



中山大學 物理与天文学院
SUN YAT-SEN UNIVERSITY SCHOOL OF PHYSICS AND ASTRONOMY

LHASSO高能粒子能量数据 AI重构和分析

报告人：黄罗琳

小组成员：朱政鑫、戴鹏辉、杨舒云

指导教师 李霄栋

报告时间：2024年11月30日

内容提要



研究背景

高海拔宇宙线观测站 (LHAASO)
水切伦科夫探测器阵列 (WCDA)
极大似然估计 (Likelihood)



原理概述

深度学习 (DL) 与神经网络 (NN)
图卷积神经网络 (GNN)
能量数据AI重构与分析



结果分析

训练流程于效果
结果比对
能量不确定度评估



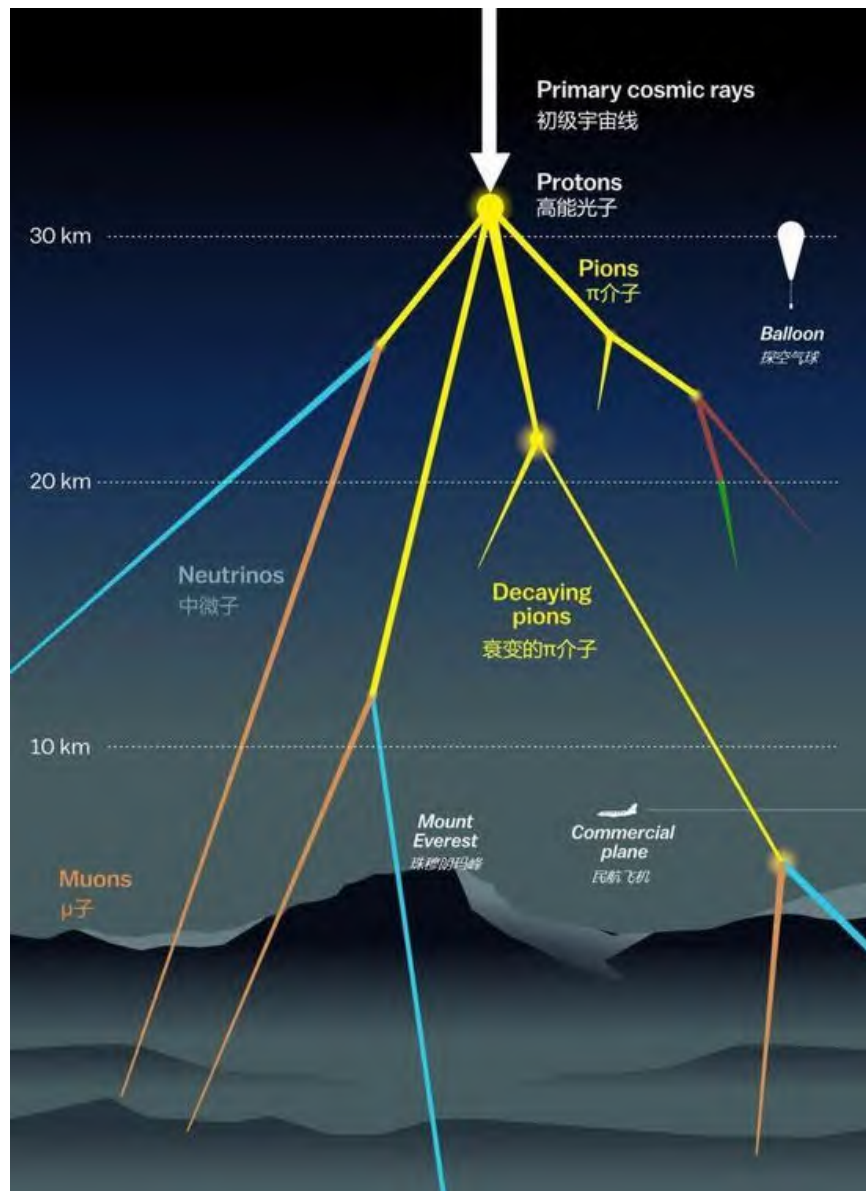
总结展望

GNN效果与AI技术回顾总结
基于新模型Transformer的实验构想
致谢

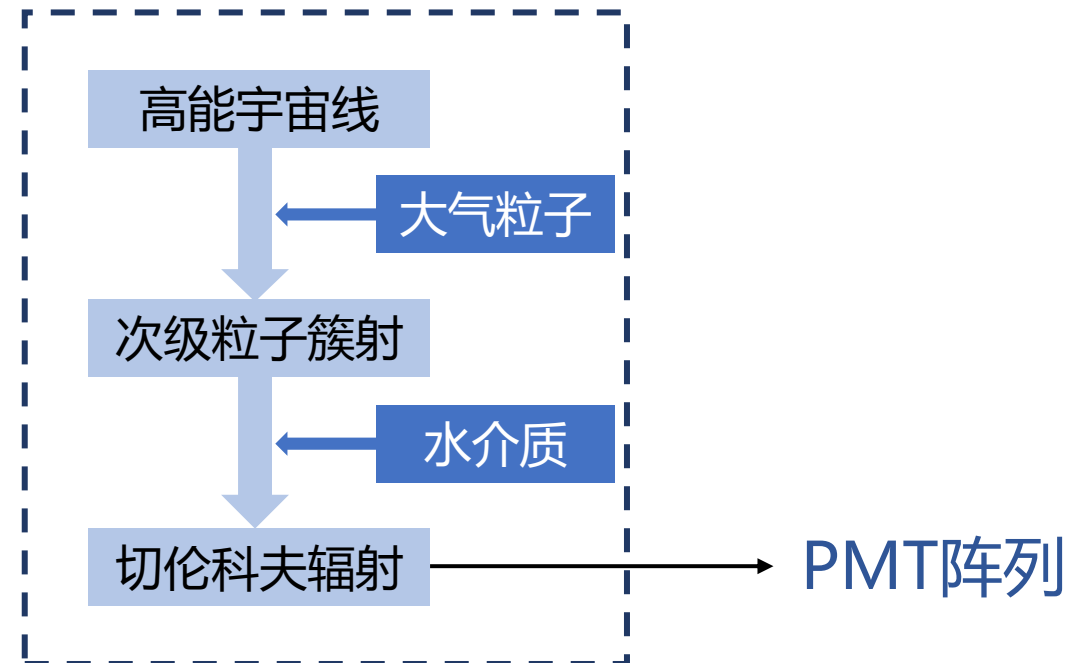
研究背景：高海拔宇宙线观测站（LHAASO）水切伦科夫探测器阵列（WCDA）



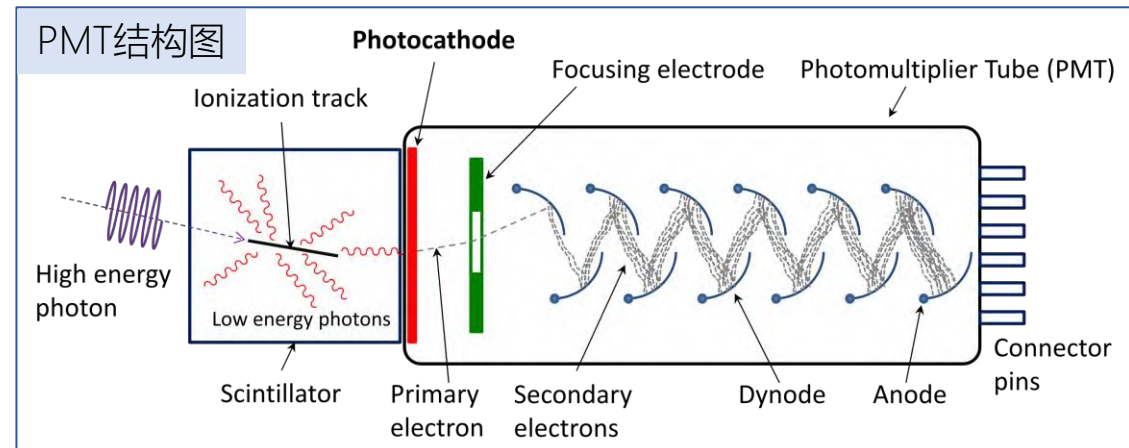
研究背景：高海拔宇宙线观测站（LHAASO）水切伦科夫探测器阵列（WCDA）



大气簇射观测



PMT结构图



研究背景：Likelihood算法

物理模型： 横向分布建模，采用“电磁级联的NKG函数”：

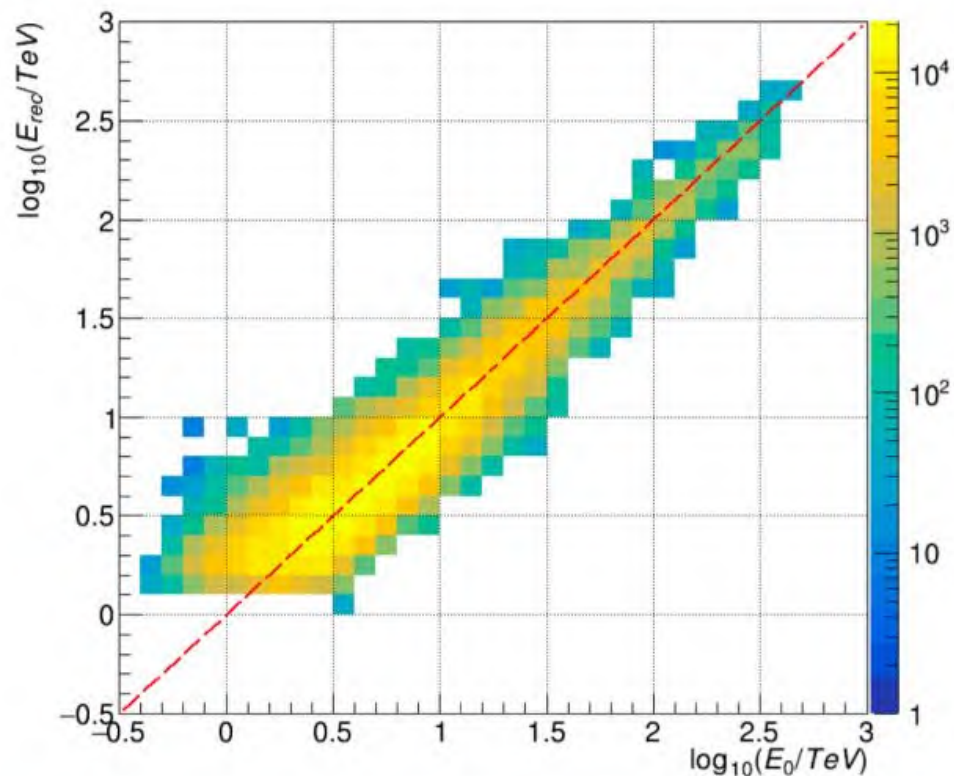
$$\rho_e(r, s, N_e) = \frac{N_e}{R_m^2} \frac{I(4.5 - s)}{2\pi\Gamma(s)\Gamma(4.5 - 2s)} \left(\frac{r}{R_m}\right)^{s-2} \left(1 + \frac{r}{R_m}\right)^{s-4.5}$$

