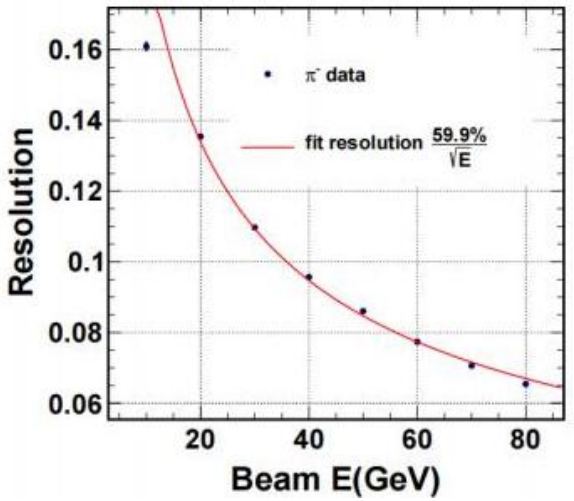
基于机器学习的 AHCAL能量重建研究

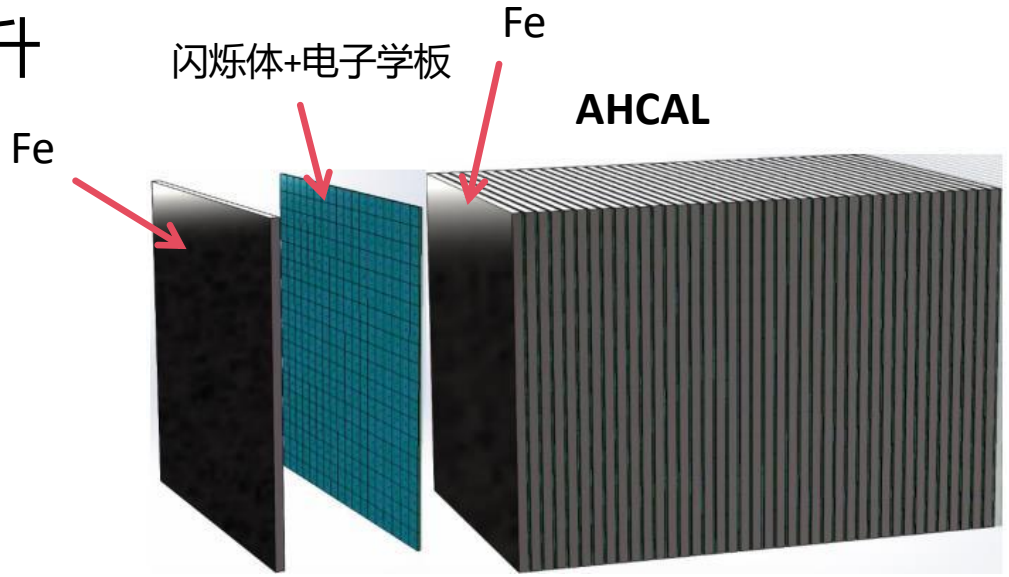
报告人：彭柏融
中国科学技术大学

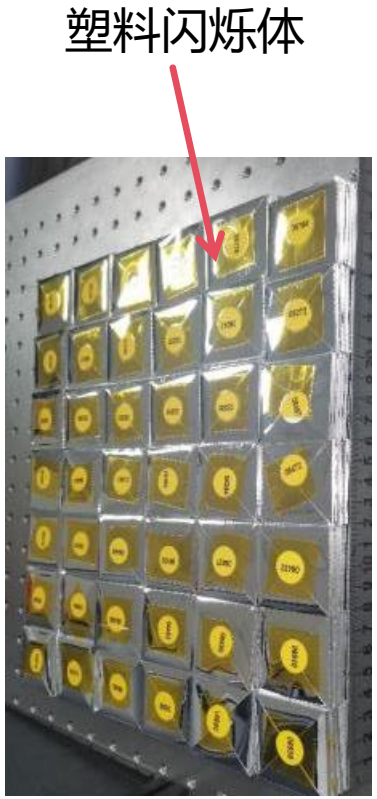
2026.3.17

研究背景

- **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







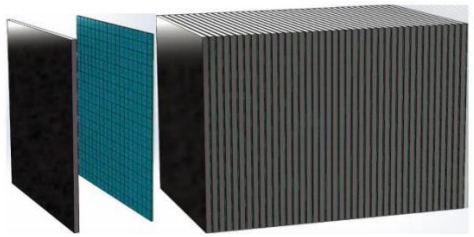
束流测试

三项束流测试,CERN

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

