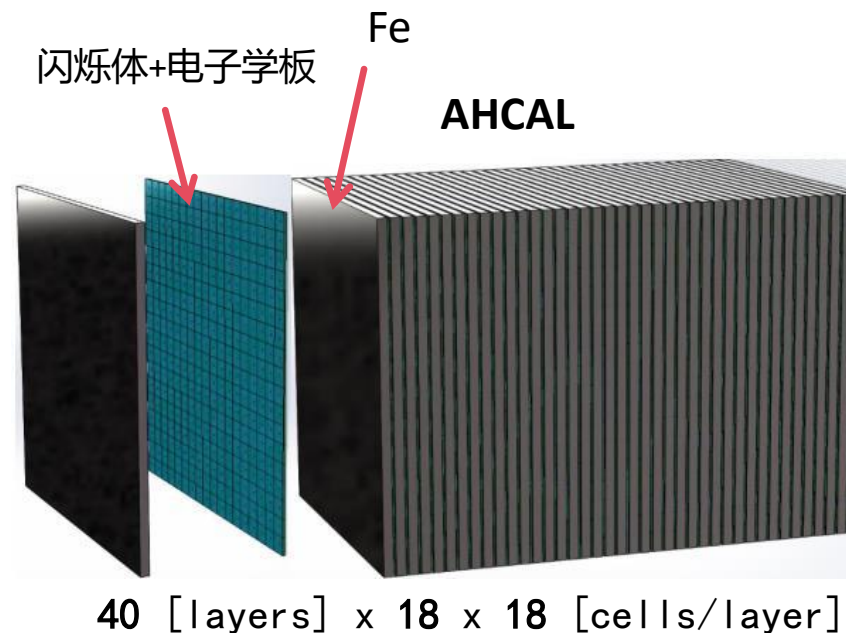
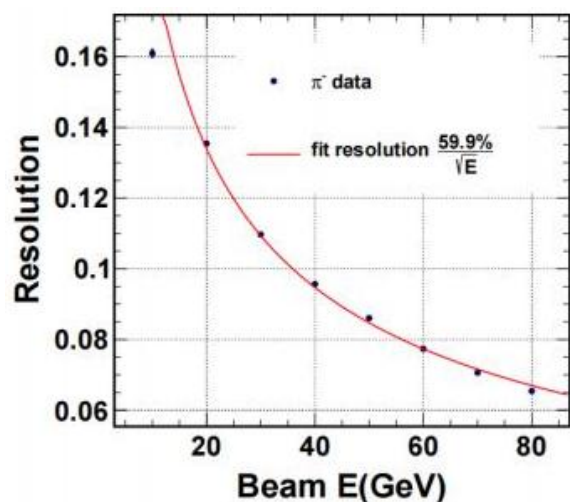


基于机器学习的 AHCAL能量重建研究

报告人：彭柏融
中国科学技术大学
2026.6.3

- **CEPC A(nalog) H(adronic) CAL(orimeter)**
- 40个灵敏层 (吸收体铁板+闪烁体层 + PCB电子学板+吸收体铁板)
- 能量分辨率达到目标值 ($\frac{60\%}{\sqrt{E} (GeV)}$)
- 机器学习——>能量重建精度提升



束流测试

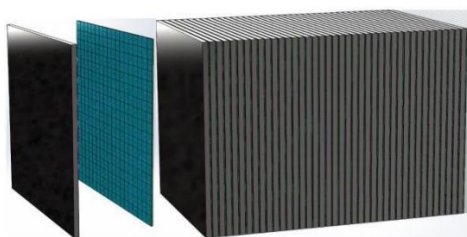
三项束流测试, CERN

- μ 子位置扫描(100GeV/c)
- pion (1-120 GeV/c)
- 电子 (1-120 GeV/c)



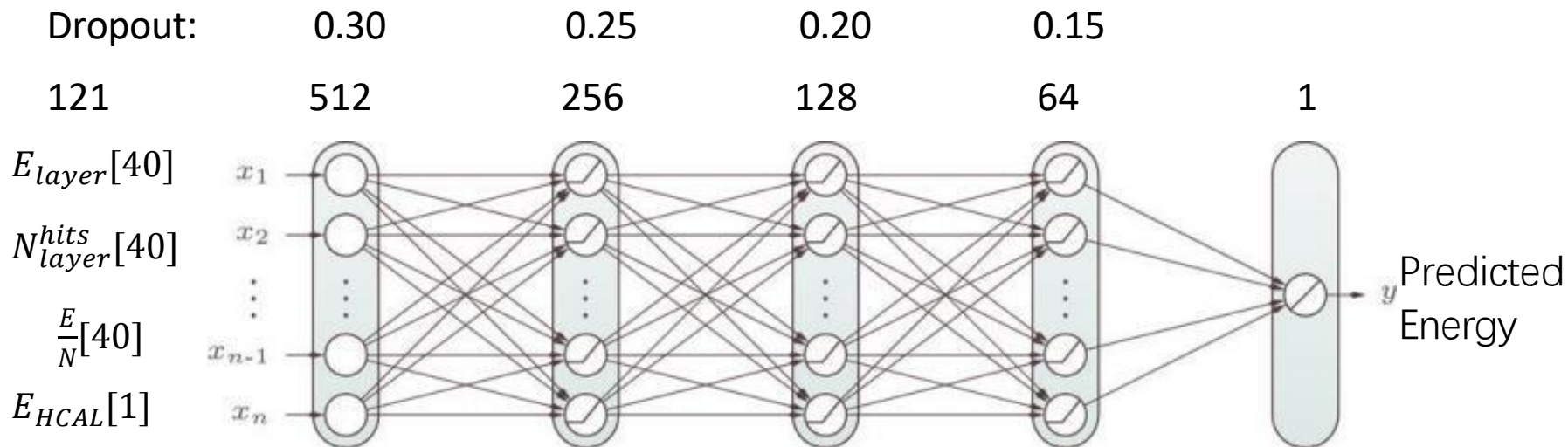
MLP模型

- 输入：探测器每层的能量(E)、击中数(N)、能量密度(E/N)和HCAL总沉积能量 E_{HCAL} ($40*3+1=121$)
- 输出：预测的入射能量 (1)



