



人工智能基础 及在高能物理领域的应用

中国科学院高能物理研究所

计算中心

张易于

2023-08-18



CONTENTS

- 1.人工智能
- 2. AI在高能物理领域的应用
- 3. 总结



01

人工智能



人工智能

人工智能（Artificial Intelligence），英文缩写为AI。它是研究、开发用于**模拟、延伸和扩展人的智能**的理论、方法、技术及应用系统的一门新技术科学。

- ✓ 结构模拟：**机器人学**
- ✓ 功能模拟：以**任务**为核心
机器学习、深度学习

自然语言处理、音频识别、定位跟踪、图像理解、知识推理、数据预测等等

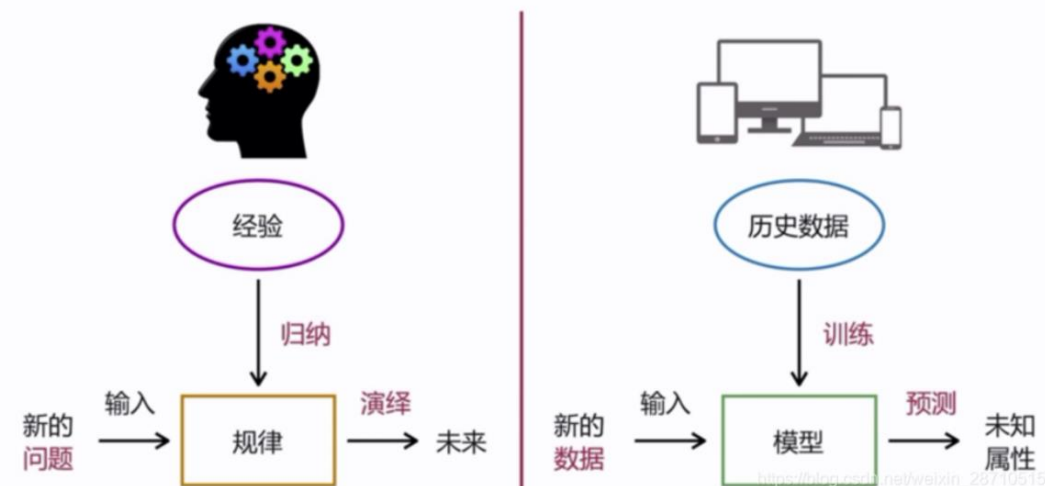


人工智能

使一部机器的反应方式像人一样进行**感知、认知、决策、执行**的人工程序或系统

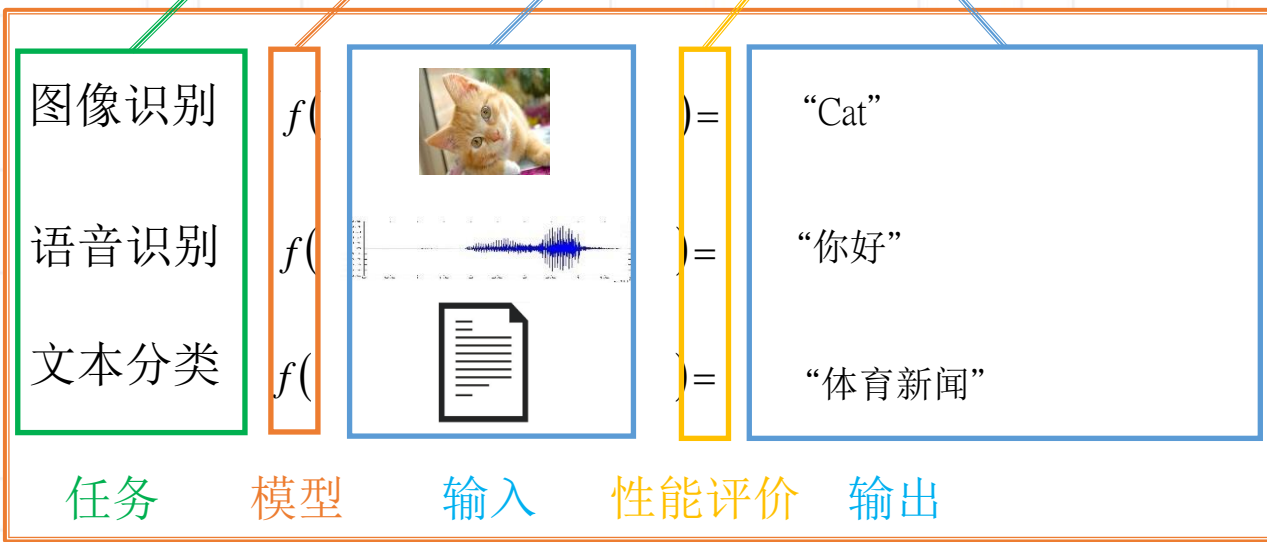
机器学习

- 机器学习是人工智能的一个分支
- 计算机有能力从数据中学习
- 数据 · 模型 · 算法: 算法通过在数据上进行运算产生模型
- 常见的机器学习算法：
 - 线性回归、逻辑回归、决策树、随机森林。

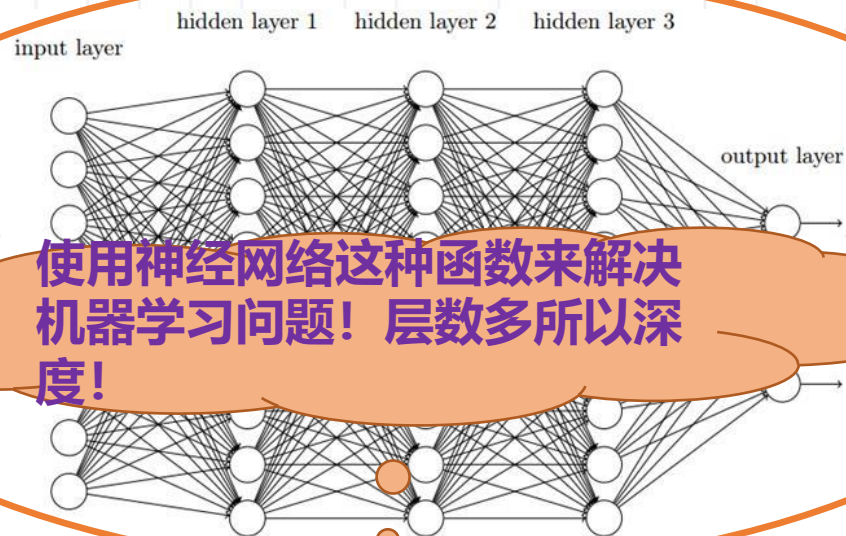


机器学习

对于某类任务T (Task) 和性能度量P (Performance Measure)，一个计算机程序被认为可以从经验E (Experience) 中学习是指，通过经验E改进后，计算机程序在任务T上由性能度量P衡量的性能有所提升



深度学习



使用神经网络这种函数来解决机器学习问题！层数多所以深度！



人是地球上具有最高智慧的动物，人类靠大脑进行思考、联想、记忆和推理判断。

建立模仿人类大脑的模型

图像识别

$f(\text{Image of a cat}) = \text{"Cat"}$

语音识别

$f(\text{Audio of "Hello"}) = \text{"你好"}$

文本分类

$f(\text{Text of sports news}) = \text{"体育新闻"}$

模型

神经元

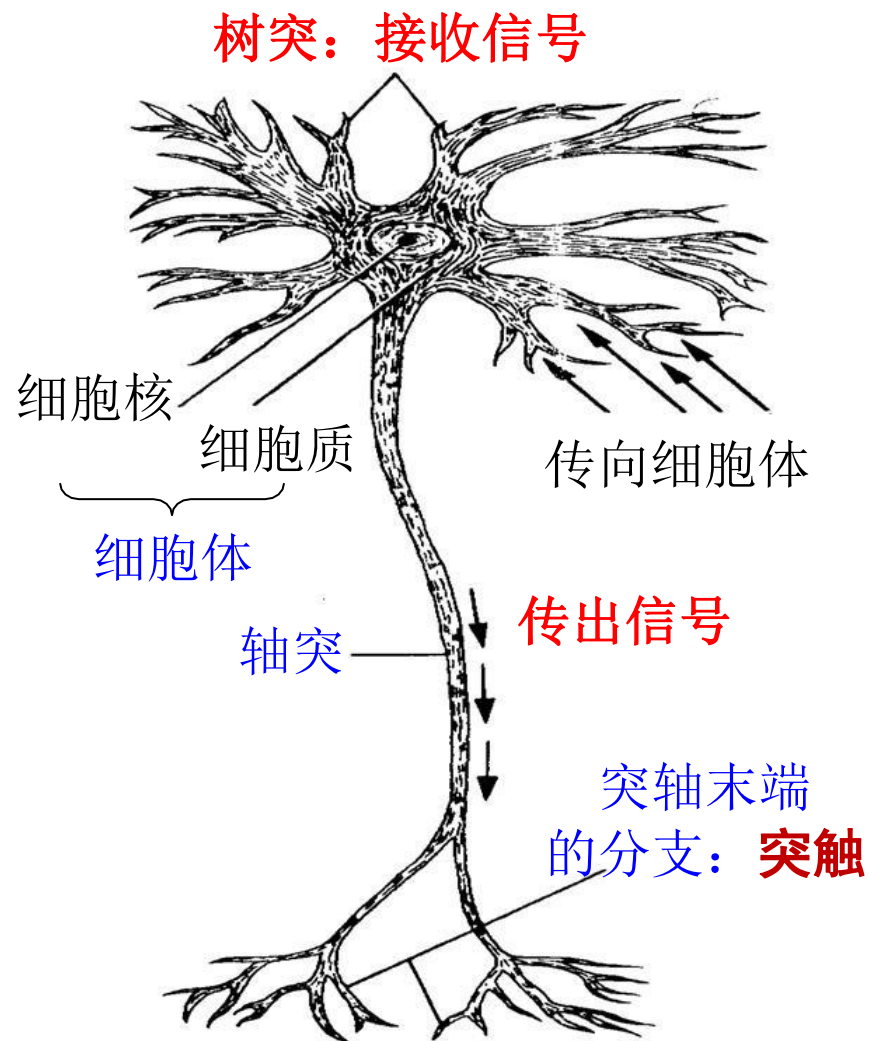
生物神经元

生物神经元

- ① 神经元间通过突触两两相连
- ② **树突**接收来自多个神经元的信号
- ③ **轴突**根据树突传递过来的综合信号的强弱是否超过某一**阈值**来决定是否将该信号传递给下一个神经元