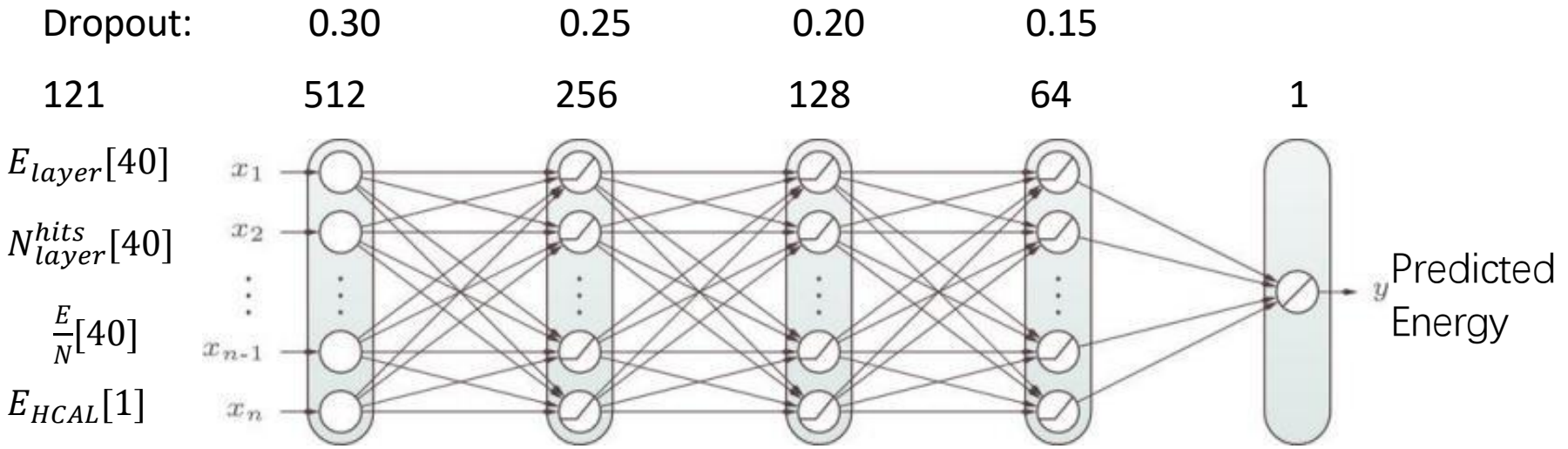
MLP模型

- 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]

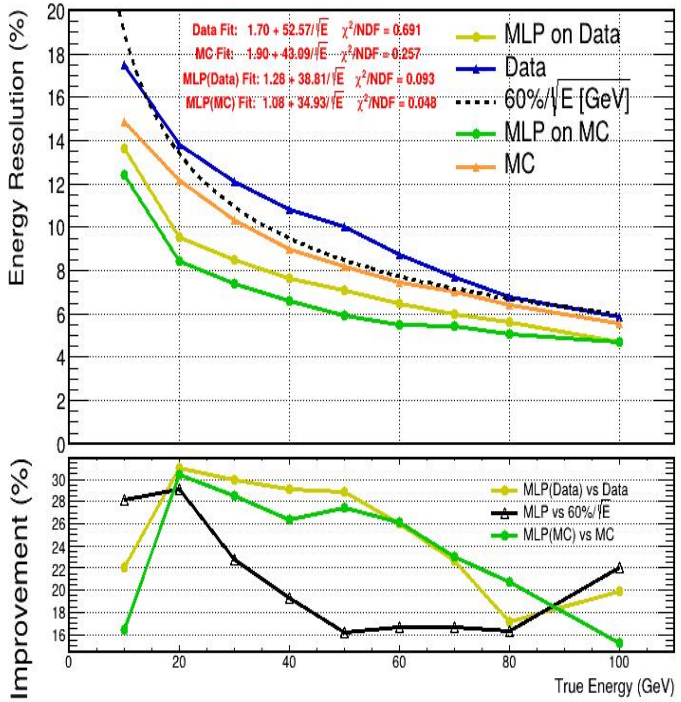
batch	64
学习率	1e-4
学习率调度	CosineAnnealing, T_max=200, eta_min=1e-7
最大epoch	200



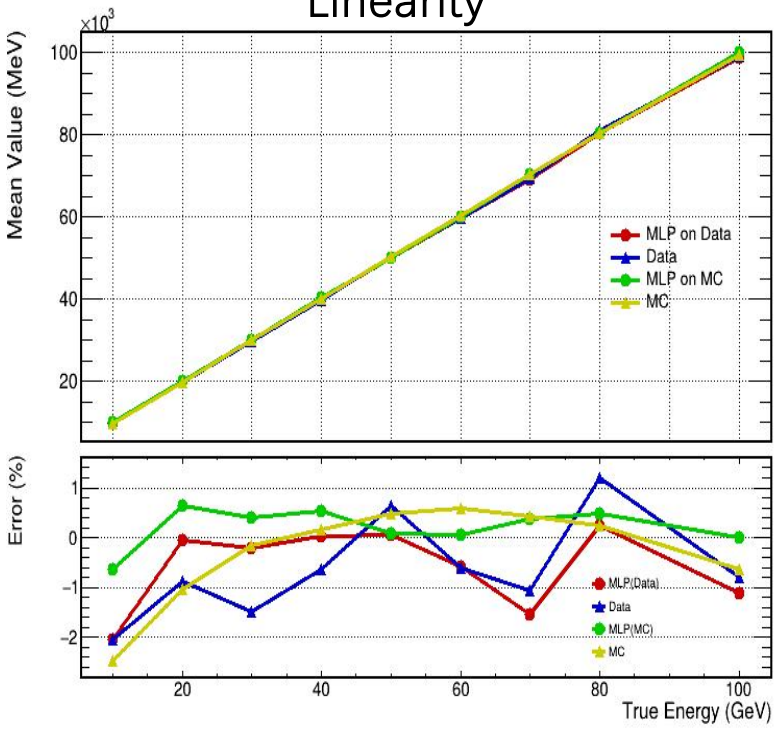
Loss: $L = \frac{1}{N}\sum_n e_n$ $e = \frac{(y_{true} - y_{pred})^2}{{y_{true}}^2}$ 均方相对误差(MSRE)