数据分析： 采用“无箱似然”：

$$\mathcal{L} = \prod_{i \text{ have sample}} f_i \cdot \prod_{i \text{ in all}} \exp(-f_i)$$

$$\ln \mathcal{L} \sim \sum_{hit} n_e^{hit} \cdot \ln(N_e \rho^{hit})$$

研究背景：Likelihood算法



Likelihood

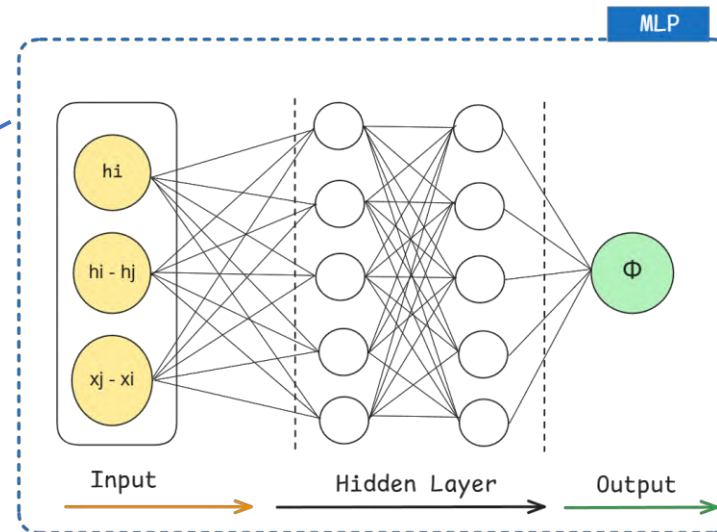
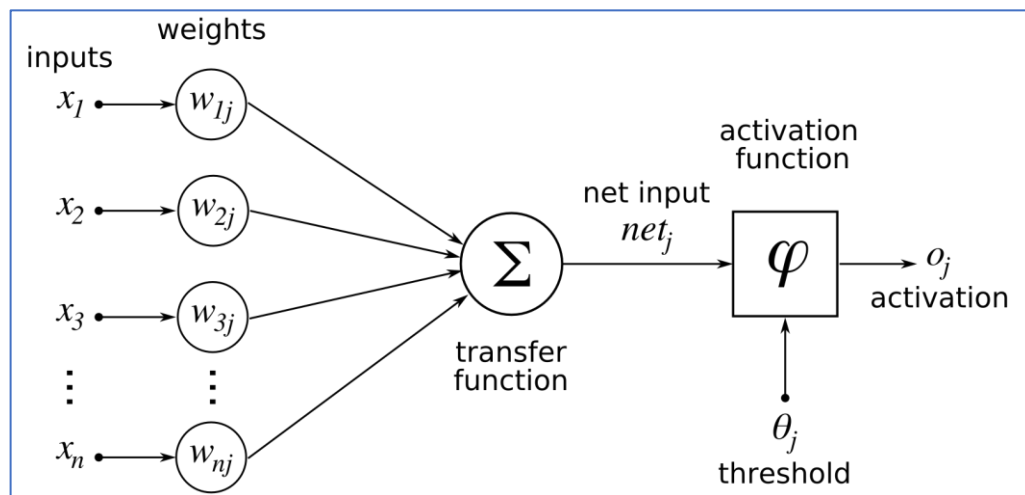
传统似然方法的缺陷：低能区域重建精度较低；简单近似准确性差；复杂建模计算成本太大且响应速度不够快。

我们的动机：我们希望改进WCDA的能量重建算法，考虑采用**GNN方法**克服传统方法的局限性，高效、准确地处理低能量事件，同时适应未来的探测器升级需求。

原理概述：图神经网络 (Graph Neural Network, GNN)

多层感知器 (Multilayer Perceptron, MLP)

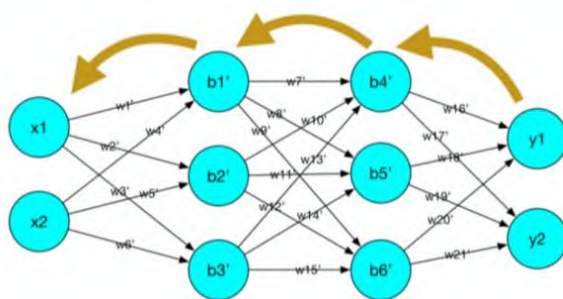
基本网络



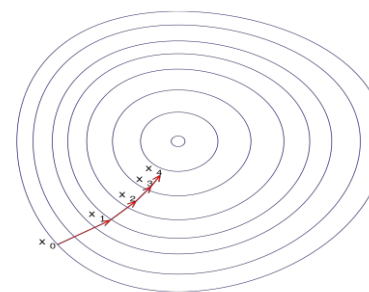
前向传播

$$\mathbf{F}(\mathbf{x}) = \begin{bmatrix} F_1(x_1, x_2, \dots, x_n) \\ F_2(x_1, x_2, \dots, x_n) \\ \vdots \\ F_n(x_1, x_2, \dots, x_n) \end{bmatrix}$$