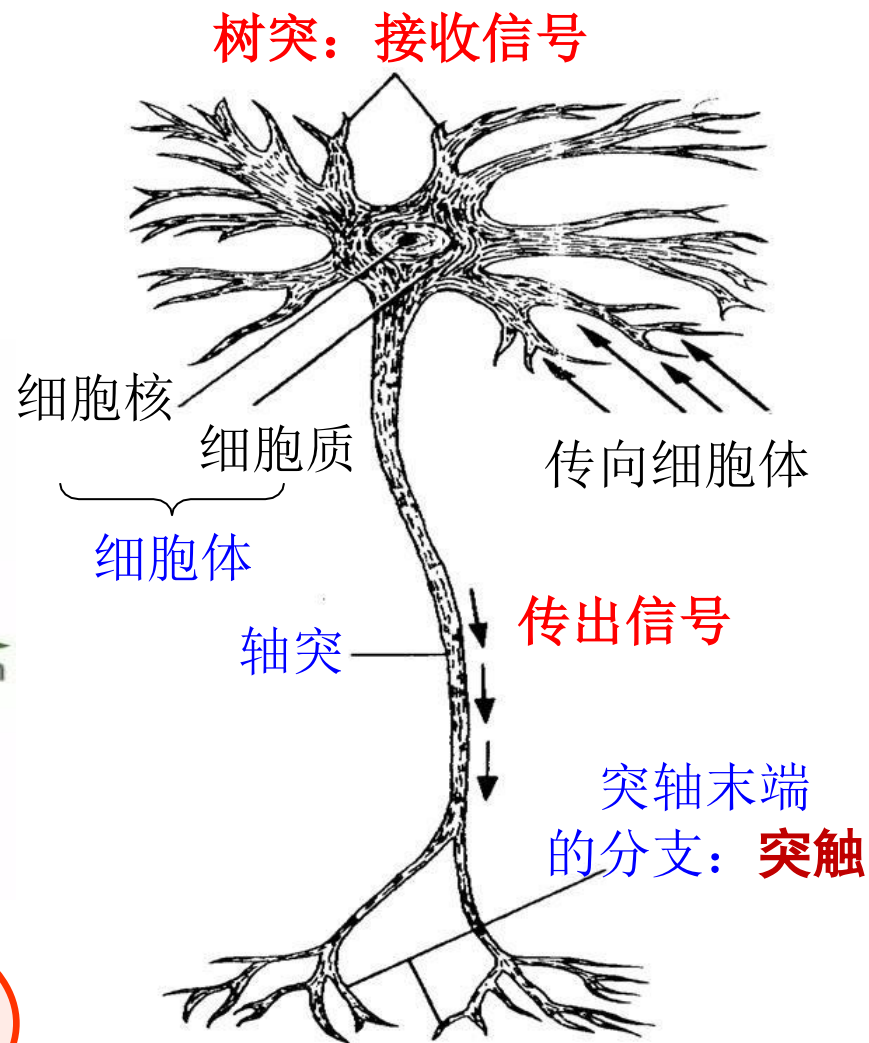
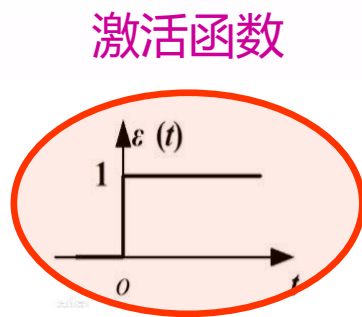
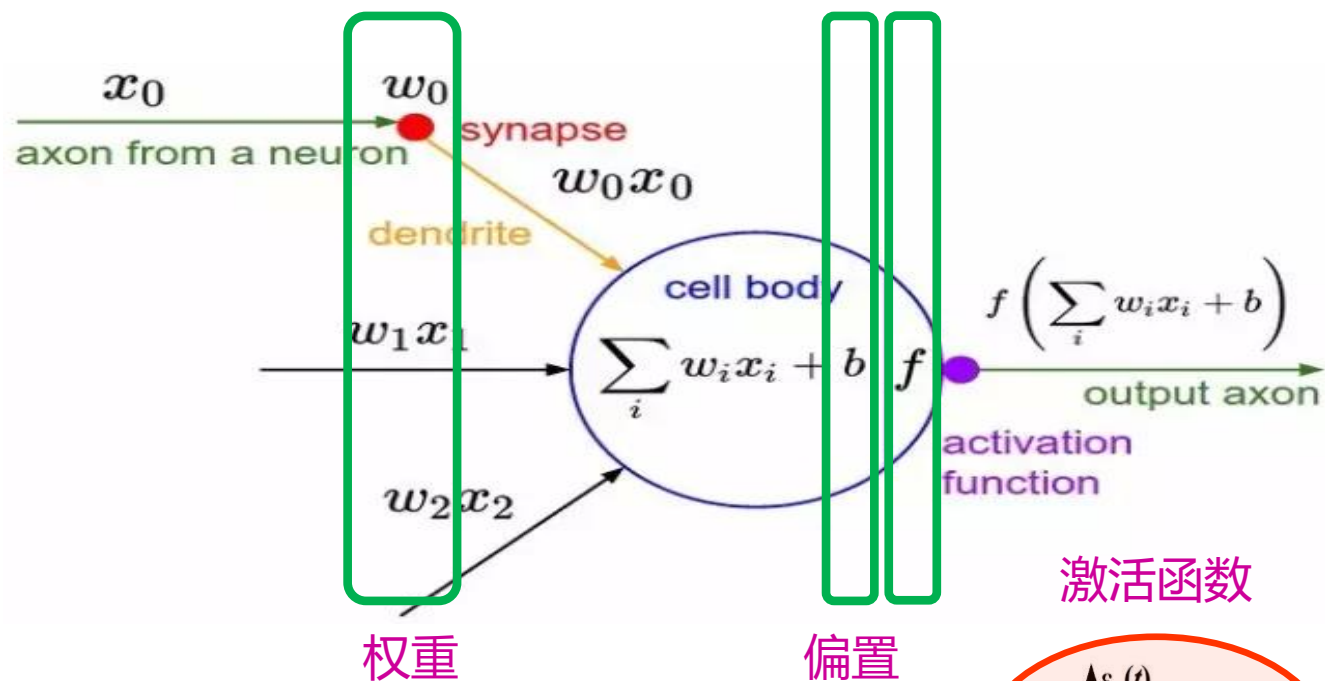
生物神经元的启示