40 [layers] x 18 x 18 [cells/layer]

batch	64
学习率	1e-4
学习率调度	CosineAnnealing, T_max=200, eta_min=1e-7
最大epoch	200

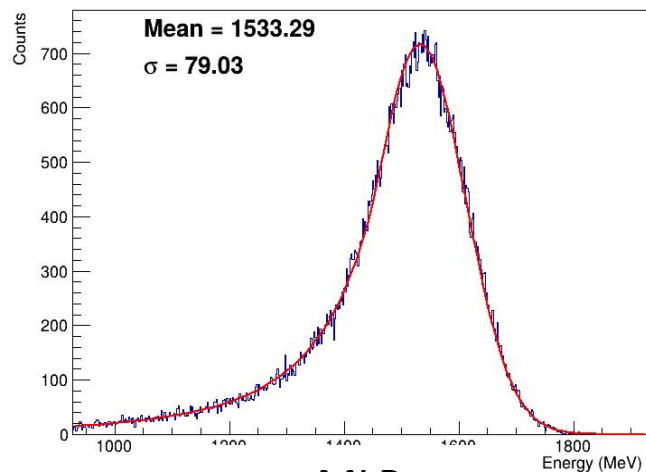


Loss: $L = \frac{1}{N} \sum_n e_n$ $e = \frac{(y_{true} - y_{pred})^2}{y_{true}^2}$ 均方相对误差(MSRE)