反向传播
算法

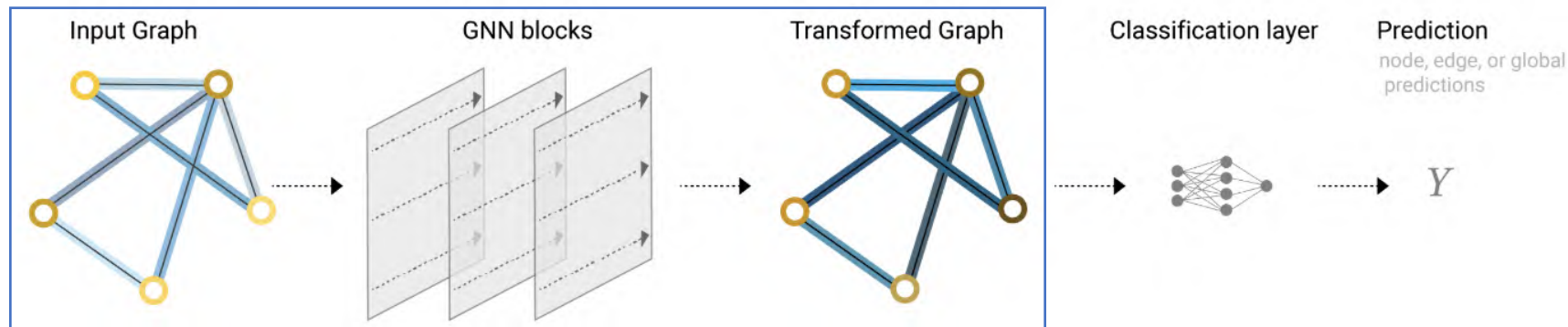


梯度下降

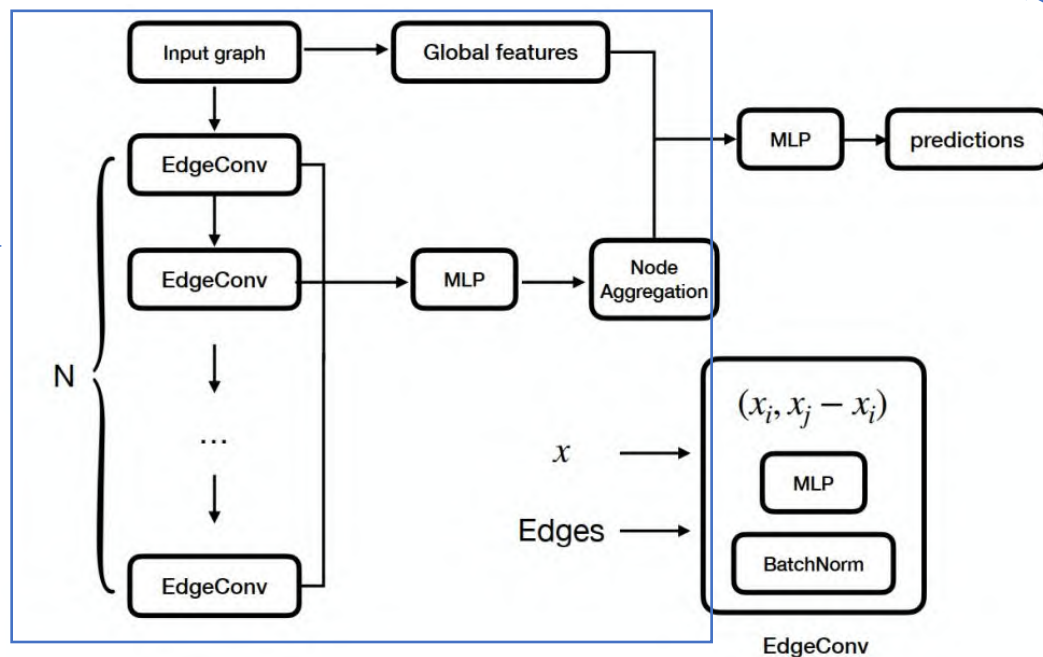


原理概述：图神经网络（Graph Neural Network，简称GNN）

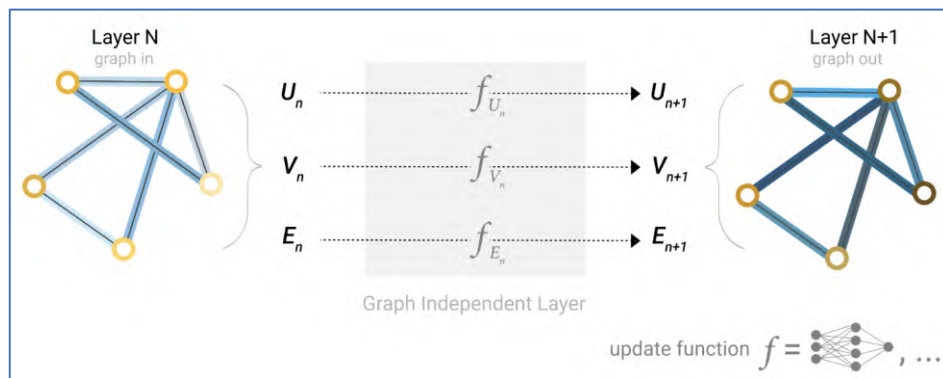
GNN的一般结构：



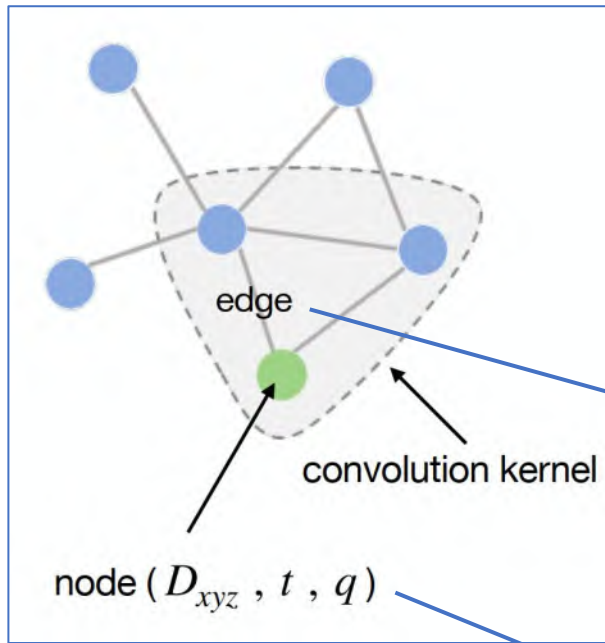
我们的模型：



基于上述基本网络（MLP架构）

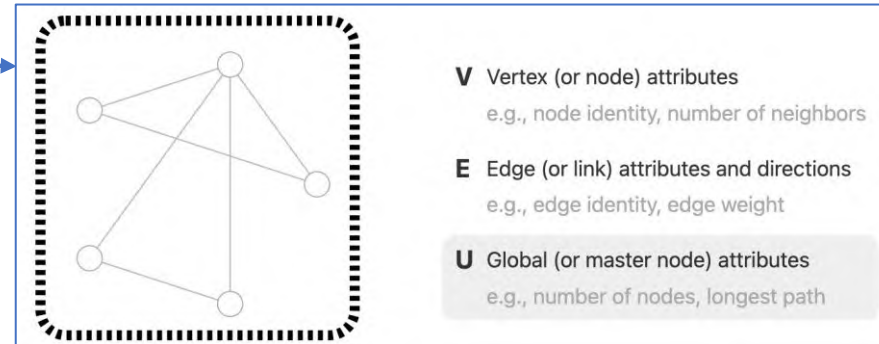
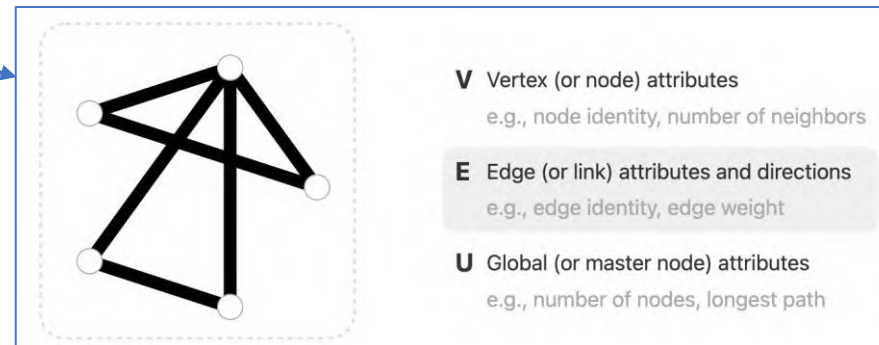


原理概述：模型搭建-输入图

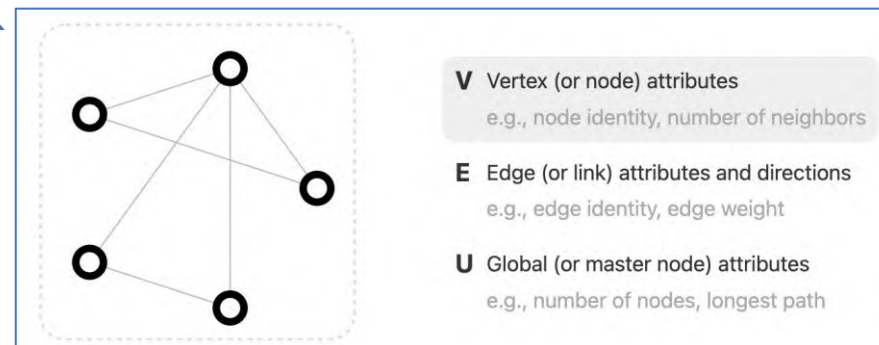


输入图

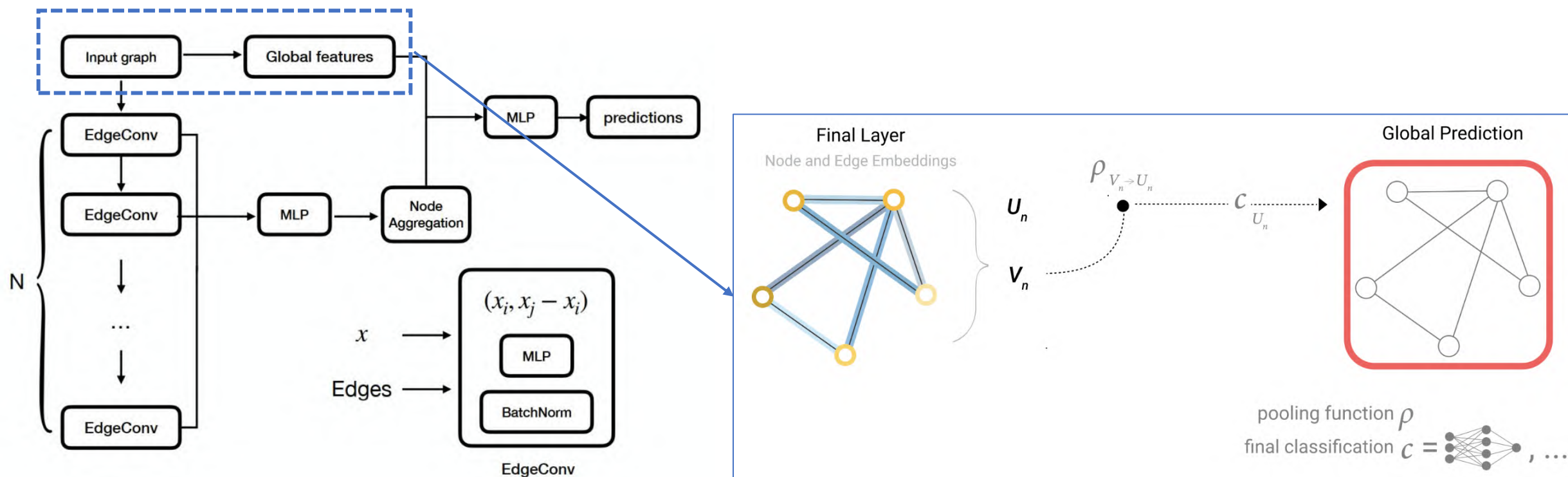
Feature	Description	Unit
D_{xyz}	Position of PMTs in wcd coordinates	m
t	Time of a pulse	ns
q	Charge of a pulse	P.E.