- ① 每个神经元都是一个**多输入单输出**的信号处理单元
- ② 神经元具有**阈值特性**



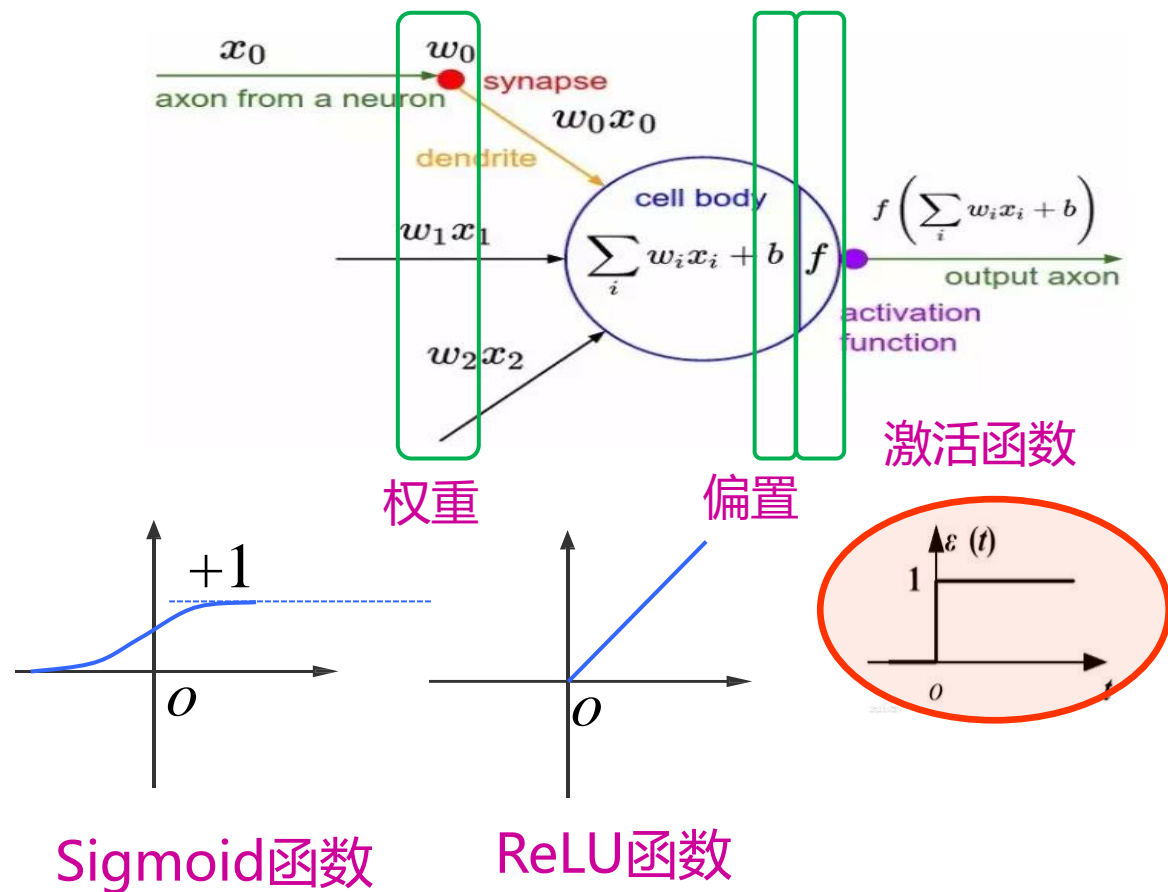
神经元

人工神经元



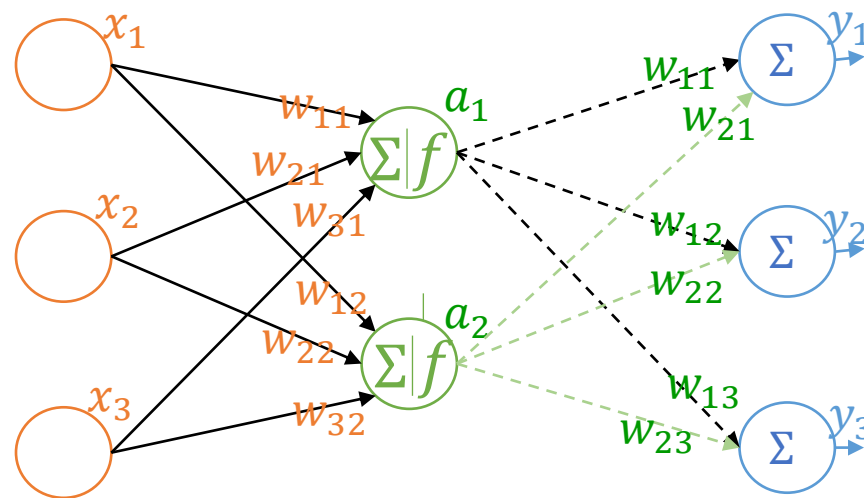
神经元

人工神经元的数学模型



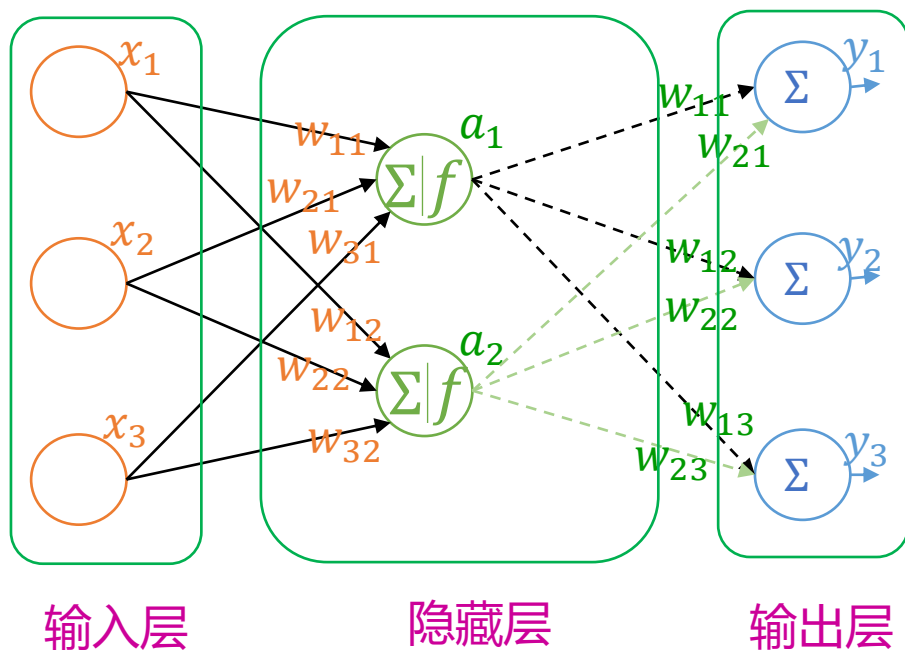
Problem : 如何连接?

- ◆ 总结: 多对多的关系
- ◆ Input: 多个上层的神经元
- ◆ Output: 多个下层神经元



神经元的连接

多层神经网络



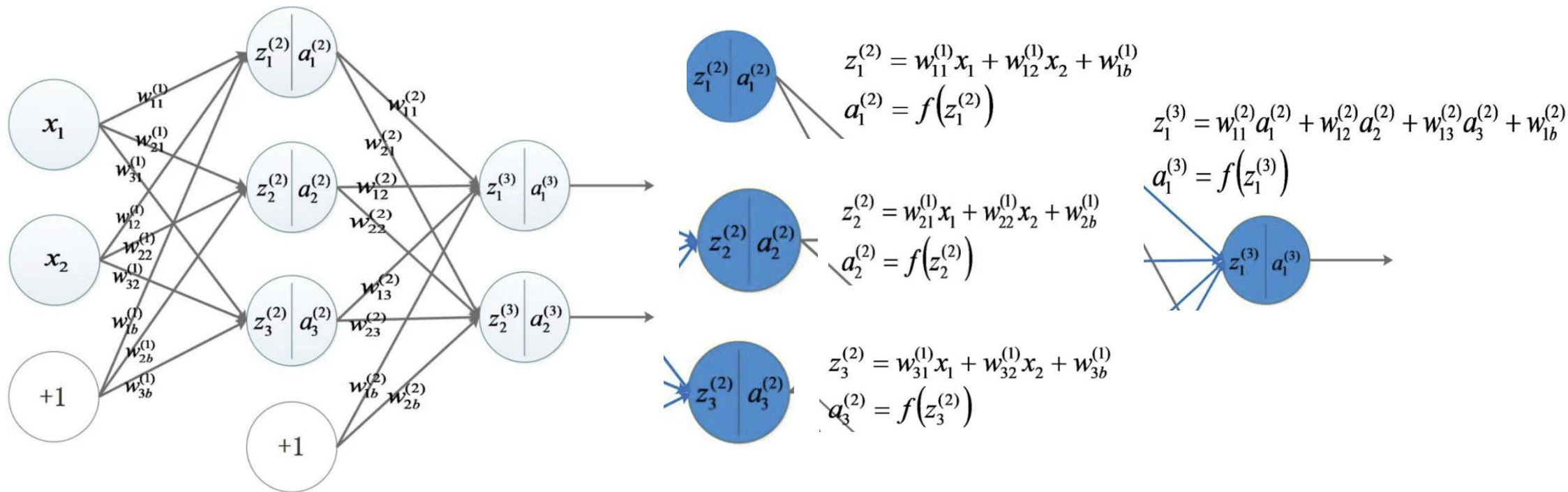
Problem : 如何连接?

- ◆ 输入层: 接收输入信号的层
- ◆ 输出层: 产生输出信号的层
- ◆ 隐含层: 同其名: 不直接与外部环境打交道
 - ✓ 隐含层的层数可从零到若干层

神经网络的训练

网络的训练是一个不断迭代的过程，在每一次的迭代中，从训练数据集中获得的新数据都会被用来修改结点的权重系数

正向传播： 信号通过隐藏层(hidden layer)进行传播



神经网络的训练

反向传播：误差反向传播

所有训练数据（共N组）的总体代价为

$$E_{total} = \frac{1}{N} \sum_{i=1}^N E(i) = \frac{1}{2} \left((y_1 - a_1^{(3)})^2 + (y_2 - a_2^{(3)})^2 \right)$$

$$\frac{\partial E}{\partial w_{11}^{(3)}} = -(y_1 - a_1^{(3)}) f'(z_1^{(3)}) a_1^{(2)}$$

调整权重和偏置，使得总体代价变小：

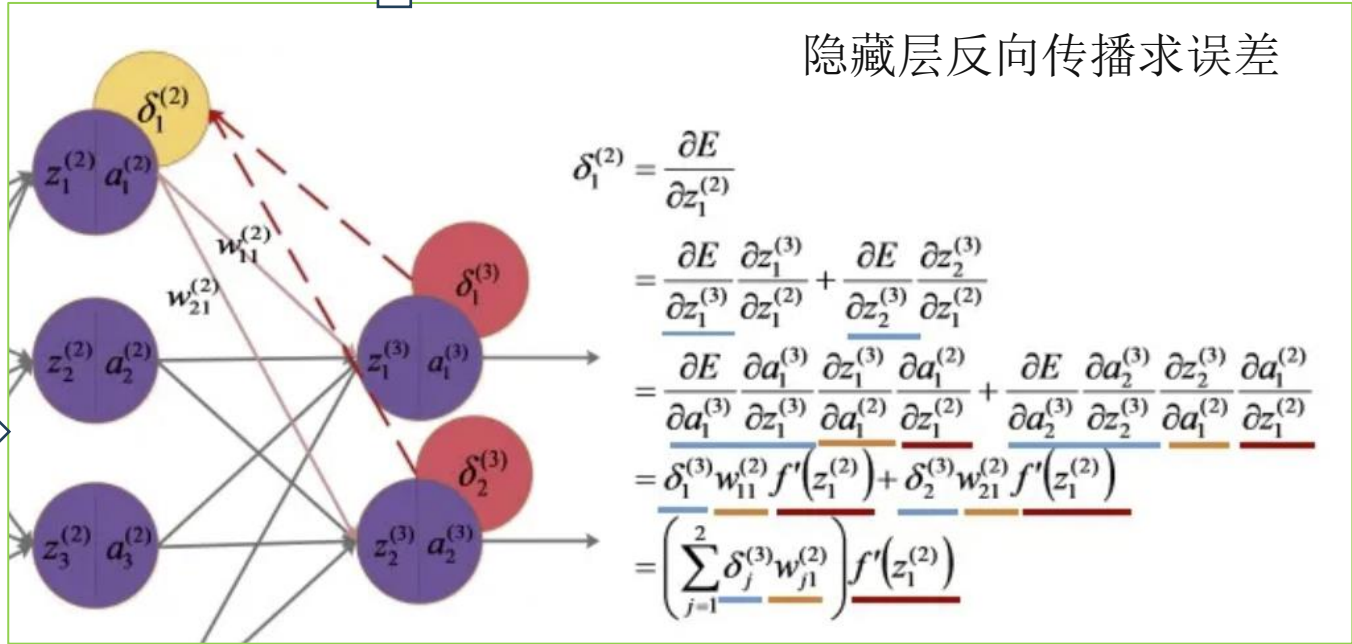
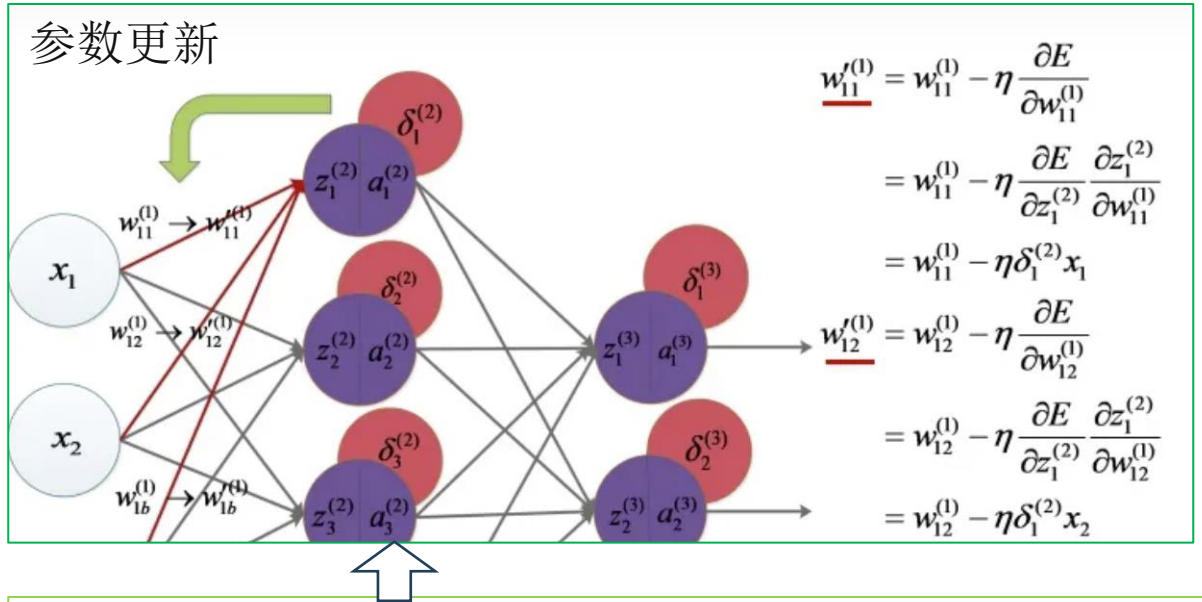
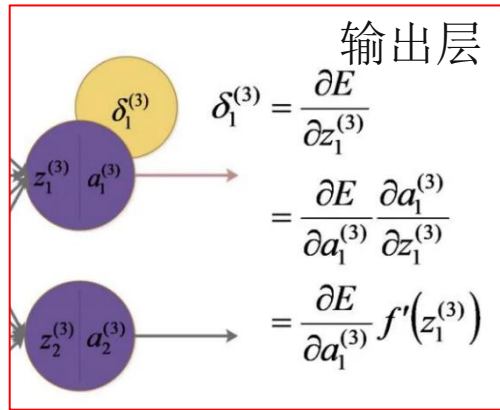
$$w_{new} = w_{old} - \eta \nabla E$$