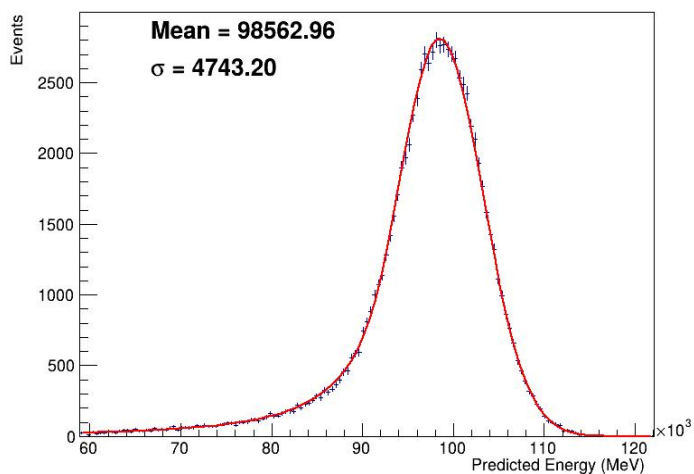
能量谱图@100GeV

MC

Raw

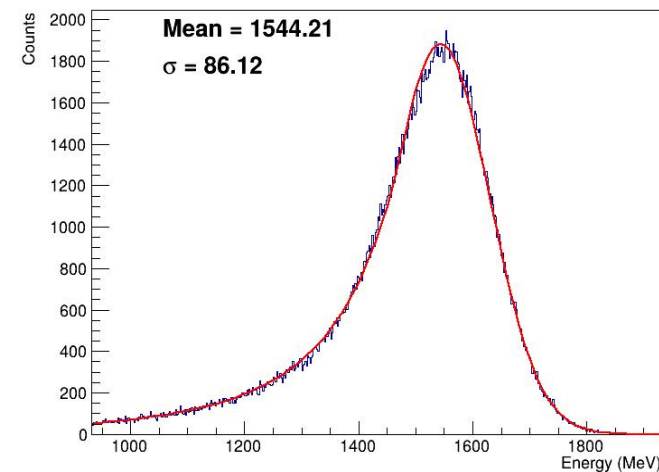


MLP

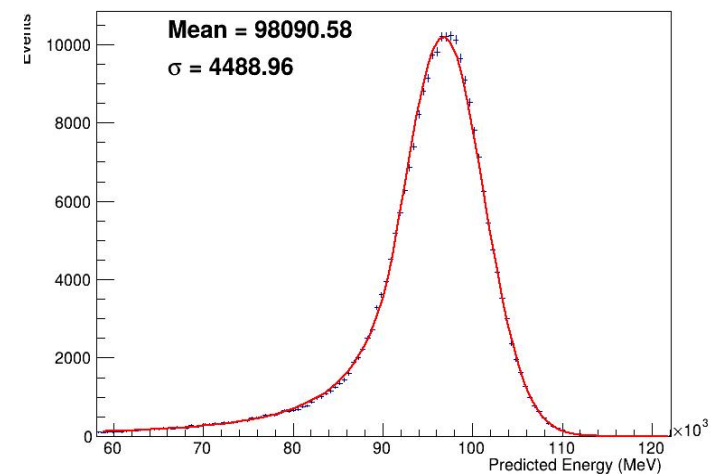


Data

Raw

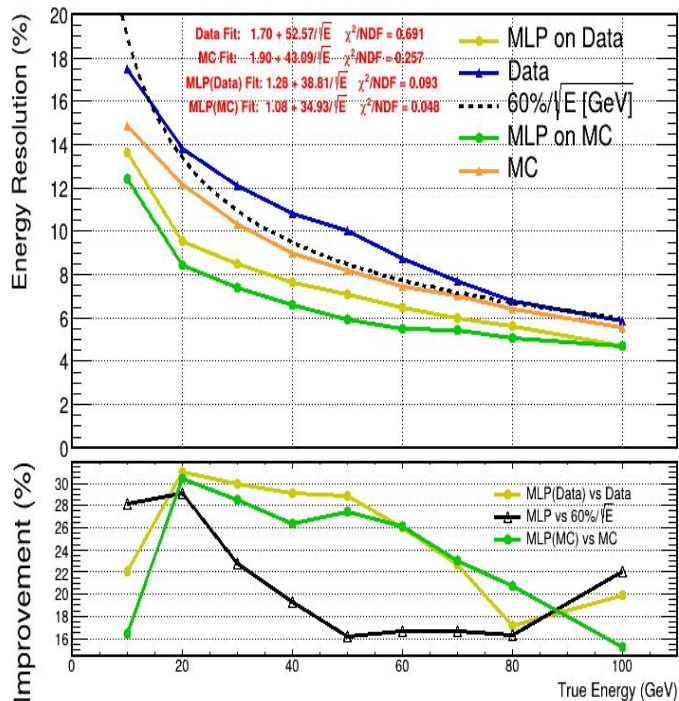


MLP

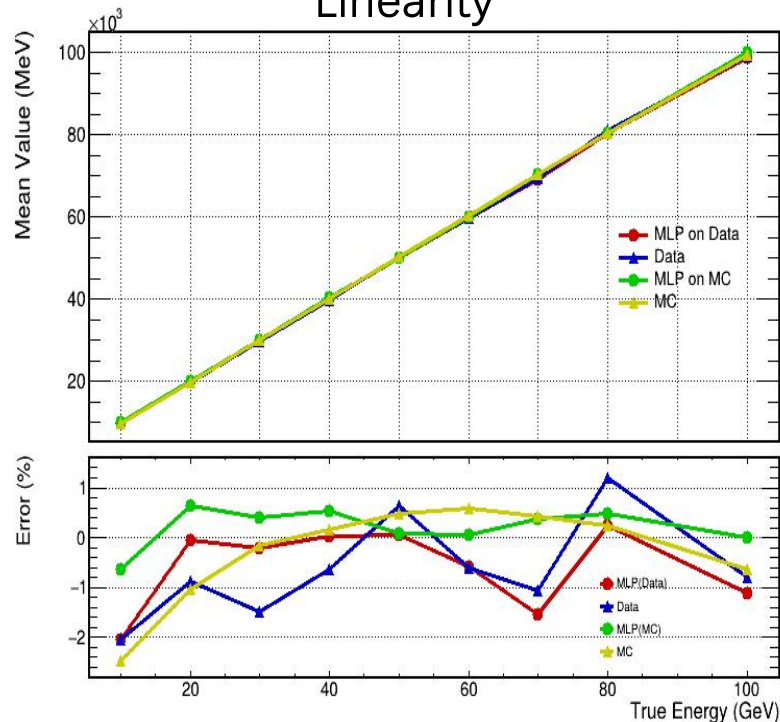


测试结果

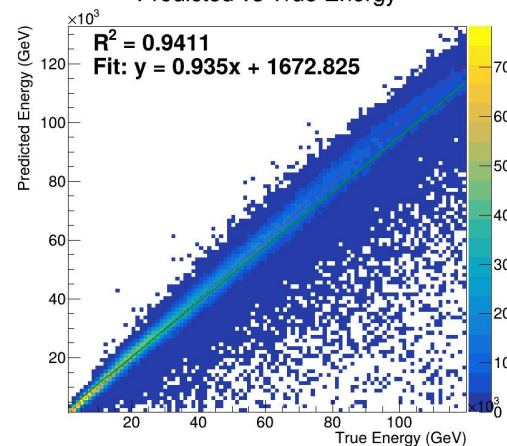
Resolution



Linearity



Predicted vs True Energy



• 训练样本

源: pion

范围: 1-120GeV

数量: train = 70.3e4

test = 12.4e4

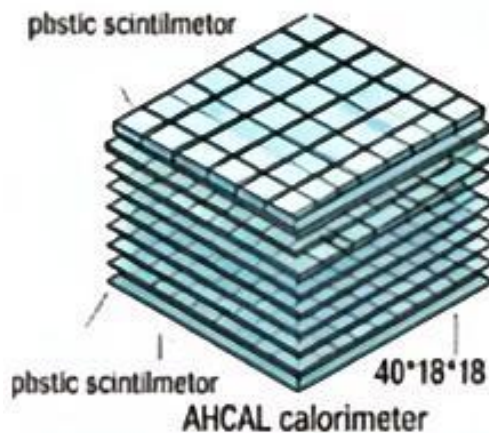
- Data 整体上略劣于MC (平均0.83%)
- 泄露事例减少
- 分辨率: MC 和 Data 都能实现约 25%的提升
- 能量线性: Data颇有改善, 平均 -0.58%
MC 保持良好, 平均 0.21%

Energy(GeV)		10	20	30	40	50	60	70	80	100
Data	Event (left) ($\times 10^4$)	8.3	30.2	48.6	45.8	49.8	36.9	32.7	28.3	25.3
	Percentage (%)	8.1	40.3	44.9	44.1	39.5	35.6	30.9	26.3	19.7
MC	Event (left) ($\times 10^4$)	20.1	21.1	19.9	18.6	16.9	15.2	13.3	11.6	8.7
	Percentage (%)	50.4	52.7	49.8	46.5	42.3	38.0	33.4	29.2	22.1