**全局特征：**全局聚合**边特征：**节点之间的物理关联(如空间相邻性或信号强度关联)用边 (edge) 表示。

$$|\mathbf{x}_i - \mathbf{x}_j|$$

**节点特征：**每个PMT (光电倍增管)可被看作图中的一个节点 (node)，每个节点携带的特征，包括PMT的位置(D_{xyz})、信号到达时间(t)和脉冲电荷量(q)。

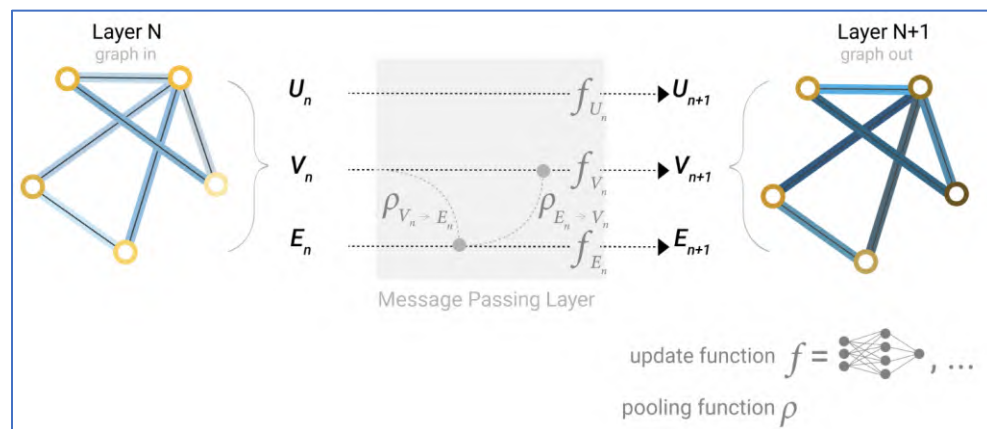
原理概述：模型搭建-全局聚合



全局聚合通过池化函数提取 5 个全局统计量，作为额外的事件信息，提供关于事件整体特性的全局信息，比如信号的紧密程度或时间分布。

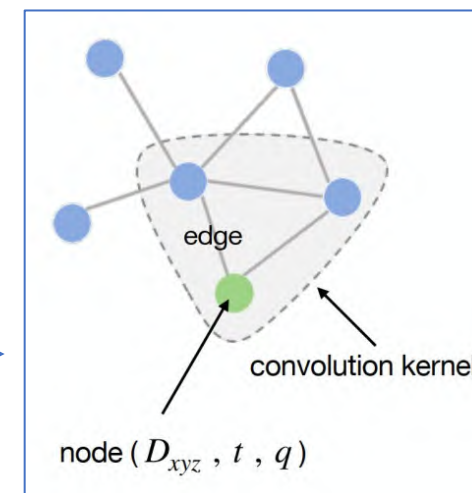
原理概述：模型搭建-EdgeConv

(1) 卷积操作：得到新一层的节点

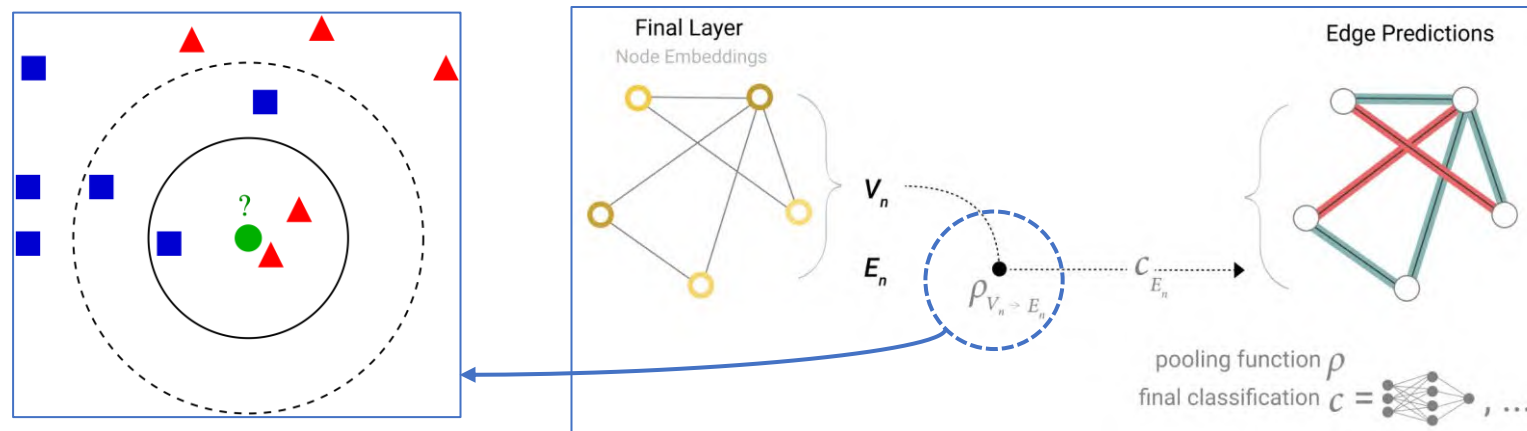


$$\mathbf{h}_i^{(k+1)} = \sum_{j \in \mathcal{N}(i)} \phi(\mathbf{h}_i^{(k)}, \mathbf{h}_j^{(k)}, \mathbf{x}_j - \mathbf{x}_i)$$

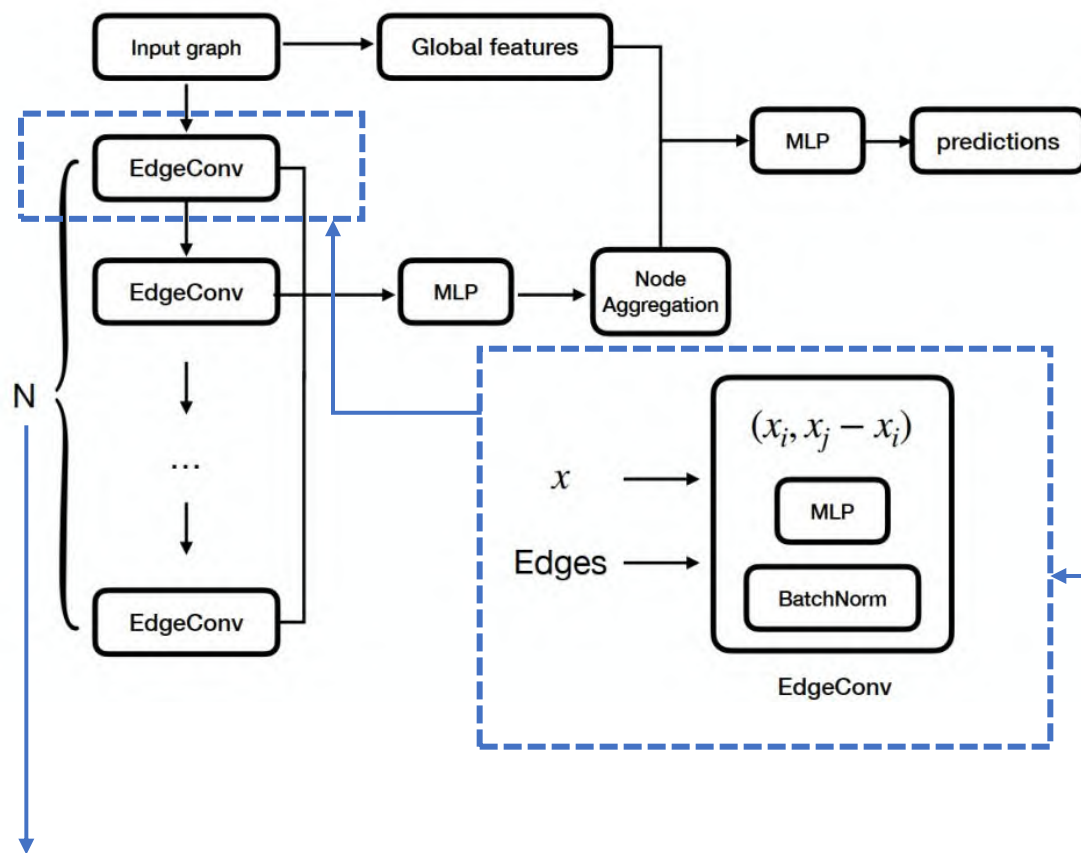
$$\phi(\mathbf{h}_i, \mathbf{h}_j, \mathbf{x}_j - \mathbf{x}_i) = \text{MLP}([\mathbf{h}_i, \mathbf{h}_j - \mathbf{h}_i, \mathbf{x}_j - \mathbf{x}_i])$$



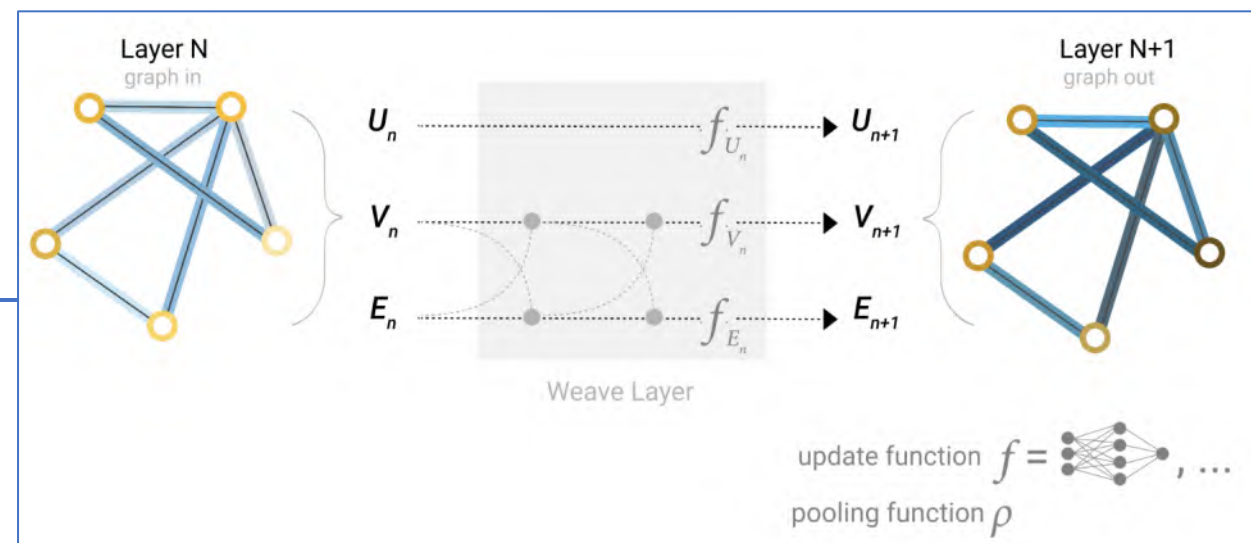
(2) KNN: 得到新一层的边，邻居的定义通过KNN计算，最初基于欧几里得距离，后续层在更高的抽象特征空间中重新定义邻居。