$$w_{ji}'^{(l)} = w_{ji}^{(l)} - \eta \frac{\partial E}{\partial w_{ji}^{(l)}}$$

$$= w_{ji}^{(l)} - \eta \frac{\partial E}{\partial z_j^{(l+1)}} \frac{\partial z_j^{(l+1)}}{\partial w_{ji}^{(l)}}$$

$$= w_{ji}^{(l)} - \eta a_i^{(l)} \frac{\partial E}{\partial z_j^{(l+1)}}$$

$$= w_{ji}^{(l)} - \eta a_i^{(l)} \delta_j^{(l+1)}$$



AI的发展趋势

★ 第一次高潮

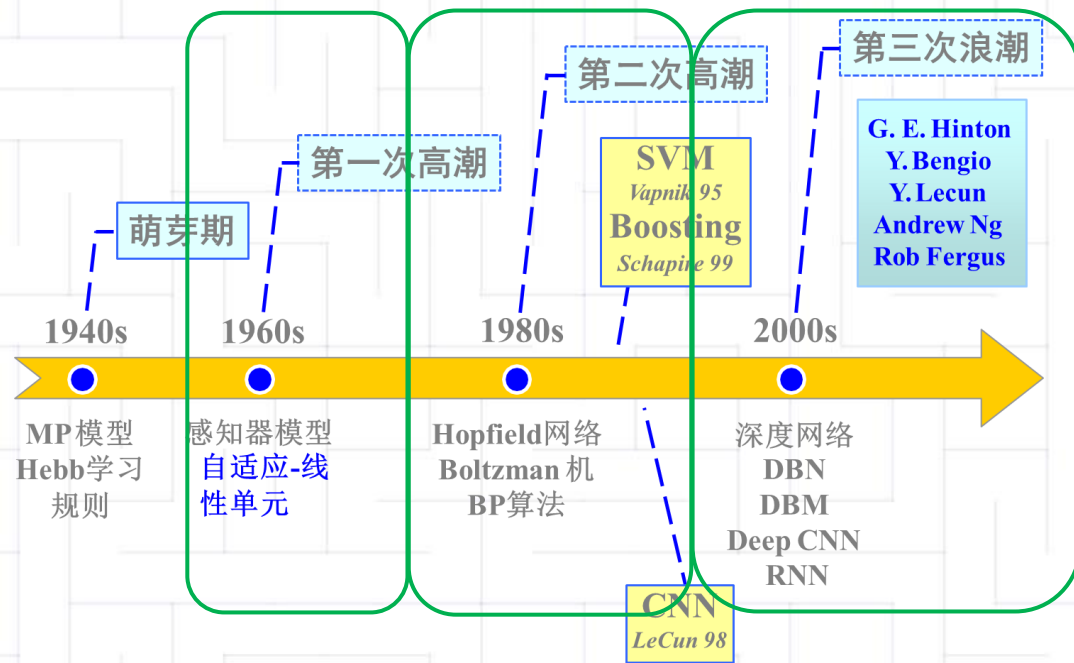
- 1962年，Frank Rosenblatt 提出感知器模型
- 1969年，M. Minsky 等人指出感知器不能解决高阶谓词问题和异或问题，将感知器拉下神坛

★ 第二次高潮

- 人工智能对自动制导车的失败，而利用神经网络有可能解决这个问题，导致了人工神经网络的第二次高潮。
- 1986年，Rumelhart等提出多层网络的学习算法—反向传播算法

★ 第三次高潮

- 2006年，深度学习被提出，开始使用更深的网络模型。
- 2012年，深度学习算法在语言和视觉识别上实现突破。
- 2020年，DeepMind的AlphaFold2达到实验水平的蛋白质结构预测，几乎颠覆了结构生物学。
- 2023年，以ChatGPT为代表的生成式大语言模型展示出卓越的学习能力和任务处理能力。



深度学习典型的网络架构

深度学习的三驾马车：**数据、算法和算力**

算法架构：

❖ **全连接神经网络 (Full Connection, FC)**

❖ **卷积神经网络 (Convolutional Neural Network, CNN)**

❖ **循环神经网络 (Recurrent Neural Network, RNN)**

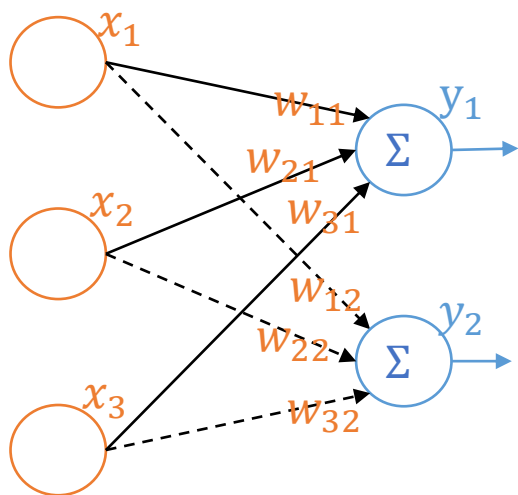
❖ **图神经网络 (Graph Neural Network, GNN)**

❖ **自注意网络 (Transformer)**



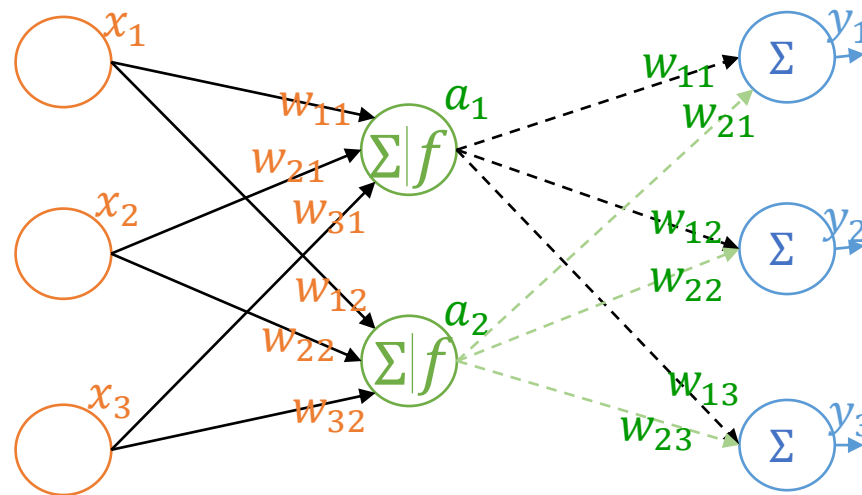
典型网络架构

全连接神经网络 (Full Connection, FC)



