声学特征变化与上下文相关、随发音人及其生理或心理状态的变化而有很大的差异、环境噪声和传输设备的差异也将直接影响语音特征的提取）

语音识别分类：按词汇量大小分、按发音方式分（孤立词识别、连接词识别、连续语音识别、关键词检出）、按说话人分（特定说话人、非特定说话人）、按识别方法分（模板匹配—利用动态时间规整 DTW 将测试语音与参考模板进行匹配、随机模型—利用 HMM 对似然函数进行估计、深度学习—利用 RNN 进行端到端学习）。

奈奎斯特采样定理：语音信号采样频率 $f_s>2f_{max}$。窄带语音信号：$f_s=8kHz$，如电话语音，可以基本保持语意，不影响对语音的感知；宽带语音信号：$f_s=16kHz$，用于对语音质量要求较高的场合。

量化精度：量化所用比特数越大，声音质量越好。人类听觉系统对声音信号强弱刺激反应不是线性的，而是成对数比例关系。

语音信号的短时分析：语音信号是一种典型的非平稳信号。语音识别中常用的帧长为 20~30ms，帧移为 10ms。短时分析做加窗处理（如汉明窗）：$S_w=\sum_m s(m)w(n-m)$。

语音信号时域分析：直接分析语音信号时域波形提取特征参数（短时平均能量、短时平均幅度、短时平均过零率、短时自相关函数）

语音信号频域分析：带通滤波器组法、傅里叶变换法、线性预测法

短时傅里叶分析 (Gabor)：得到短时功率谱与语谱图。

线性预测分析：语音信号当前数据可以用过去若干数据的线性组合来预测；利用实际采样值与预测值之间的均方误差对线性组合系数进行优化求解，可得线性预测分析系数。

梅尔频率倒谱系数 MFCC：梅尔频率与实际声音频率成对数关系，对低频比高频敏感。Mel(f)=2595 lg(1+f/700)。

特征提取：对信号空间的原始数据通过处理及变换得到在特征空间中最能反映分类本质的特征。特征提取是模式识别的重要环节，其目的是提取类内差别小，类间差别大的鉴别力强的特征向量（图像特征提取—Gabor 滤波器组特征等；语音特征提取—梅尔频率倒谱系数等）

一维信号的时频分析：傅里叶变换、短时傅里叶变换（Gabor 变换）、小波变换

特征变换：在给定精度下，准确地对某些变量的函数进行估计，所需样本量会随着特征维数的增加而呈指数形式增长。通过特征变换等手段，实现特征降维。克服维数灾难；获取本质特征；节省存储空间；去除无用噪声；大数据可视化。

特征降维：特征降维常见方法：特征选择（直接选择法）、特征变换（线性—PCA、LDA；非线性—自动编码器）

主成分分析(PCA)：最小均方误差准则下保留原始数据信息。优点：采用样本协方差矩阵的特征向量作为变换的基向量，与样本的统计特性完全匹配；在最小均方误差准则下，是最佳变换。缺点：变换矩阵随样本数据而异，无快速算法。

线性判别分析(LDA)：最小均方误差准则下区分多类数据。

类内散度矩阵 $S_w=\sum_i\sum_k(x_k^i-\mu_i)(x_k^i-\mu_i)^T/N$

类间散度矩阵：$S_b=\sum(\mu_i-\mu)(\mu_i-\mu)^T/N$

总散度矩阵：$S_T=S_w+S_b=\sum_i(x_i-\mu)(x_i-\mu)^T/N$

准则函数：$J(w)=\max tr\,(W^TS_bW)/tr\,(W^TS_wW)$

优化问题：$S_bW=\lambda S_wW,\lambda\ge0\Rightarrow(S_w^{-1}S_b)W=\lambda W$。

比较 PCA 与 LDA：相同点：降维方法，特征值分解，高斯假设；不同点：PCA 无监督/LDA 有监督；LDA 可用于分类、LDA 取分类性能最好的投影方向，PCA 取投影点方差最大的方向。

自动编码器：一种尽可能复现输入信号的神经网络，找到代表输入数据的最重要的因素。线性自动编码器等同于 PCA。

其他实用数据降维方法：t-SNE—是一种考虑数据概率分布的非线性特征降维方法；UMAP—与 t-SNE 类似，一种快速的非线性降维方法；