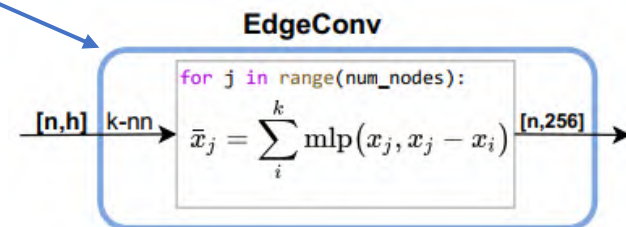
原理概述：模型搭建-EdgeConv



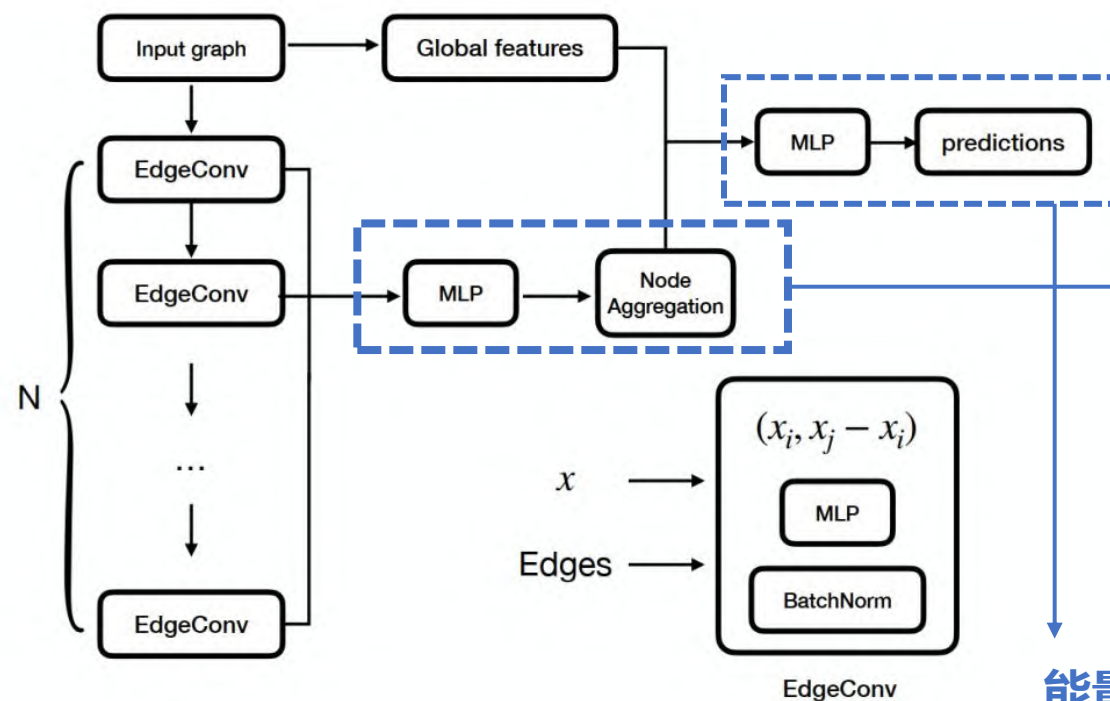
(3) EdgeConv: 将卷积得到的新节点和KNN得到的新边合起来，生成新一层的图，这体现了模型对图特征的学习。



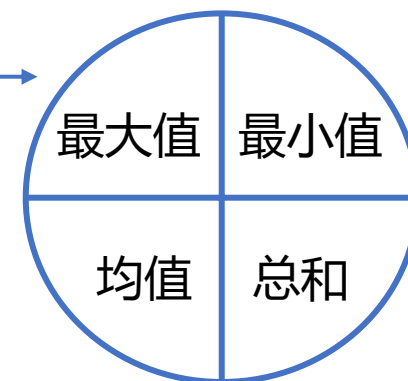
多层EdgeConv: 层与层之间允许图的结构逐步变化，使得模型能够学习节点的高层次关系。结合邻域的动态调整，模型可以自动发现对特定任务最相关的局部和全局特征。



原理概述：模型搭建-节点聚合与能量重建任务

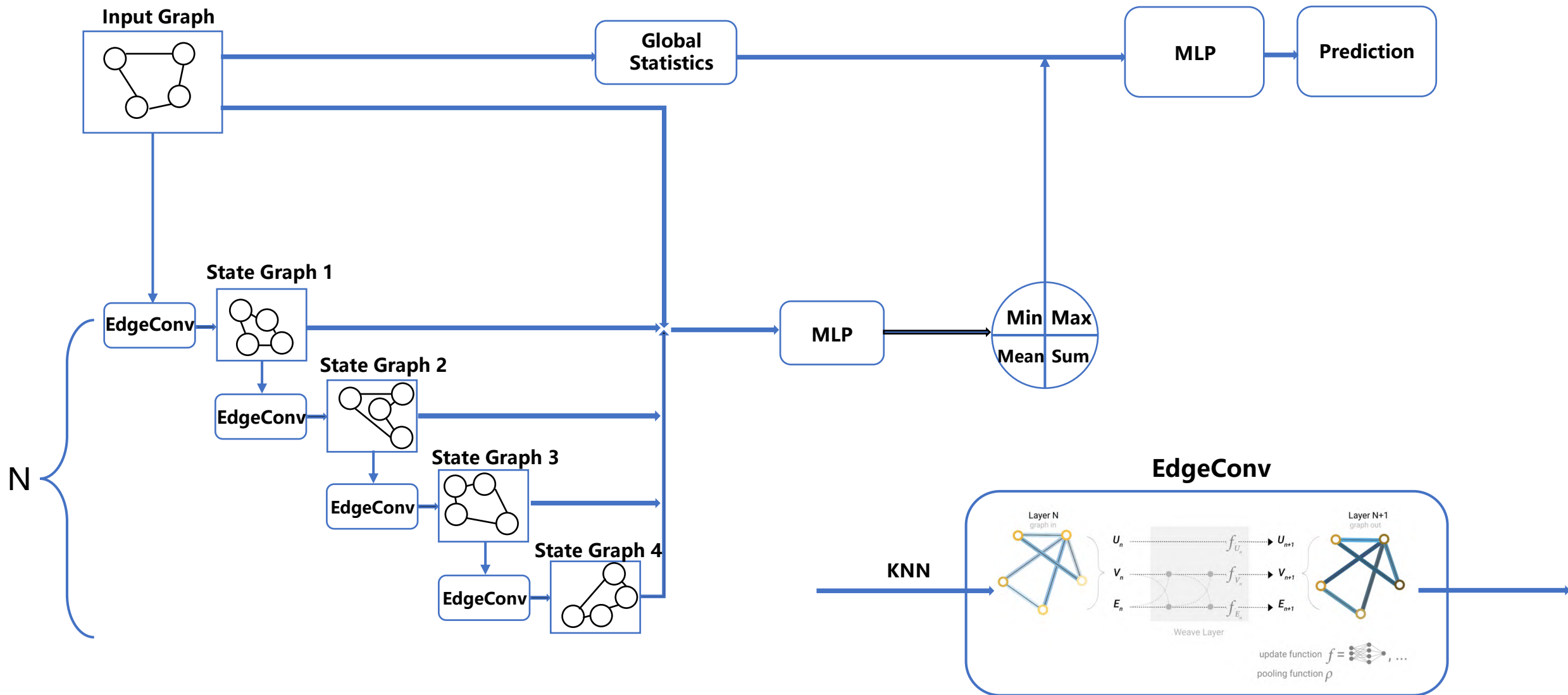


节点聚合：卷积处理完成后，节点特征被聚合为全局统计信息，表示整个事件的特征。



能量重建任务：聚合后的特征与全局统计量一起输入到一个全连接的多层感知器中，最终输出预测值（也就是重建结果）。

原理概述：模型搭建-完整模型



原理概述：训练模型

(1) 优化器和学习率：

(2) 数据集：

数据集	个数
训练集	103758
测试集	25940
验证集	19839

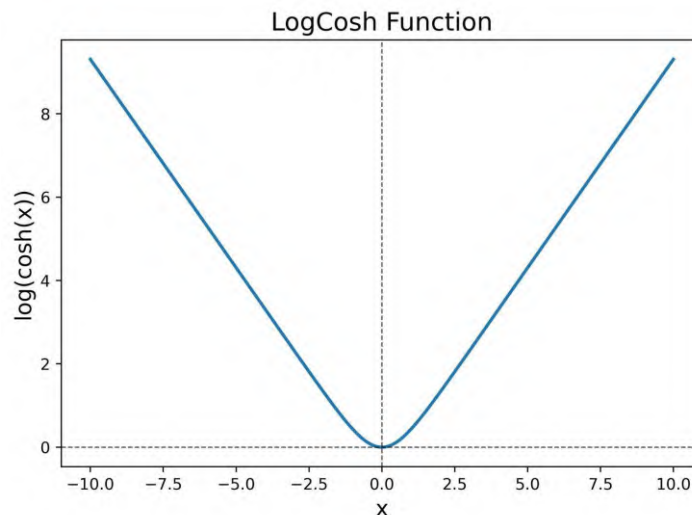
(3) 损失函数：

LogCosh函数

$$\mathcal{L}(y, \hat{y}) = \sum_{i=1}^n \log(\cosh(R_E))$$

能量对数残差

$$R_E = \lg(E_{\text{reco}}) - \lg(E_{\text{sim}})$$



Data

MC: Gamma (4.5m_p115mp250m_wmag_CE08_all)