单层全连接网络

◆ 前馈/全连接神经网络：每层每个节点均和上一层的所有节点相连



多层全连接网络

全连接/前馈神经网络是能解决很多问题，但是本身存在一个问题，对于某些数据而言，一是参数量巨大，二是不能充分应用数据的某些特性

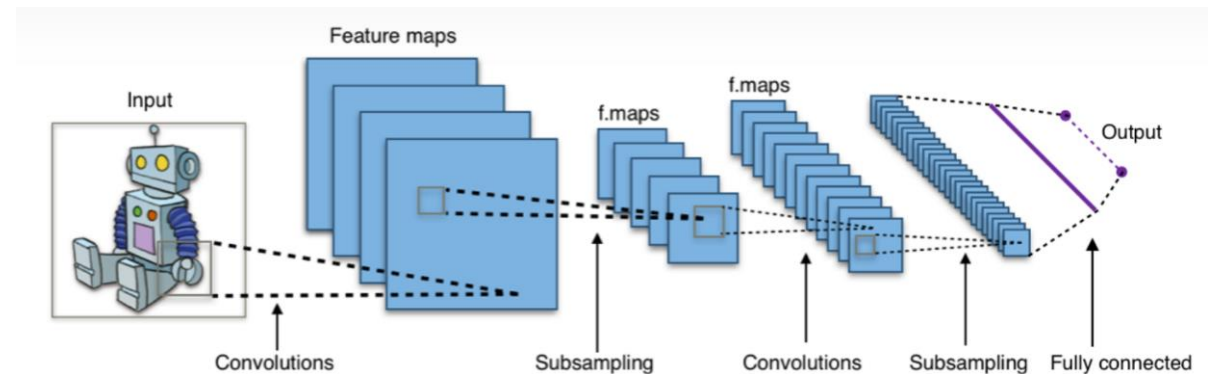
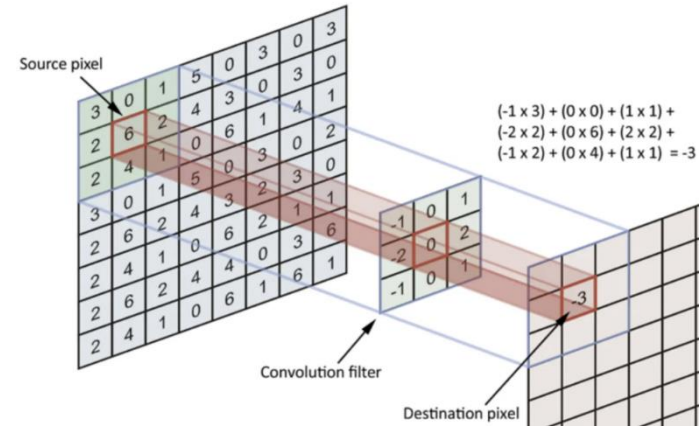
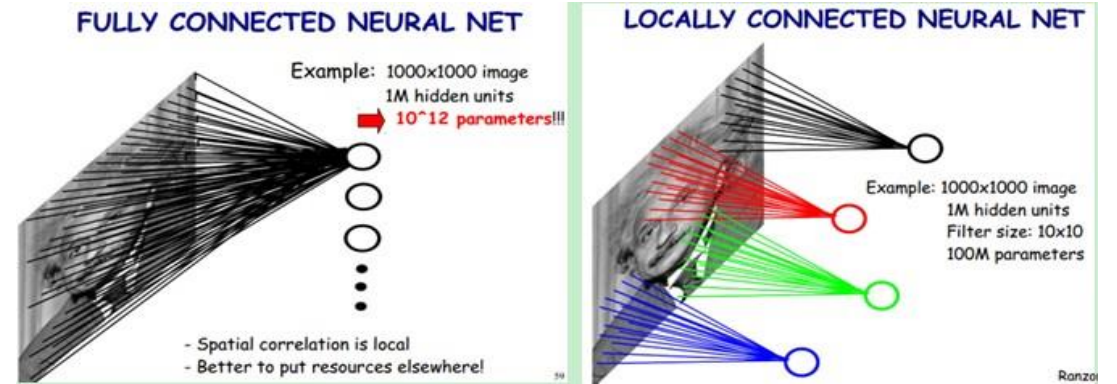
典型网络架构

卷积神经网络 (CNN)

- ◆ 引入特性：局部连接 (图像局部上下文)
- ◆ 输入层：不是向量，而是一个三维数组 (图像)
- ◆ 卷积层：对三维数组及其权重的计算方式。卷积核 (参数) 在通过逐一滑动窗口计算而得
- ◆ 池化/采样层：直接抽样选取极小局部的某一元素作为下一层的元素
- ◆ 黑盒 → 特征抽取的转变

更多经典CNN网络:

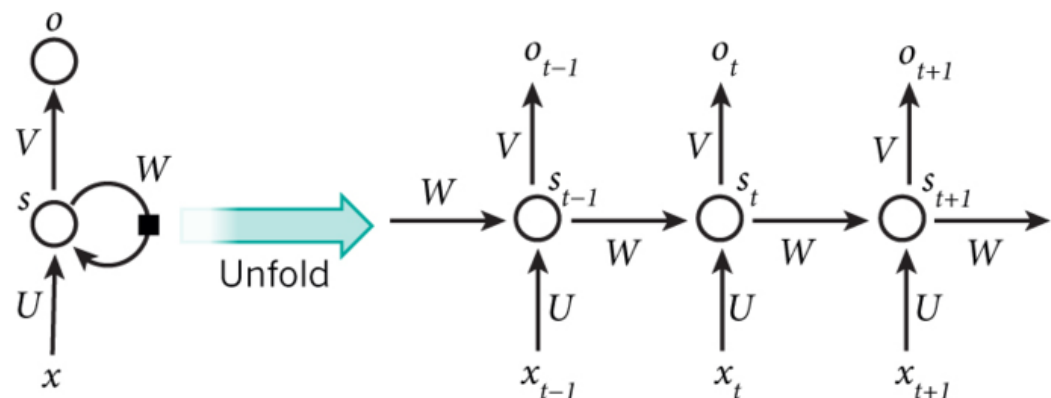
LeNet、AlexNet、VGG、GoogLeNet、ResNet



典型网络架构

循环神经网络 (Recurrent Neural Network, RNN)

- ◆ **引入特性:** 时序, 处理序列数据 (文本上下文)
- ◆ **对序列数据的处理:** 每次输入序列中的一个单元, 然后保存每一个隐层神经元的计算结果, 留给下一个输入时该神经元进行使用
- ◆ **历史信息:** 保存的隐层单元计算结果, 含有上一次输入的信息



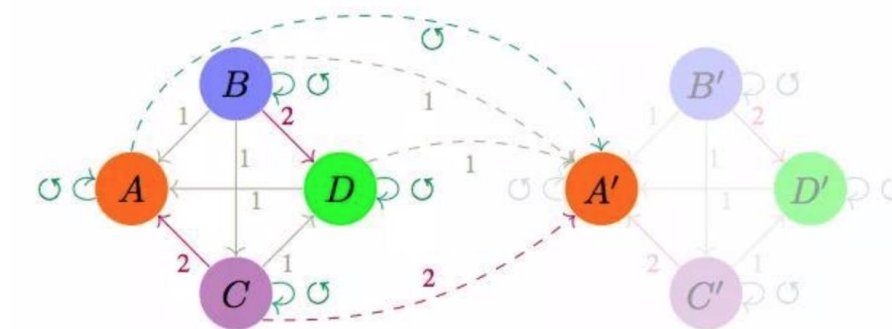
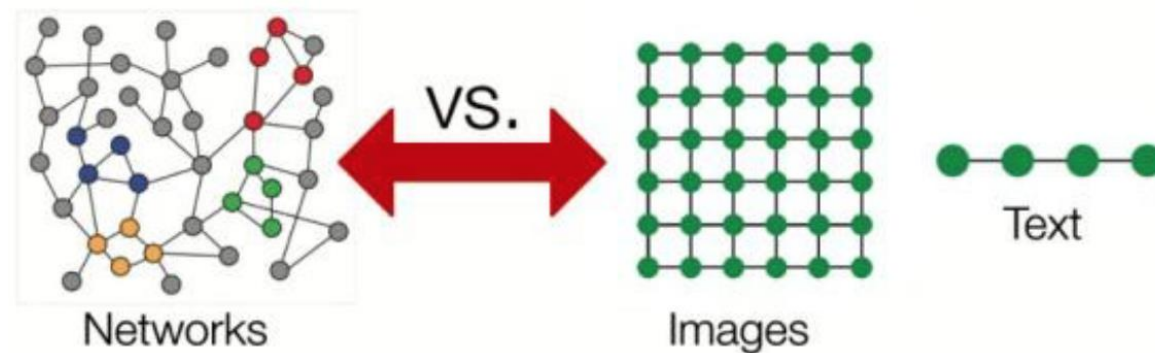
更多经典RNN网络:

LSTM、GRU、Bi-RNN、Seq2Seq

典型网络架构

图神经网络 (Graph Neural Network, GNN)

- ◆ 引入特性：任意尺寸和复杂的拓扑逻辑结构
- ◆ 思想：节点及其近邻构成了一个计算图 (Graph)
- ◆ 构成： $G(V, E)$ 表示图， V 是所有节点的集合， E 是所有边的集合。节点包含自身的特征，边包含节点与其近邻节点的关系。
- ◆ 图卷积：图论中用顶点和表建立起相应关系的拓扑图



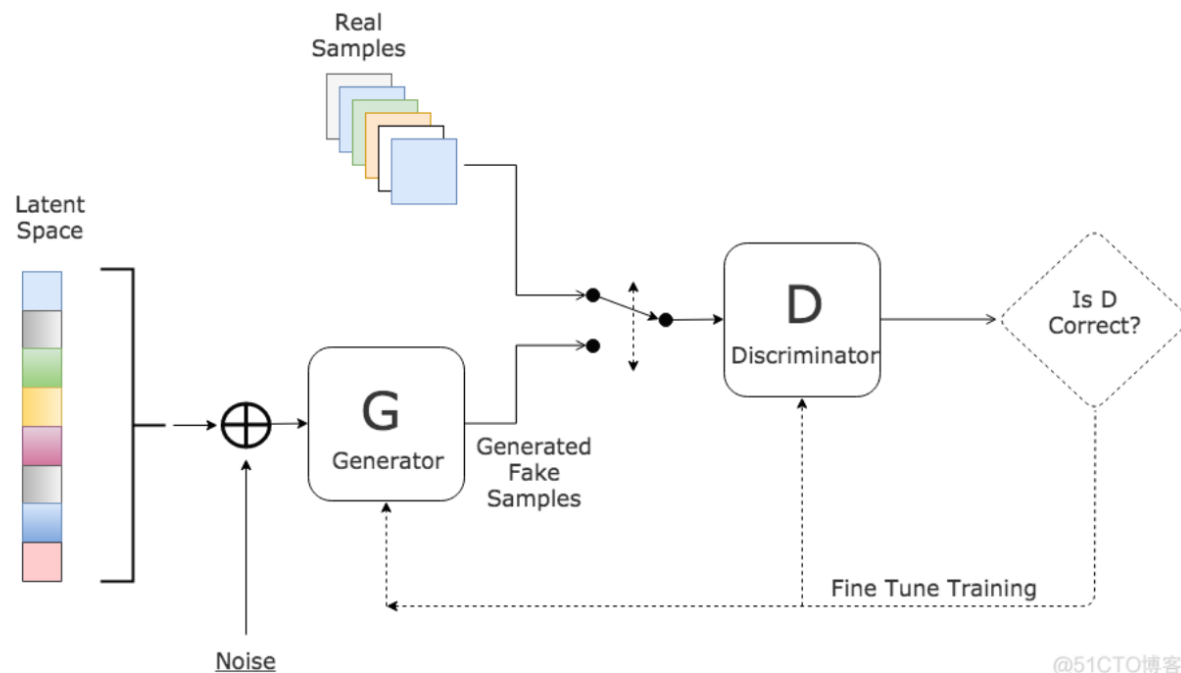
更多经典GNN网络：

GCN、NN4G、GraphSAGE、GIN、3WL-GNN

典型网络架构

对抗神经网络 (Generative Adversarial Network, GAN)