$$Imp = \frac{|R_{MC} - R_{MLP}|}{R_{MC}}$$

$$Error = \frac{|Mean - True|}{True}$$

1. Physical Input: Orthogonal Grids



2. Spatial Extraction - Mean Pooling

2D grid hits

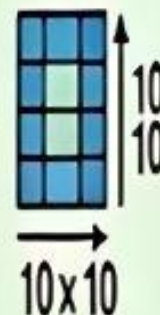
Conv2d → GroupNorm → GELU



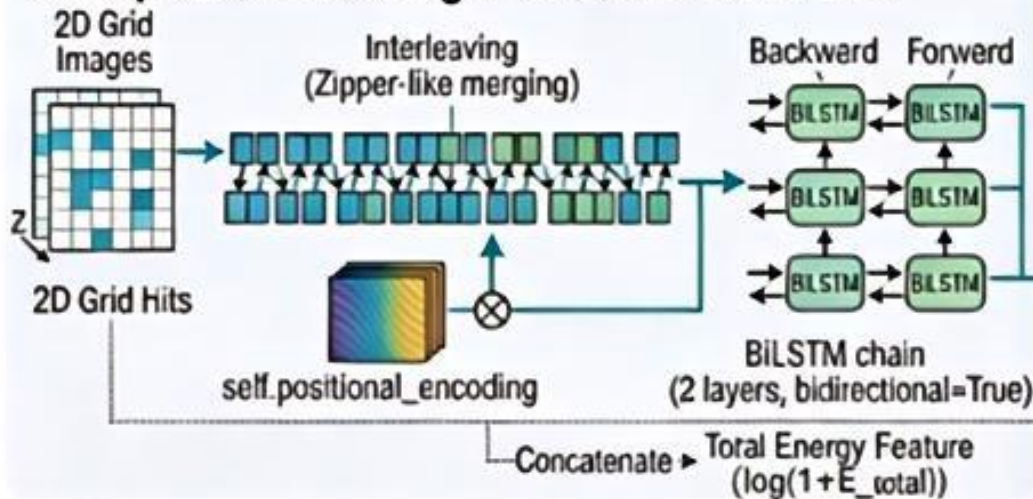
Adaptive Mean Pooling



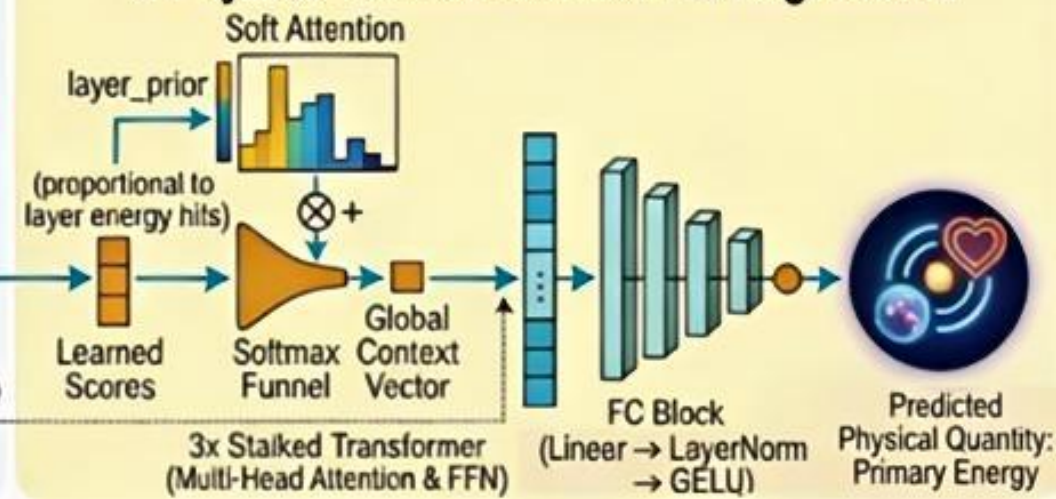
Output 10x10



3. Sequence Encoding: Positional & BiLSTM

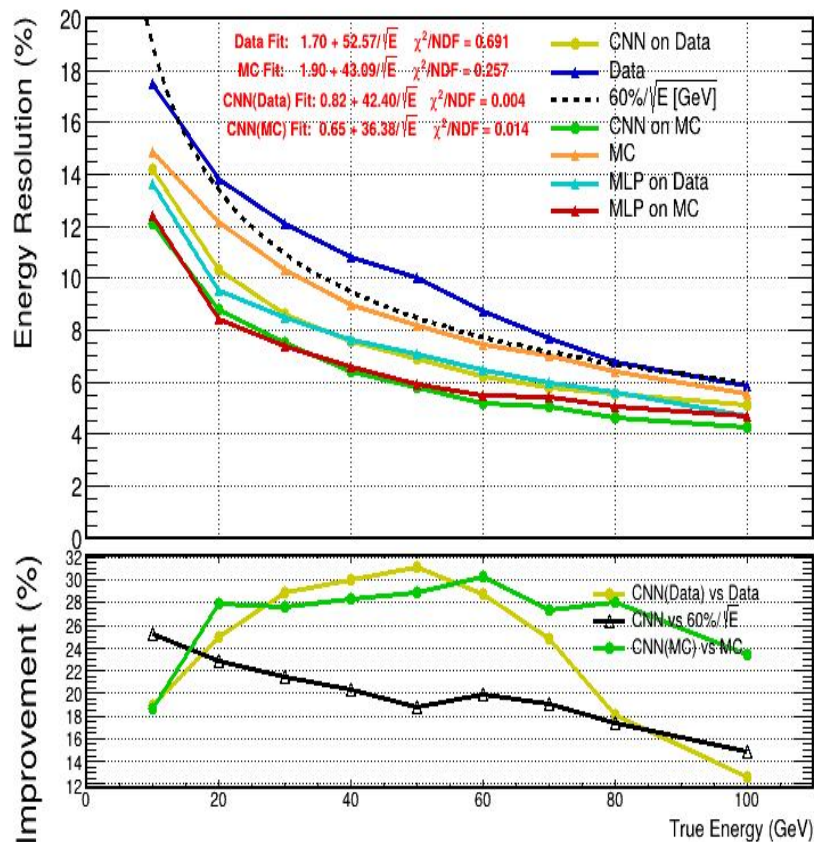


4. Physics-Guided Attention & Regression

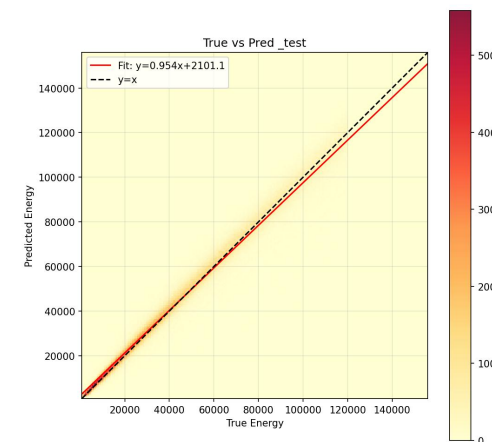
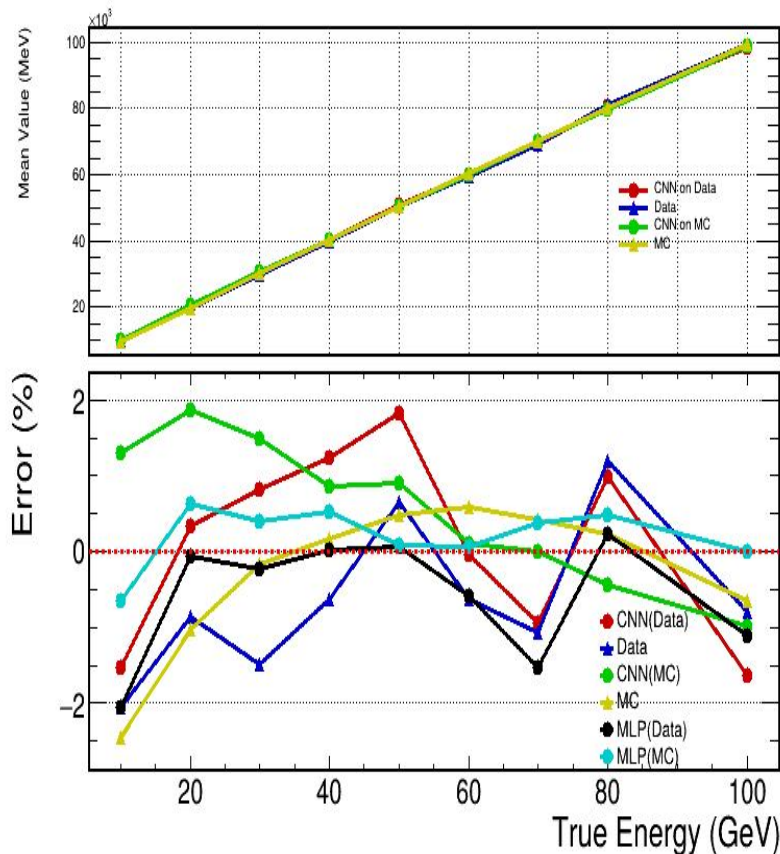