Energy: 0.1TeV - 1000TeV

Selection Criteria

likelihood:

Nhit > 200

angdiff < 1deg

abs(xc) < 100m

abs(yc) < 100m

compactness > 0.9

GNN:

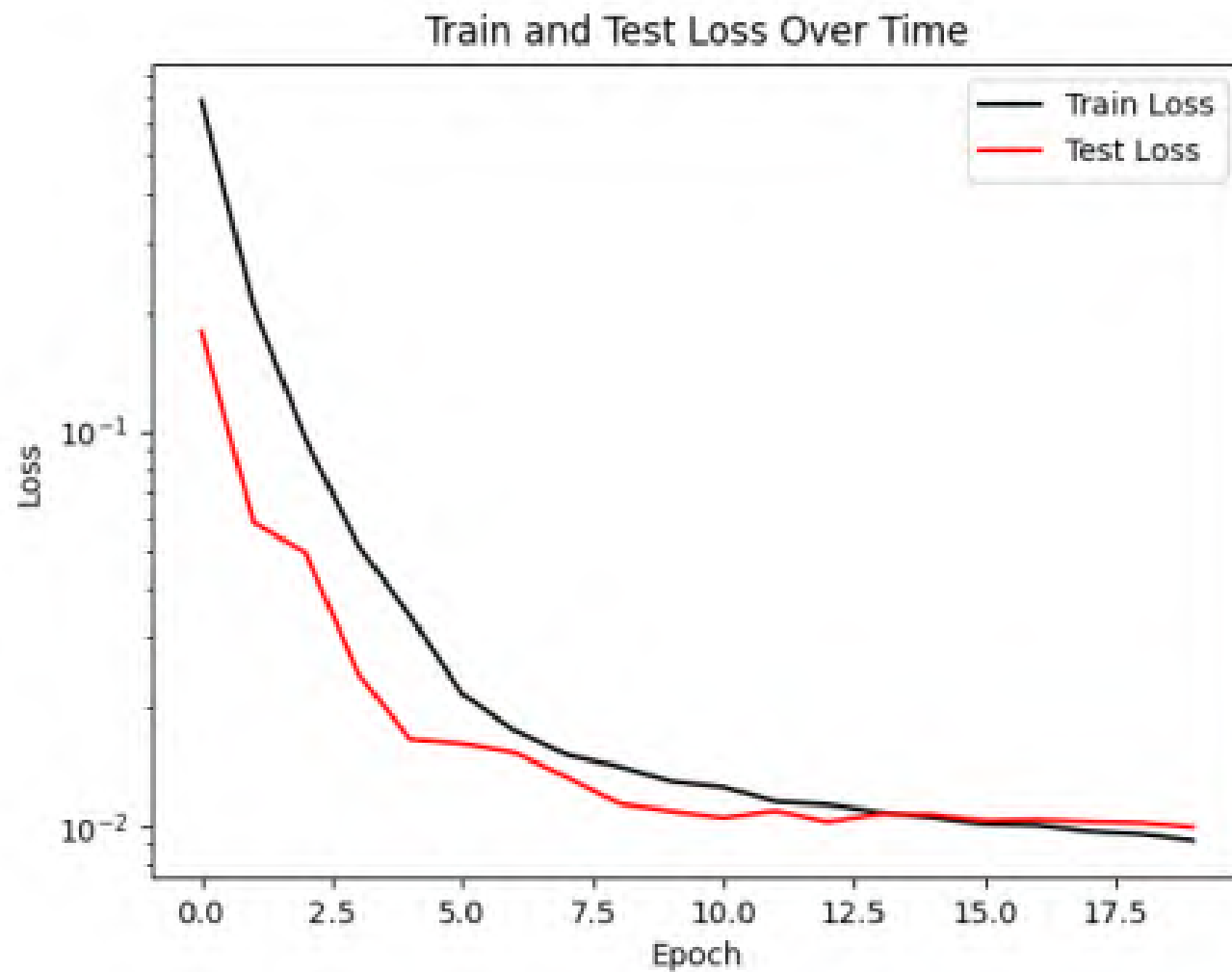
Nhit > 200

原理概述：训练流程

数据集	个数
训练集	103758
测试集	25940
验证集	19839

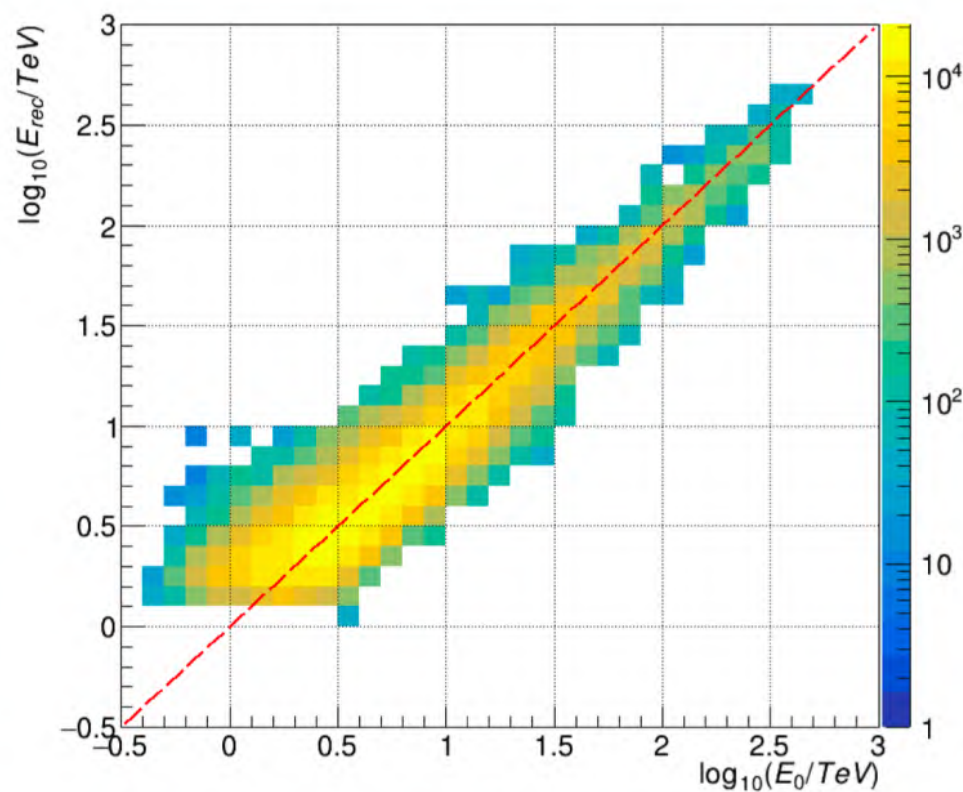
$$\mathcal{L}(y, \hat{y}) = \sum_{i=1}^n \log(\cosh(R_E))$$