- ◆ 引入特性：无监督学习
- ◆ 核心思想：生成对抗网络由一个生成网络G与一个判别网络D组成，通过两个神经网络相互博弈的方式进行学习
- ◆ 作用：生成式任务，如生成、重建、超分辨率等



@51CTO博客

更多经典GAN网络：

GAN, DCGGAN, WGAN, LSGAN, Style-GAN



典型网络架构

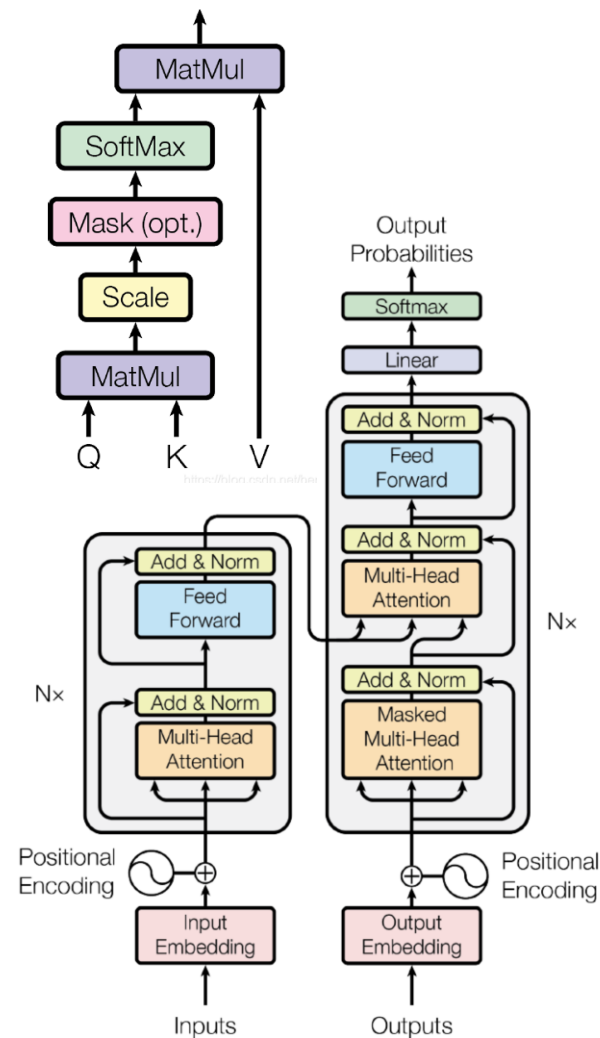
$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V$$

自注意力网络 (Transformer)

- ◆ 引入特性：自注意和多头注意力机制
- ◆ QKV：三个矩阵Query, Key, Value
- ◆ QK内积：Q和K中每个向量的相关度
- ◆ Attn：应用相关度对V进行缩放，即注意对V中每个向量的不同重要程度
- ◆ 自注意：Q和K相同时，为自注意

更多经典Transformer网络：

Transformer, Swin-Transformer, Vision Transformer



图像技术

- 图像分类
- 目标检测
- 分割算法
- 人脸与人体识别
- 文字识别
- ...

语音技术

- 声纹识别
- 实时语音识别
(内容、情绪)
- 语音合成
- 智能问答
- ...

自然语言处理

- OCR转写
- 情感倾向分析
- 文章标签、分类
- 机器翻译
- 文本审核
- ...

科学研究

- 蛋白质结构预测
- 加速分子动力学模拟
- 新冠病毒药物筛选
- 提升气象预报准确率
- 助力育种技术换代升级
- 加速偏微分方程求解
- 加速矩阵计算
- 喷注事例重建
- 加速ALTAS模拟
- 识别强子 τ 衰变
- 实验装置控制
- 快速CT重建
- ...

AI的优势

- 缓解“维数灾难”，复杂度 $\Omega(n^3) \rightarrow \Omega(n)$

AI的缺陷

- 可解释性问题
- 小样本问题



02

AI在高能物理领域的应用



机器学习在高能物理领域的应用

- 高能物理:

- 以越来越高的精度探索标准模型 (SM)
- 寻找超出标准模型的物理相关的新粒子
- 在庞大的背景中识别罕见信号
- 对海量数据进行统计分析

- 深度学习:

- 需要大量训练数据样本
- 通用函数模拟器, 替代复杂的算法, 性能和成本
- 并行化是高能物理计算的趋势, 深度学习算法天然地支持并行计算
- 训练完成后, 深度学习的推理计算比传统的算法要简单得多。一次训练多次调用, 可以节约计算资源

- 应用领域

- 模拟
- 实时分析和触发
- 事例重建、鉴别和刻度
- 端到端深度学习
- 理论应用
- 监测探测器

[HEP ML Living Review \(iml-wg.github.io\)](https://iml-wg.github.io)



Simulation

- 蒙特卡洛模拟工具，例如**GEANT4**

- 精确，可解释，包含所有细节
- 计算复杂度：需要进行大量的随机数生成和事件采样，这导致计算复杂度较高。对于复杂的物理过程或大规模的模拟，需要投入大量的计算资源和时间

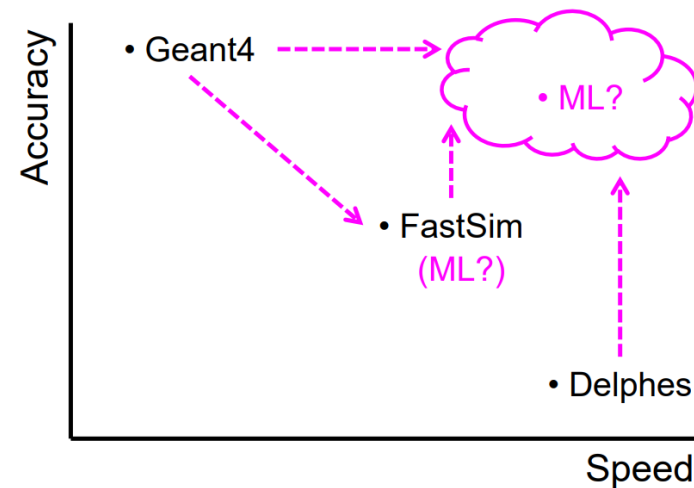
- 快速模拟：

- 通过简化参数化或粒子shower表进行近似

- 高保真度的快速生成模型：

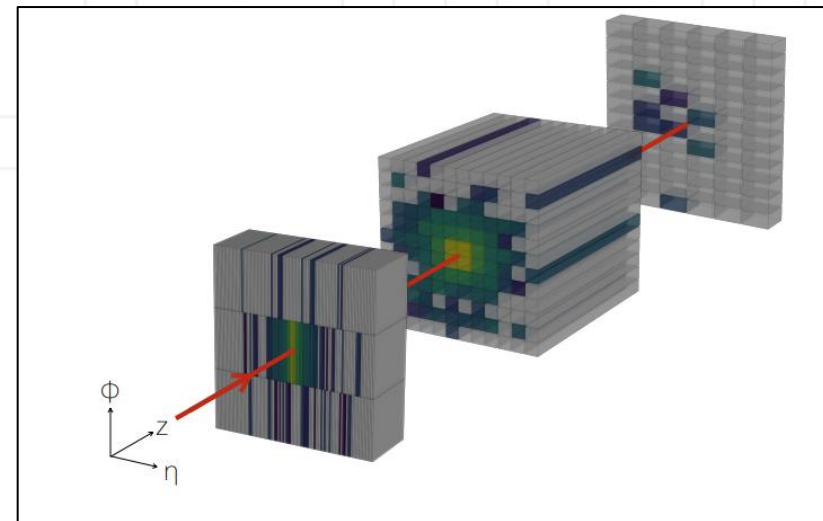
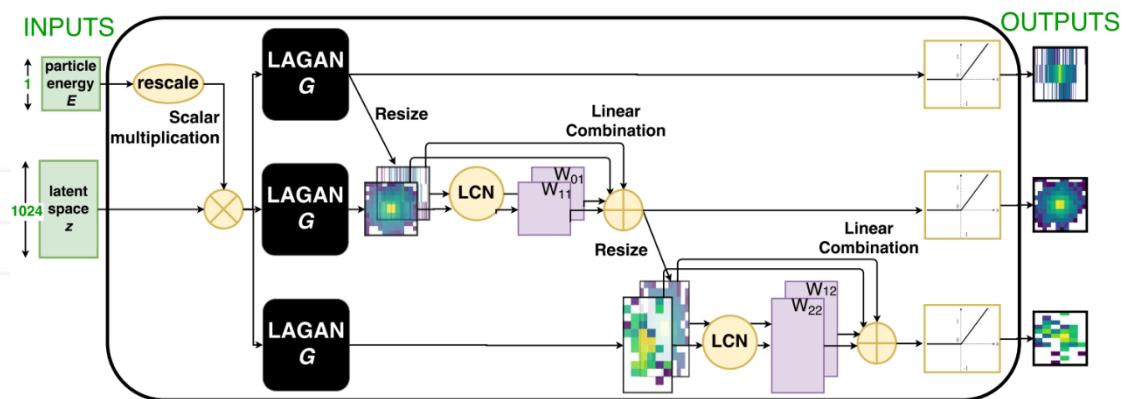
- CaloGAN:

使用生成对抗网络在多层电磁量能器中模拟三维高能粒子shower
(Phys.Rev.D 97 (2018) 1, 014021)