CNN测试结果

Resolution



Linearity



- 分辨率:
MLP和CNN都实现约25%的提升
- 能量线性:
MLP颇有改善, 平均0.65%
CNN改善有限, 平均1.04%
- 与MLP相比,

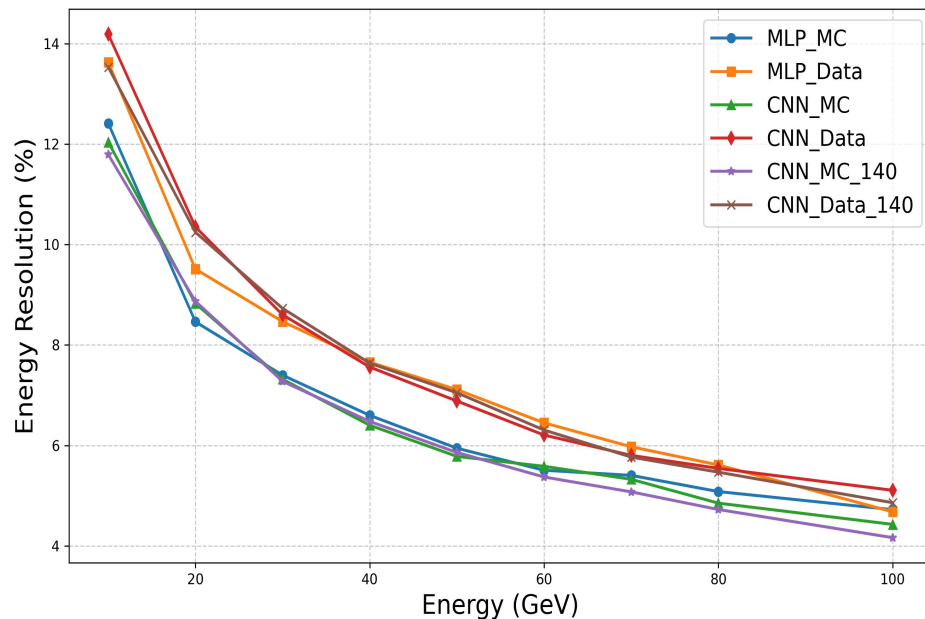
CNN在AHCAL上暂无明显优势

$$Imp = \frac{|R_{MC} - R_{MLP}|}{R_{MC}}$$