$$R_E = \lg(E_{\text{reco}}) - \lg(E_{\text{sim}})$$

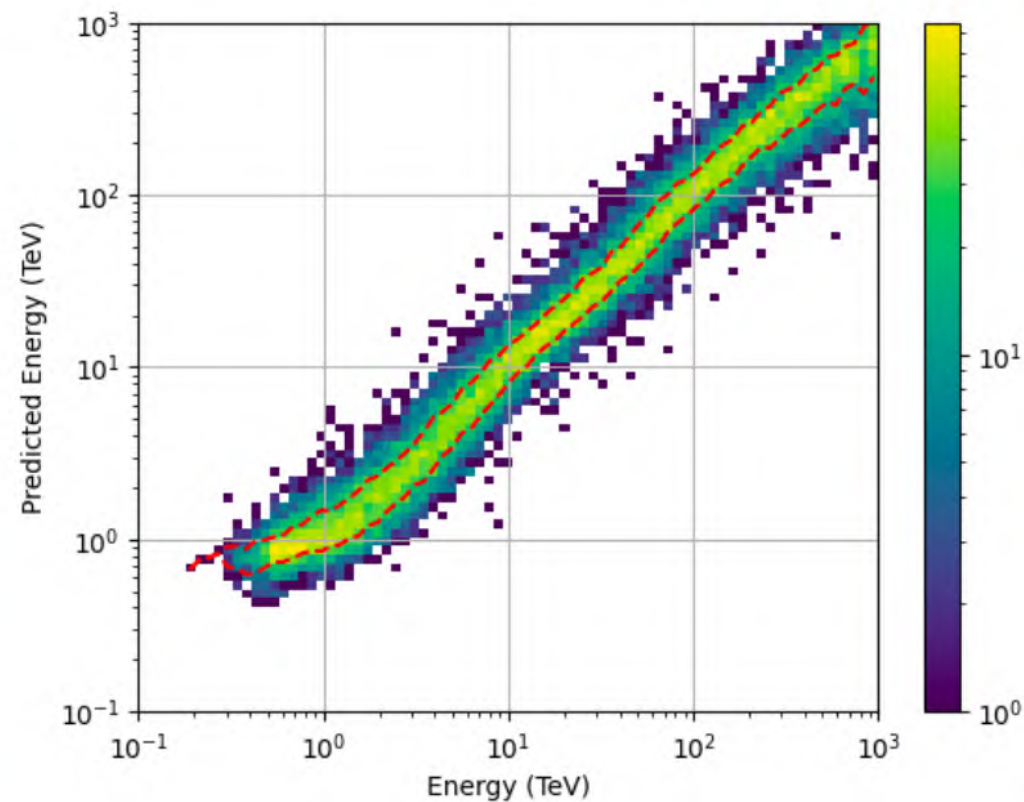


结果分析：重建结果与模拟数据比对

E_{sim} **vs** E_{rec}

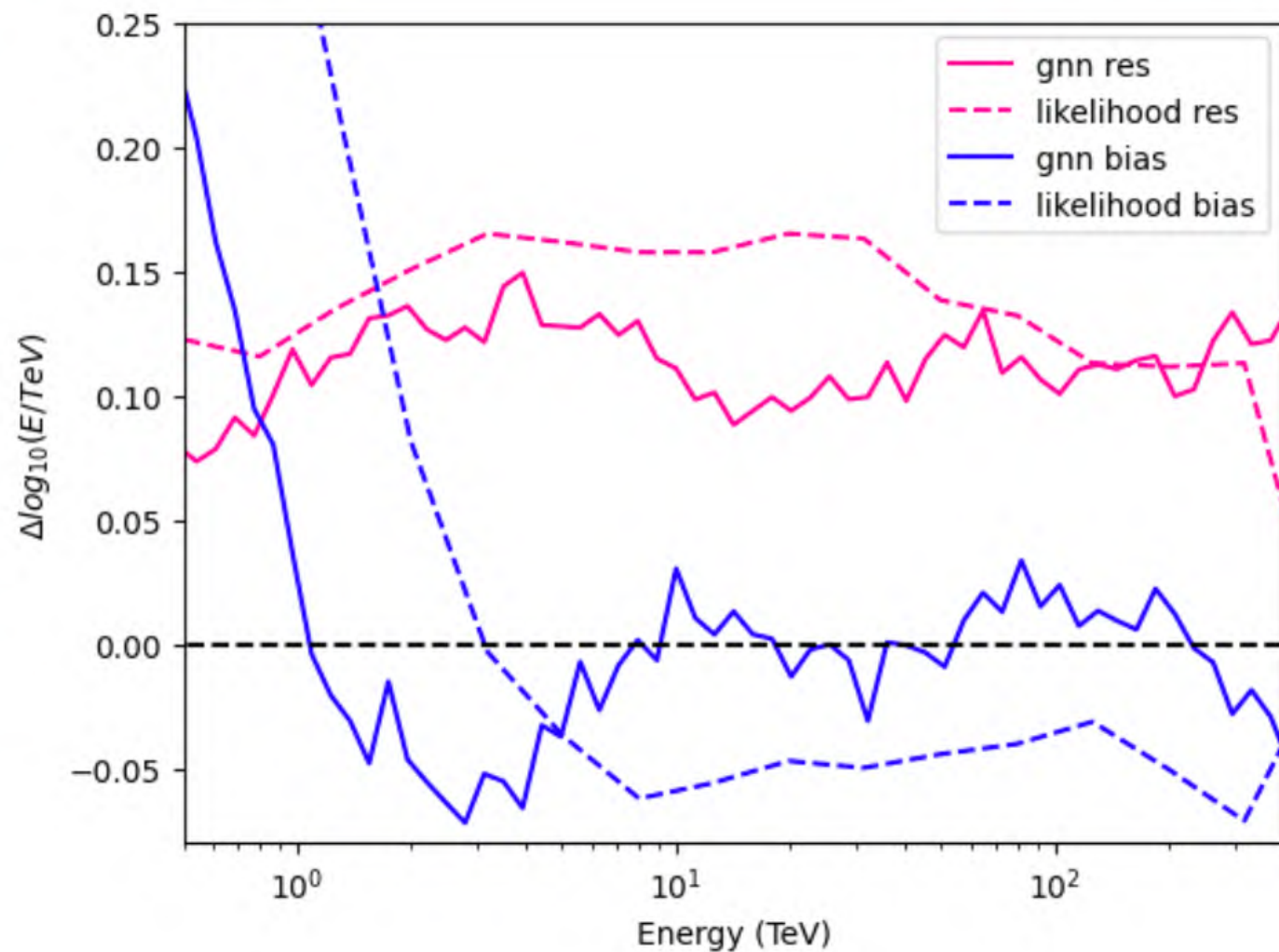


Likelihood



GNN

结果分析：误差评定



偏差值:

$$bias = E_{mean}^{rec} - E_{mean}^{sim}$$

分辨率:

$$res = \text{Resolution of Data}$$

项目总结

自己写!