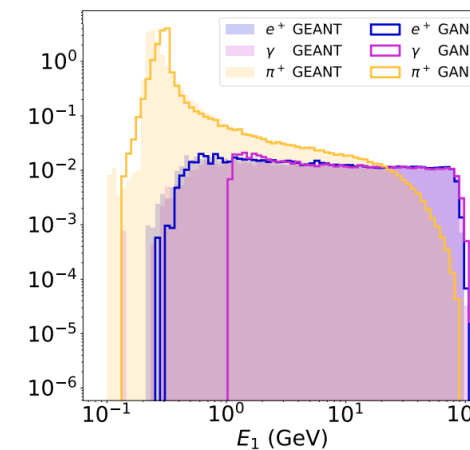
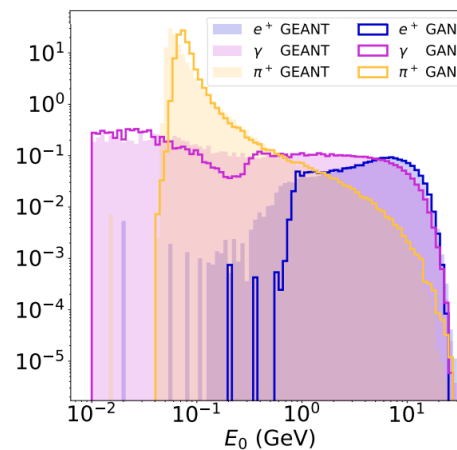
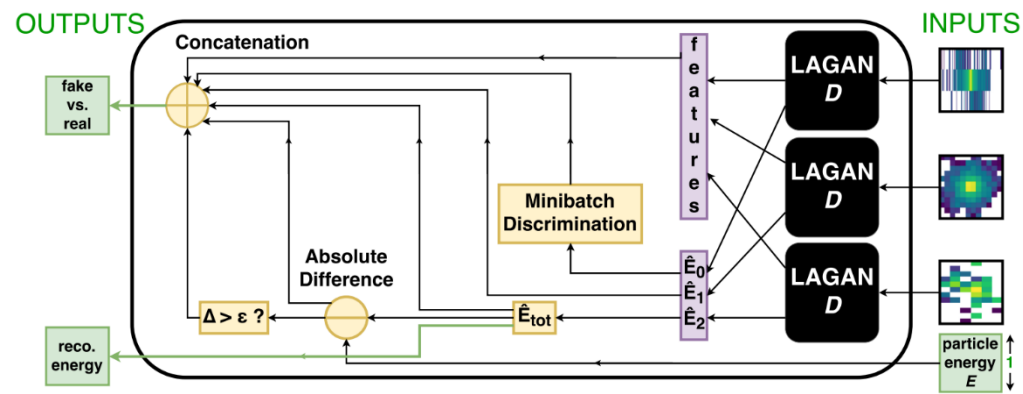
Simulation

Composite Generator



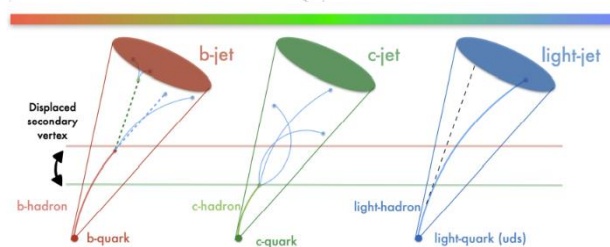
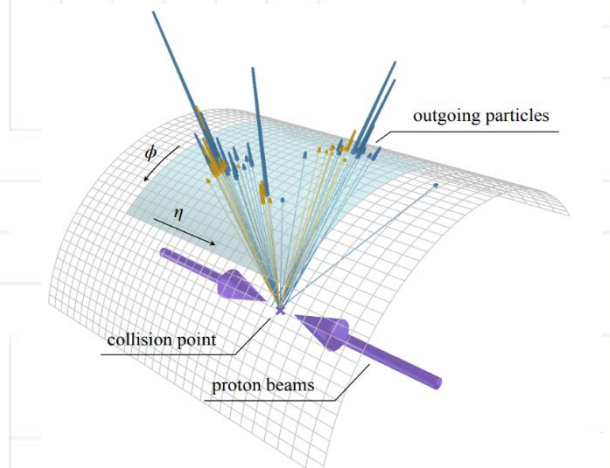
Three-dimensional representation of a 10 GeV e^+ incident perpendicular to the center of the detector

Composite Discriminator



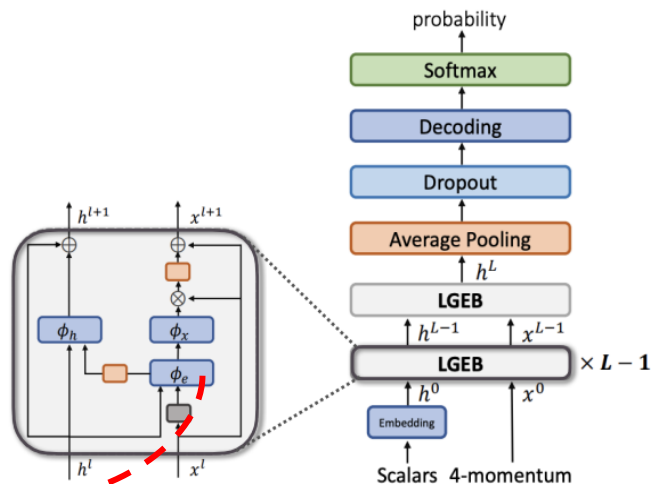
Identification (Classification)

- Jet tagging



	Accuracy	# params	FLOPs
PFN	0.772	86.1 k	4.62 M
P-CNN	0.809	354 k	15.5 M
ParticleNet	0.844	370 k	540 M
LorentzNet	0.855	233 k	2.01 G
ParT	0.861	2.14 M	340 M
ParT (plain)	0.849	2.13 M	260 M

LorentzNet



MLP Sum Pooling Minkowski Norm & Inner Product

Lorentz Group Equivariant Block (LGEB)

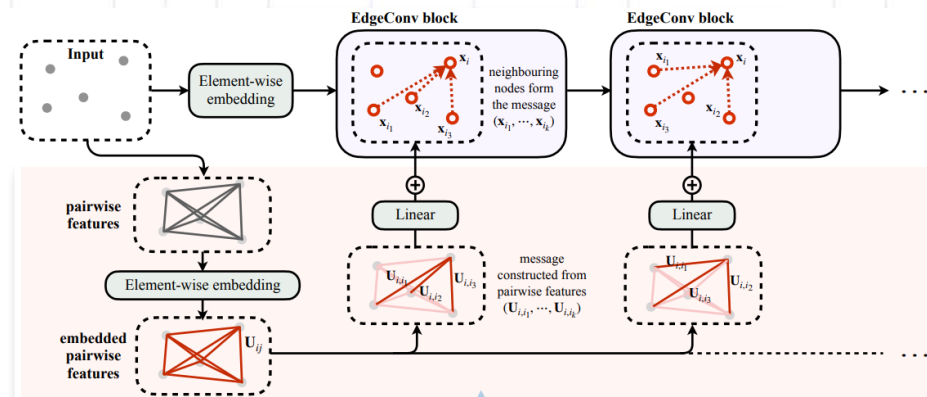
LorentzNet

- Construct edge features from
- (1) scalars from two nodes
 - (2) Lorentz inner product from two vectors

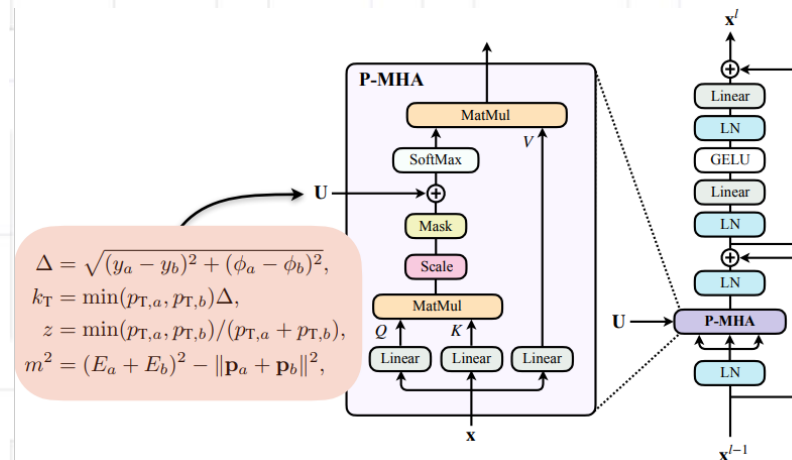
$$m_{ij}^l = \phi_e \left(h_i^l, h_j^l, \psi(\|x_i^l - x_j^l\|^2), \psi(\langle x_i^l, x_j^l \rangle) \right)$$

hence it's a Lorentz scalar

ParticleNet



ParT



$$\Delta = \sqrt{(y_a - y_b)^2 + (\phi_a - \phi_b)^2},$$

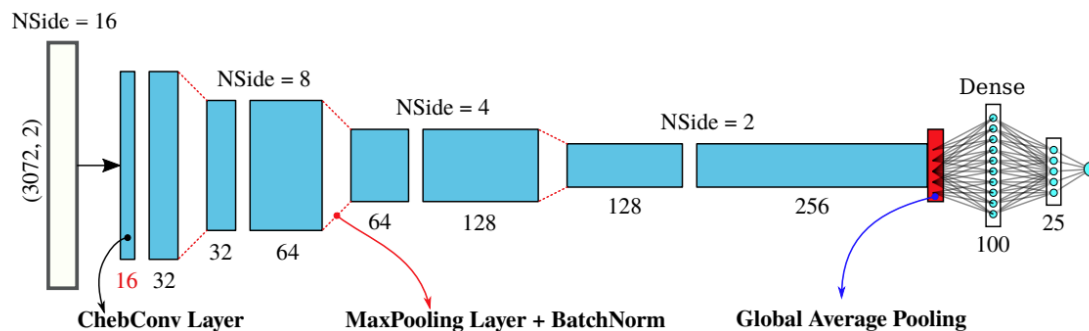
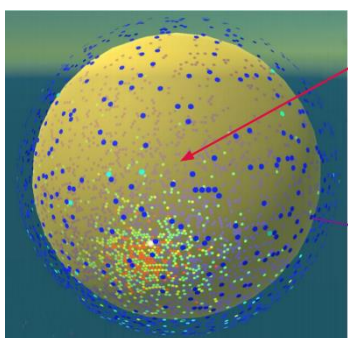
$$k_T = \min(p_{T,a}, p_{T,b}) \Delta,$$

$$z = \min(p_{T,a}, p_{T,b}) / (p_{T,a} + p_{T,b}),$$

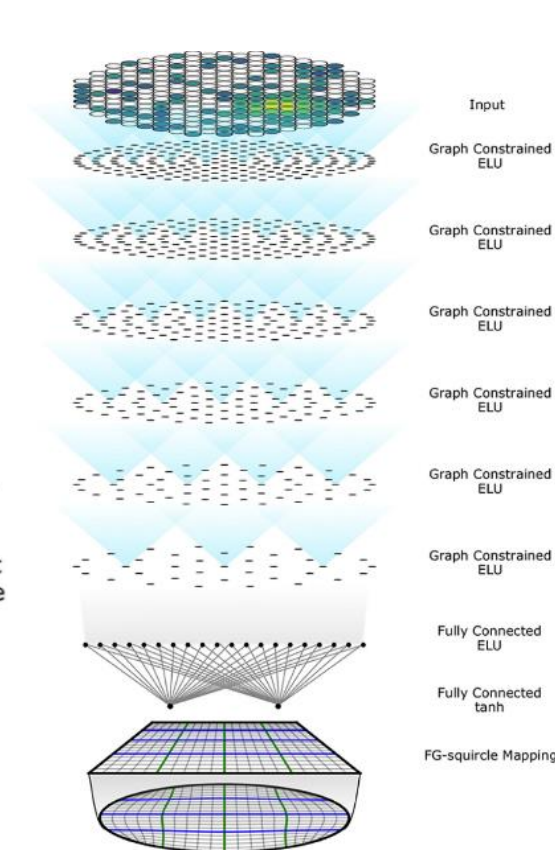
$$m^2 = (E_a + E_b)^2 - \|\mathbf{p}_a + \mathbf{p}_b\|^2,$$