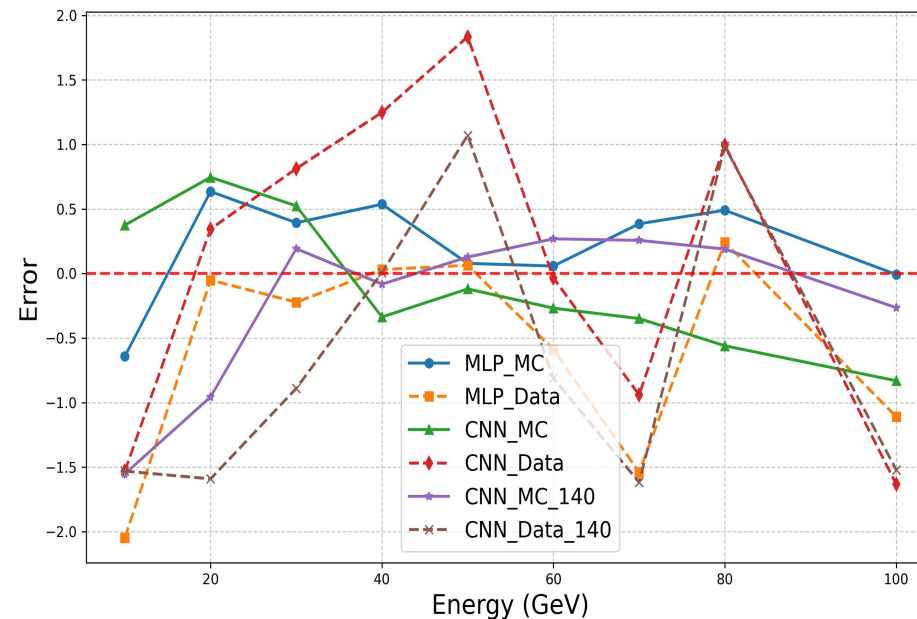
$$Error = \frac{|Mean - True|}{True}$$

more event测试结果

Resolution

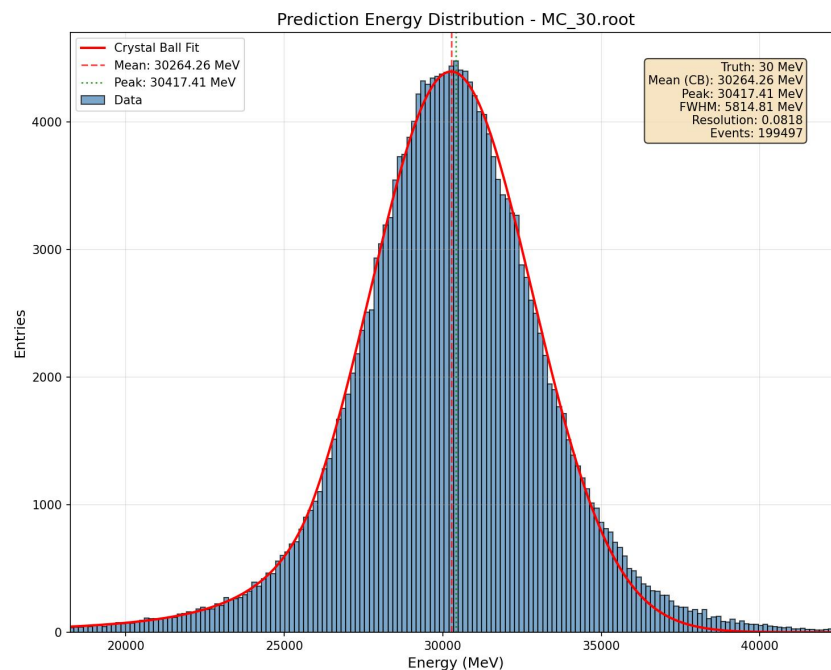
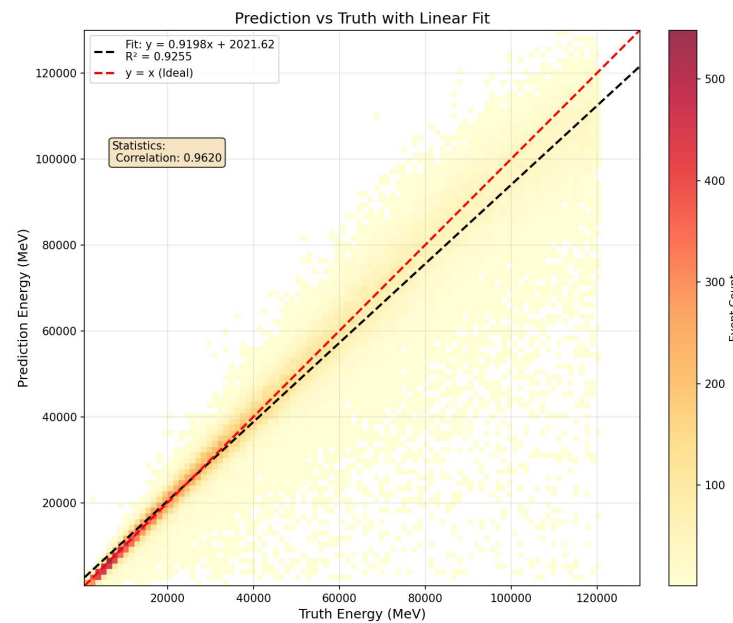
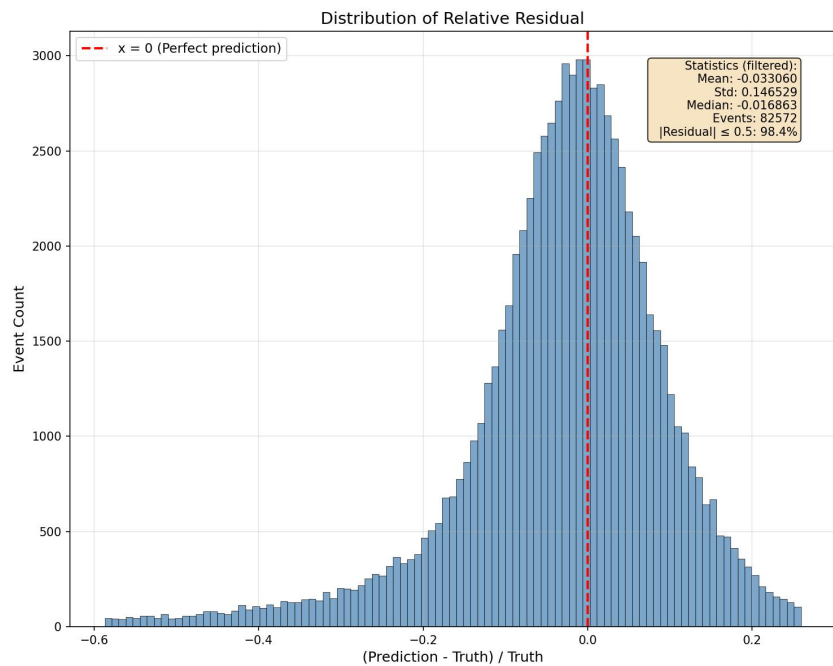