后续计划：基于Transformer架构的新模型

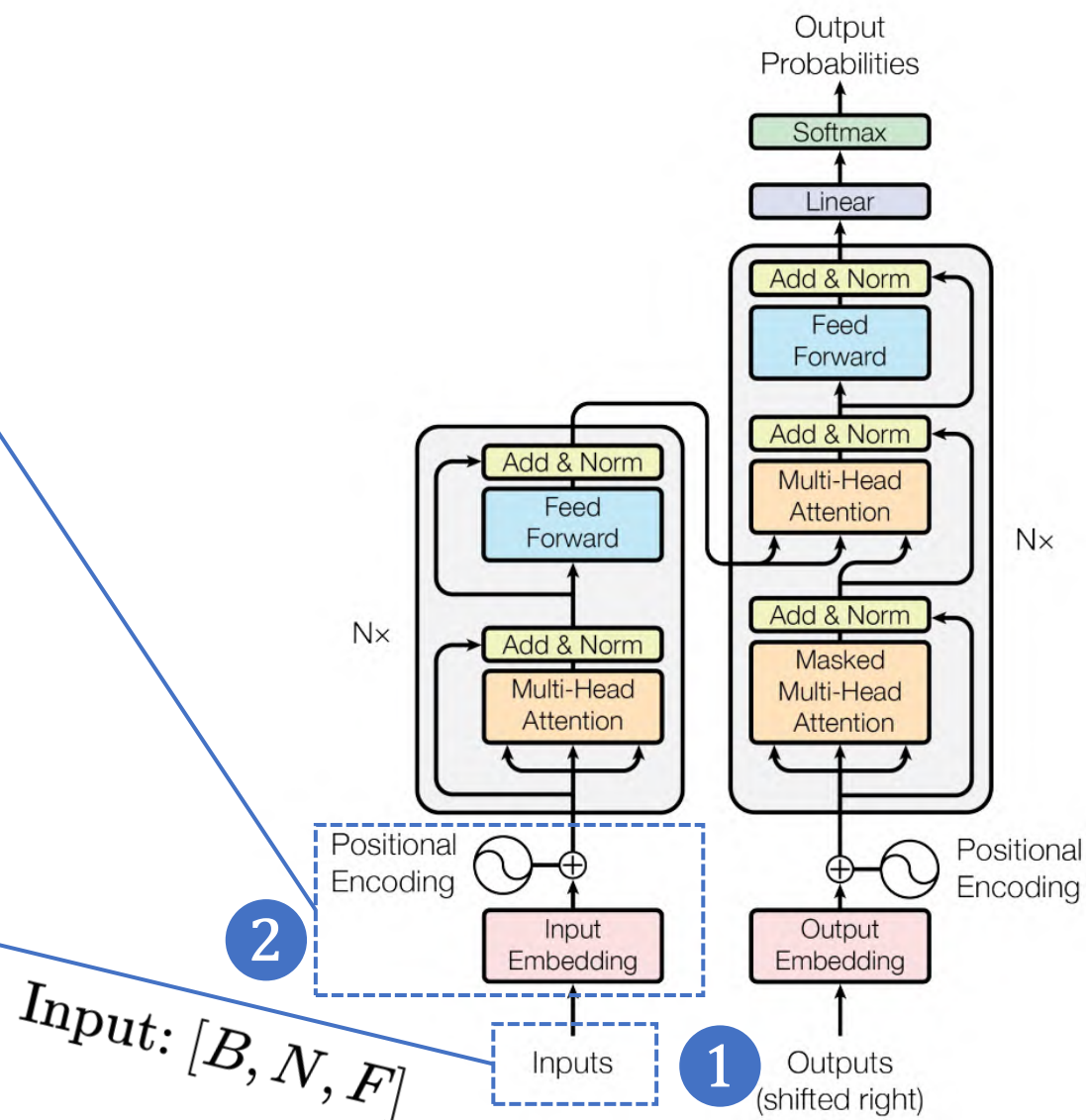
输入层：线性嵌入 + 位置编码

线性嵌入层：每个特征通过线性变换映射到一个高维嵌入空间。

位置编码 (Positional Encoding)：为节点序列加入位置编码，使模型感知脉冲的空间位置关系，位置编码采用标准正余弦方法。

输入数据：Transformer天然适合处理序列数据，因此我们将事件转化为序列输入，每个节点对应一个序列元素，包含空间位置、脉冲时间和电荷量。

B：批量大小；
N：节点数量（可变长度）；
F：每个节点的特征数。



后续计划：基于Transformer架构的新模型

Transformer编码器

多头自注意力层：

多头自注意力：Transformer的核心是多头自注意力机制，可以捕获全局脉冲关系。

输入：嵌入后的特征 $[B, N, D]$ ，每个头计算节点间的全局依赖：

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

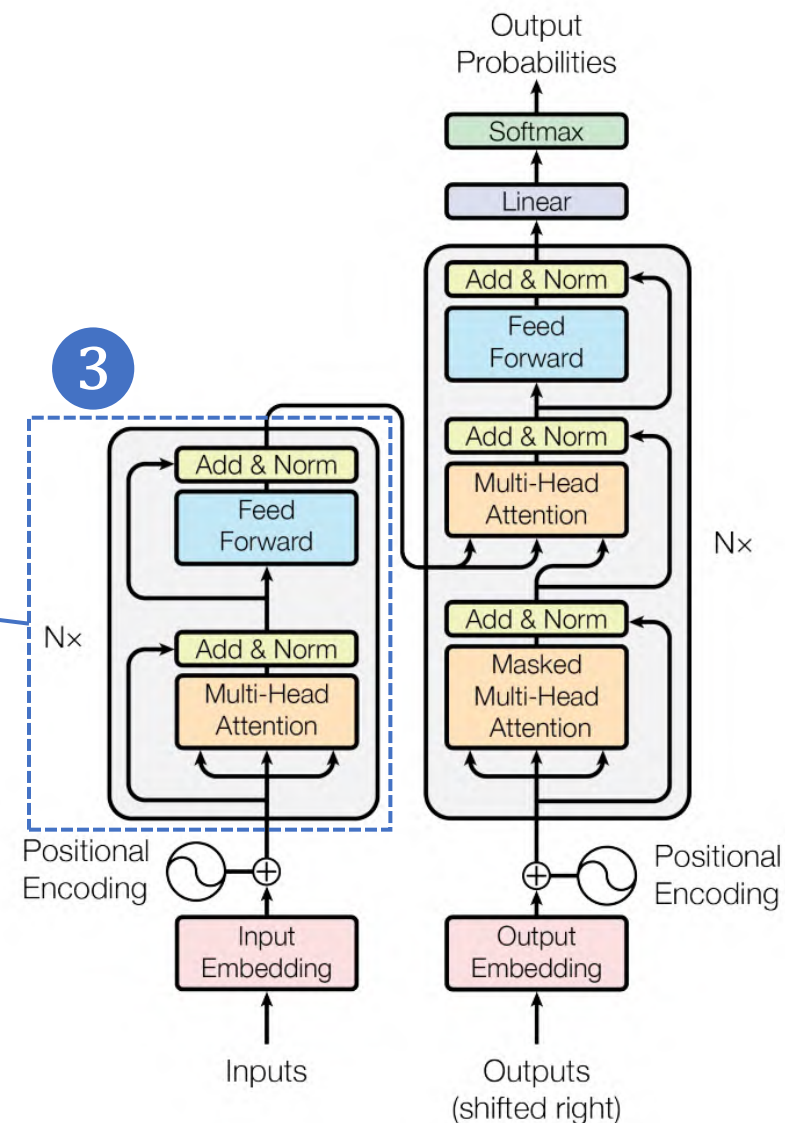
(Q, K, V)：由输入特征线性投影得到的查询、键和值矩阵。

前馈网络 (FFN)：逐节点应用两层全连接网络，激活函数为ReLU。

层归一化与残差连接：每层后加LayerNorm和残差连接以稳定训练。

堆叠多层编码器：堆叠多层Transformer编码器。

动态节点：引入掩码矩阵。



后续计划：基于Transformer架构的新模型

能量重建任务的**损失函数**：选取与之前相同的损失函数。

$$\mathcal{L}(y, \hat{y}) = \sum_{i=1}^n \log(\cosh(R_E))$$

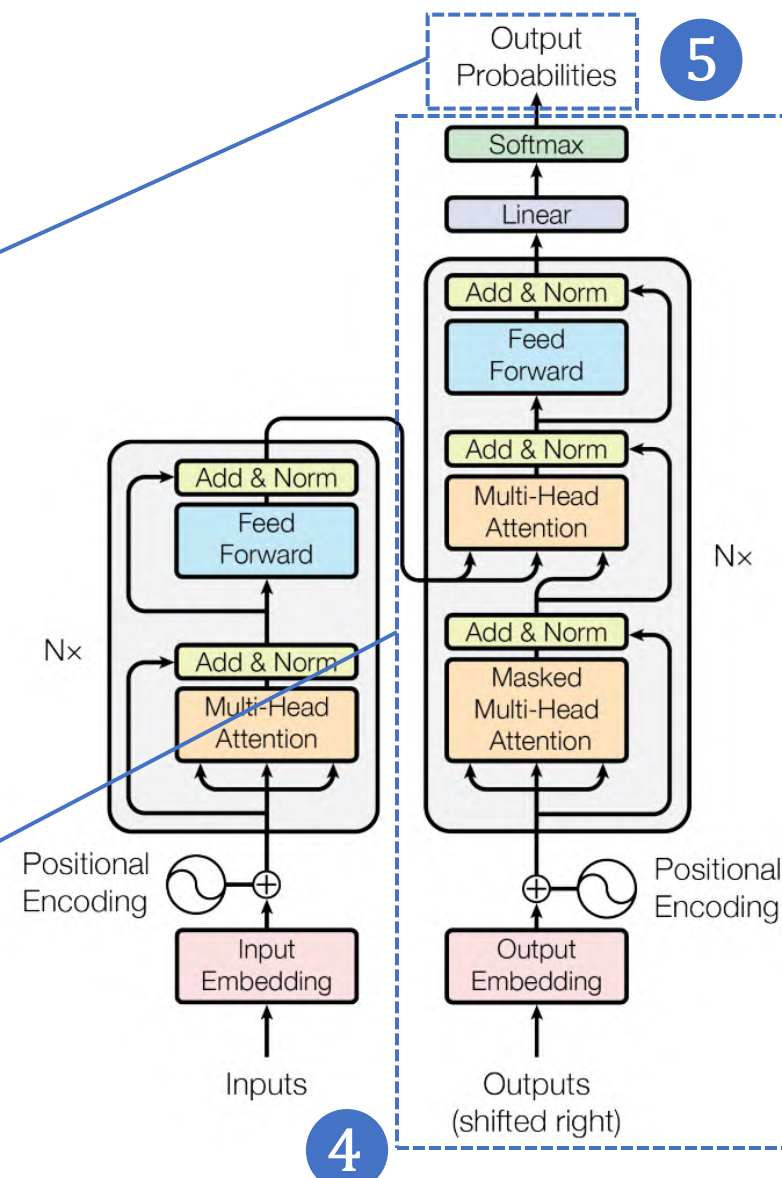
$$R_E = \lg(E_{\text{reco}}) - \lg(E_{\text{sim}})$$

特征聚合：

Transformer输出的是每个节点的高维特征。为了生成事件级别的全局特征，我们引入特征聚合策略。

全局池化：对所有节点特征进行 (Max, Min, Mean, Sum) 聚合，输出形状为 [B, D_global]。

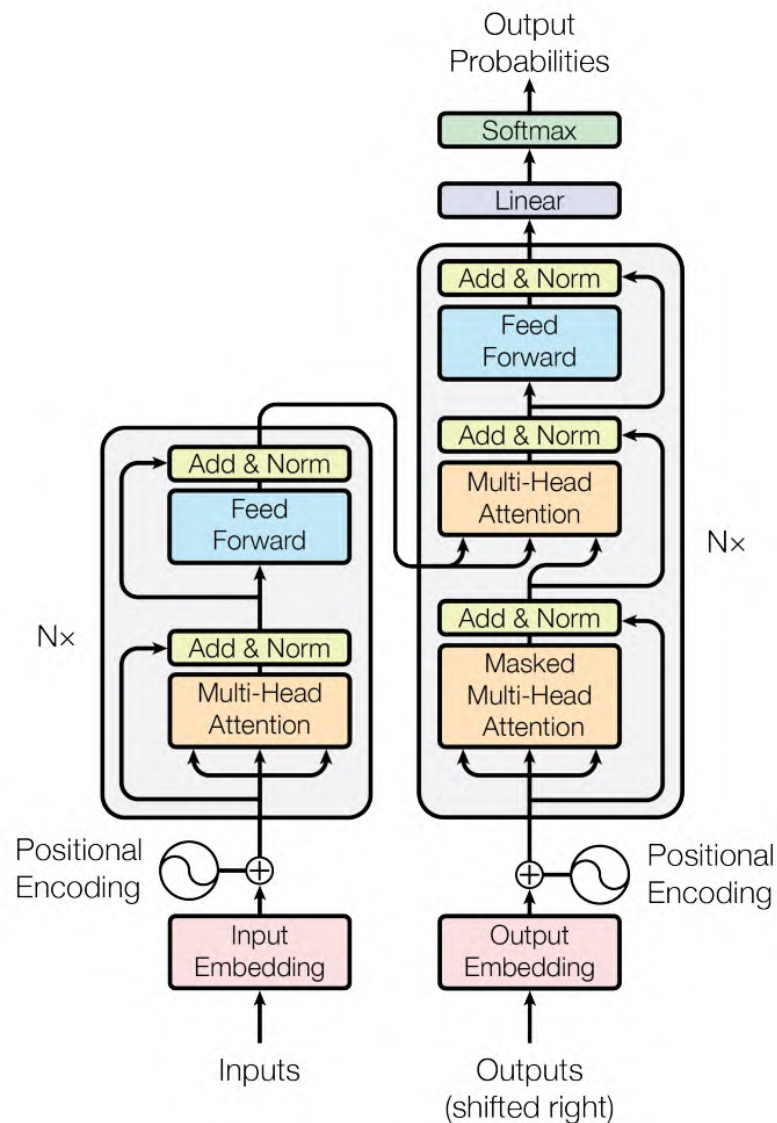
能量重建任务特定的头部：使用全连接层输出回归值。



后续计划：新模型预期优势与局限

优势： 具有全局建模能力，可以灵活处理动态数据，并且会有高度泛用性即可用于其它任务。此外由于无需预定义图结构，它的硬件效率与扩展性也非常好，具有鲁棒性与适应性。

局限： 注意力机制计算复杂度较高，对大规模训练数据依赖较强，但可通过优化方法和数据增强解决



致谢

自己写!