Reconstruction

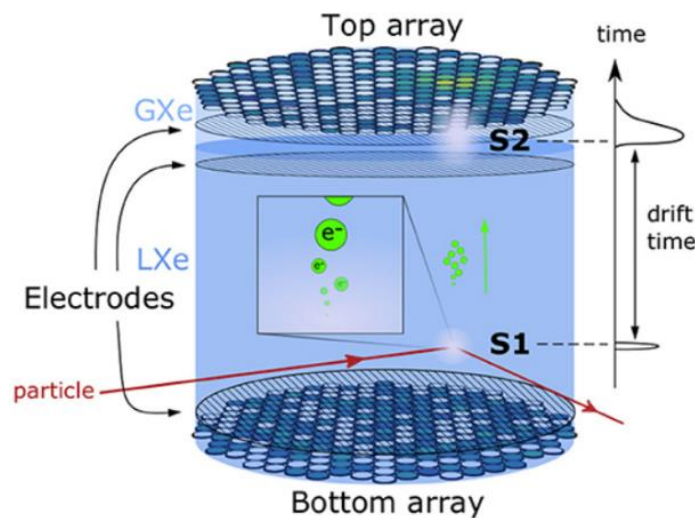
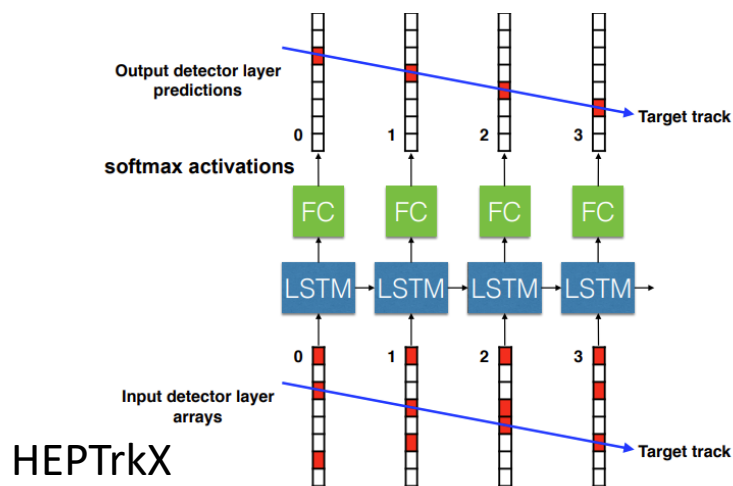
- 用机器学习方法重建JUNO实验中的顶点和能量



Astroparticle Experiments



- Hit with LSTMs in 2D

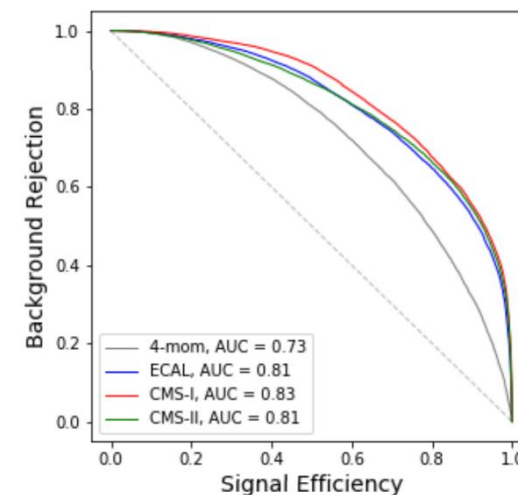
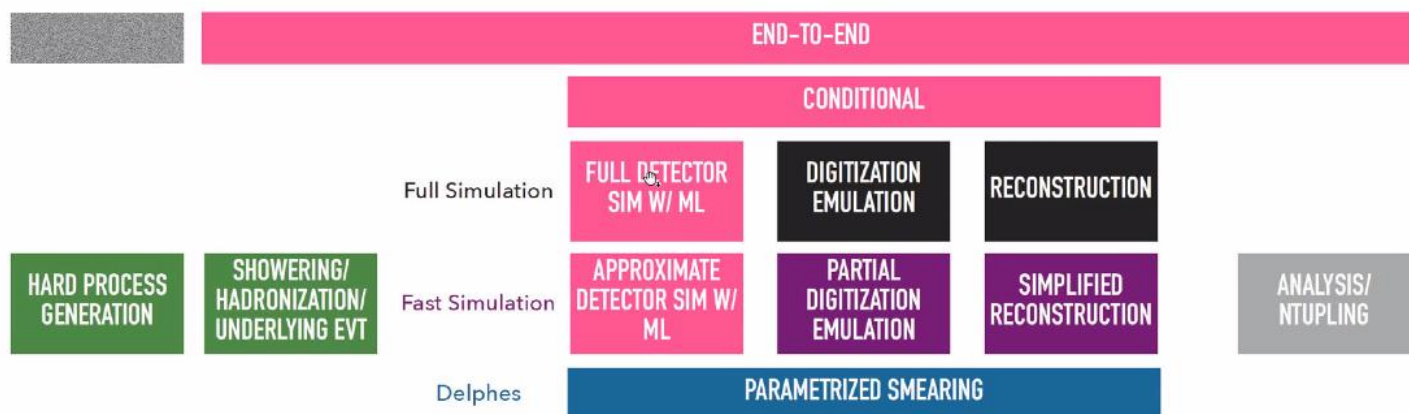


End-To-End Deep Learning

- End-to-End Physics Event Classification with CMS Open Data (Comput.Softw.Big Sci. 4 (2020) 1, 6)

$$gg \rightarrow H \rightarrow \gamma\gamma$$

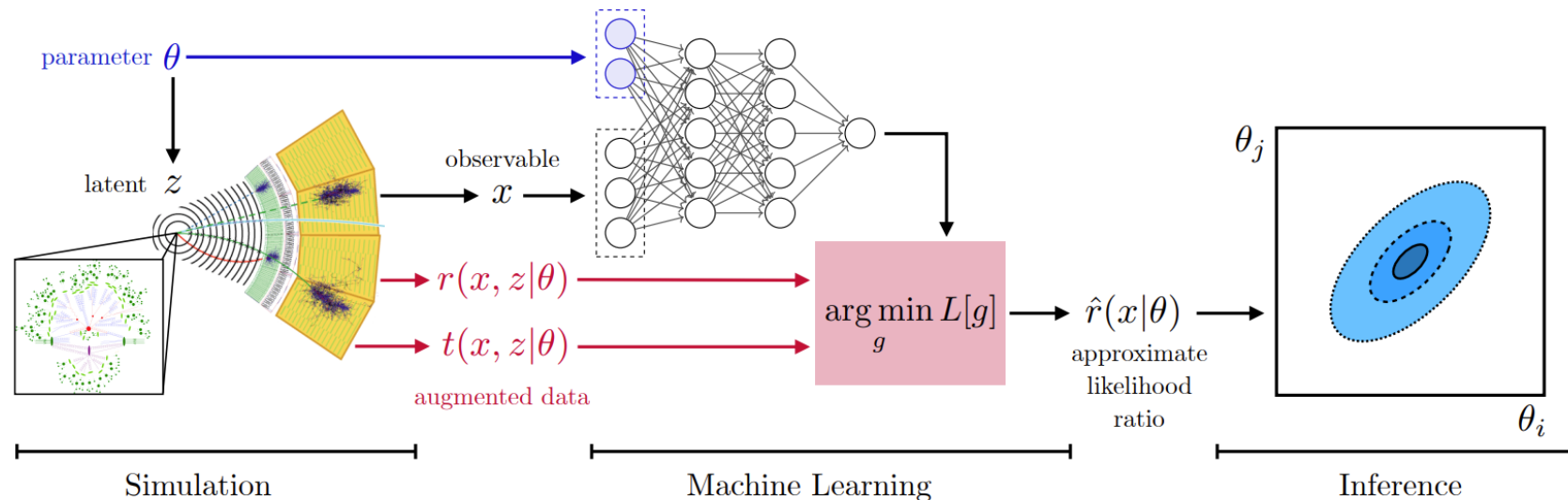
Residual Net-type networks (ResNet-15)



- 直接利用low-level simulated detector 数据来区分信号和背景过程
- High-level Particle Reconstruction:
 - probable particle identity, kinematics, and shower shape features

Theory Applications

- 用机器学习约束有效场论
 - 学习 likelihood ratios



- Feynman 圈图积分的数值计算
 - sampling of the phase space in integrals (arXiv: 2209.01091)
- Machine learning spectral functions in lattice QCD (arXiv: 2110.13521)
- Sequence-to-sequence transformer model to compute a key element of the cross section calculation

03

总结



总结

- I. 人工智能
 - 人工智能-机器学习-深度学习
 - 优势：
 - 能够处理大量的数据和复杂的模式，以及在决策和预测方面具有高精度和效率各个领域的应用
 - 关键技术与典型网络架构
- II. 人工智能在高能物理中的应用
 - 高能物理数据模拟，重建，分析...
 - 加速科学发现和推动理论模型的发展
 - 高能物理的研究需要高度精确的模拟和预测，对AI算法的准确性和可解释性提出了更高要求
 - AI在高能物理实验中的应用还需要与传统的物理实验技术和方法进行结合和融合。





THE END

THANKS