Linearity



$$Error = \frac{|Mean - True|}{True}$$

- CNN, 训练Event数目翻倍
- 分辨率:
MC上25%-->27%; Data上24%-->25%
- 能量线性:
MC上1.04%-->0.43%; Data上暂无



- 只训练2个epoch
- 模型学习相关性良好

- AHCAL : 40 layers \times 18 \times 18 cells/layer
- MLP Model : $E_{layer}[40] + N_{layer}^{hits}[40] + \frac{E}{N}[40] + E_{HICAL}[1]$ (121) \rightarrow Predicted Energy (1)
- CNN Model : 18*18*40 cells \rightarrow Predicted Energy (1)

- 测试情况:

Thanks for your attention!

MLP: MC 和 Data 都有约25%的分辨率提升

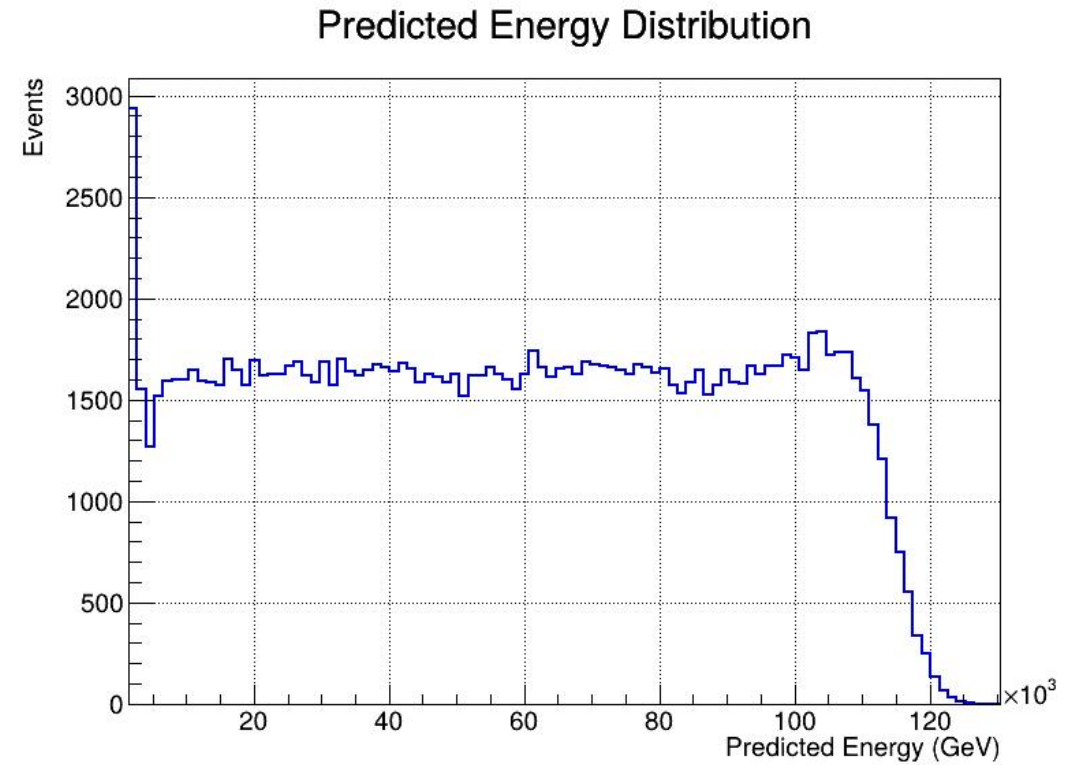
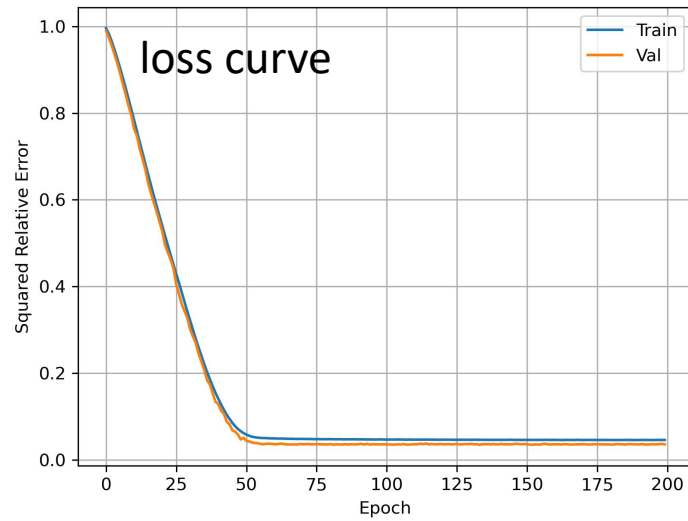
MC能量线性巨大提升, Data相对MC略差, Error平均 -0.58 %

CNN: MC 和 Data 也都有约25%的分辨率提升

MC和 Data 能量线性都提升有限, Error平均 1.04 %

- CNN在AHCAL上相对于MLP暂无明显优势

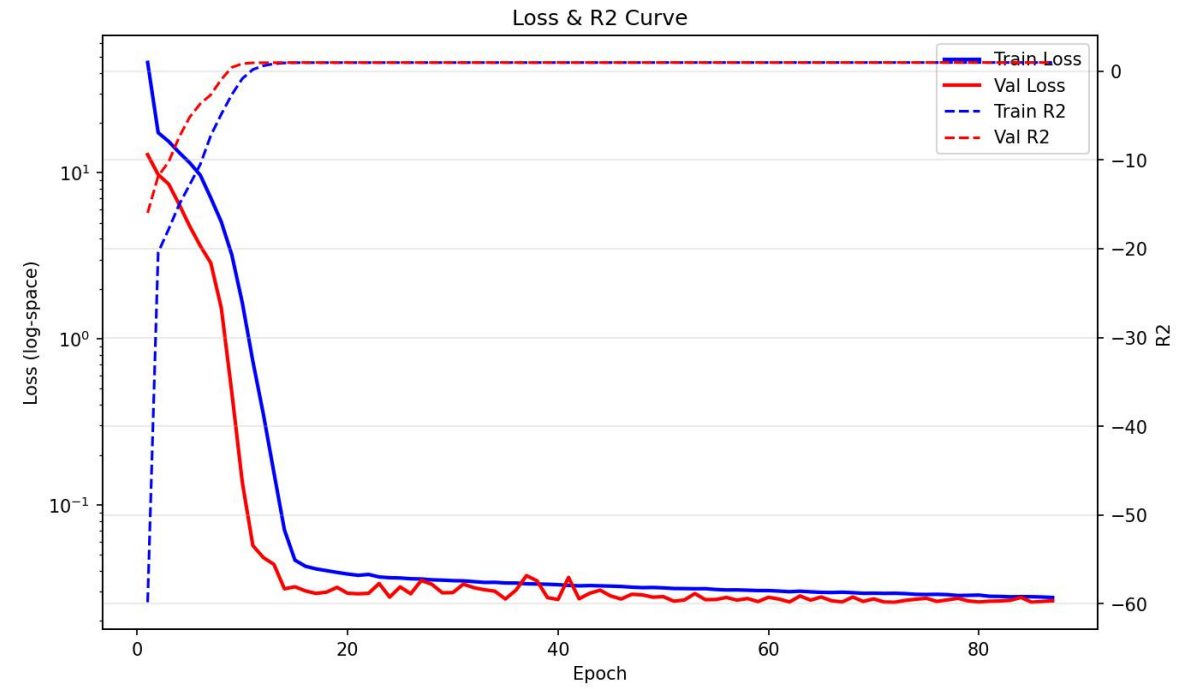
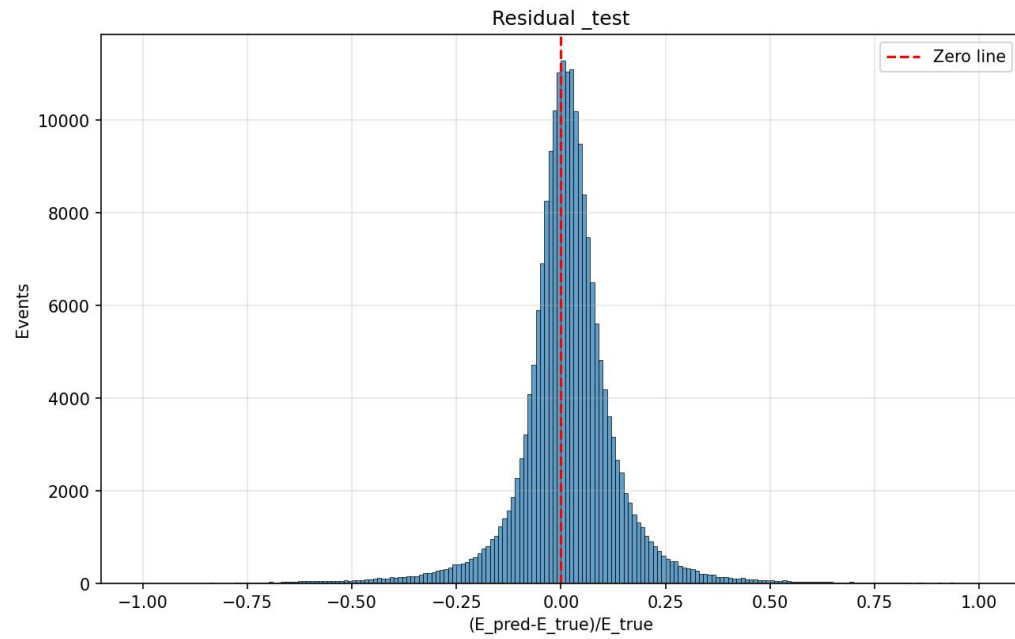
back up



back up

event	energy(G eV)	10	20	30	40	50	60	70	80	100
Data	cut前	1,027,814	749,791	1,081,725	1,038,984	1,258,699	1,034,938	1,058,822	1,073,061	1,286,608
	cut后	83,860	302,303	486,653	458,874	498,094	369,281	327,979	283,120	253,900
	占比 (%)	8.16	40.32	44.99	44.17	39.57	35.68	30.98	26.38	19.73
MC	cut前	400,000	400,000	400,000	400,000	400,000	400,000	400,000	400,000	394,000
	cut后	201,676	211,115	199,497	186,119	169,367	152,117	133,628	116,822	87,346
	占比 (%)	50.42	52.78	49.87	46.53	42.34	38.03	33.41	29.21	22.17
Train : 703,747					Test : 124,035					

back up



back up