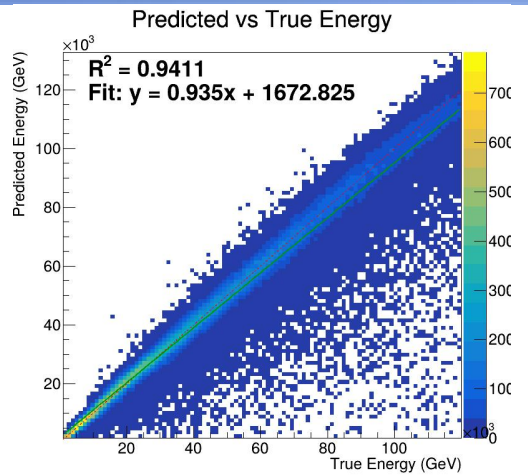
测试结果

Resolution



Linearity





- **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

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

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

总结

- AHCAL : 40 layers × 18 × 18 cells/layer
- MLP Model : $E_{layer}[40] + N_{layer}^{hits}[40] + \frac{E}{N}[40] + E_{HCAL}[1](121)$——>Predicted Energy (1)
- MC数据情况

  约有24%的分辨率提升

  能量线性巨大提升，Error平均0.21%
- 束流数据情况(pion,1-120GeV/c)

  约有25%的分辨率提升

  能量线性也有提升，但相对MC略差，Error平均-0.58%

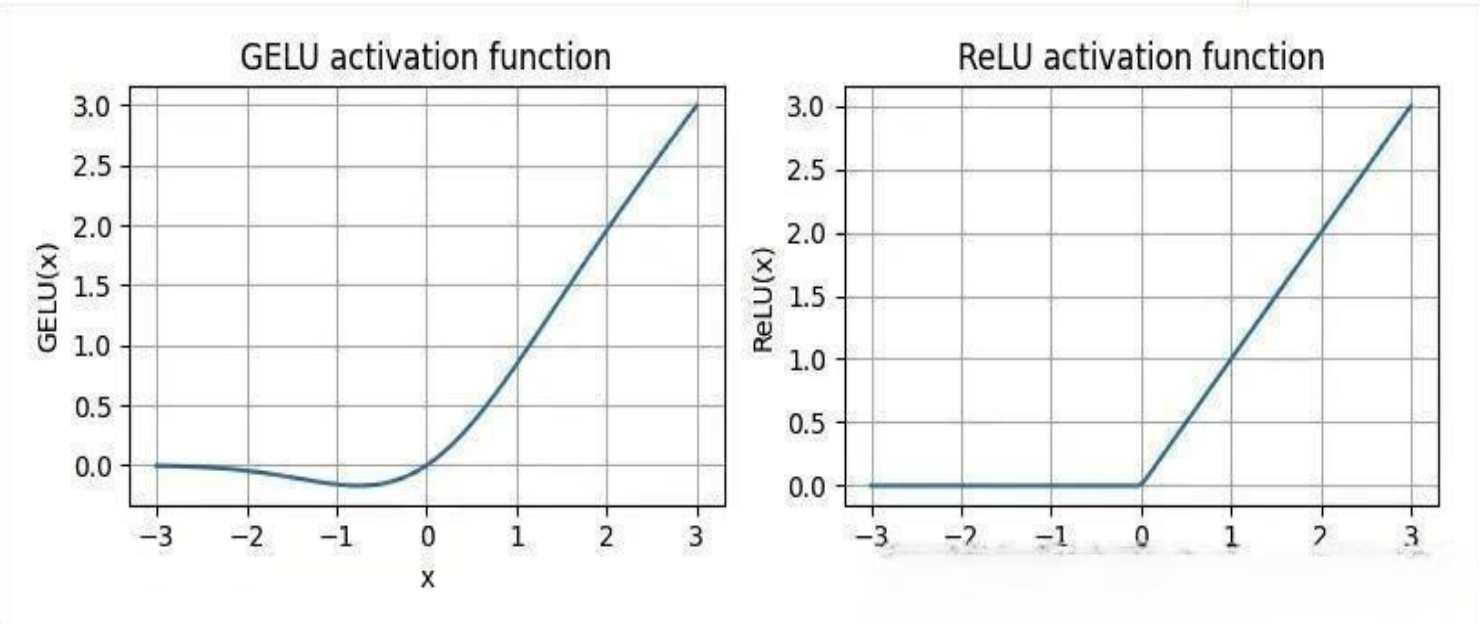
Thanks for your attention!

展望

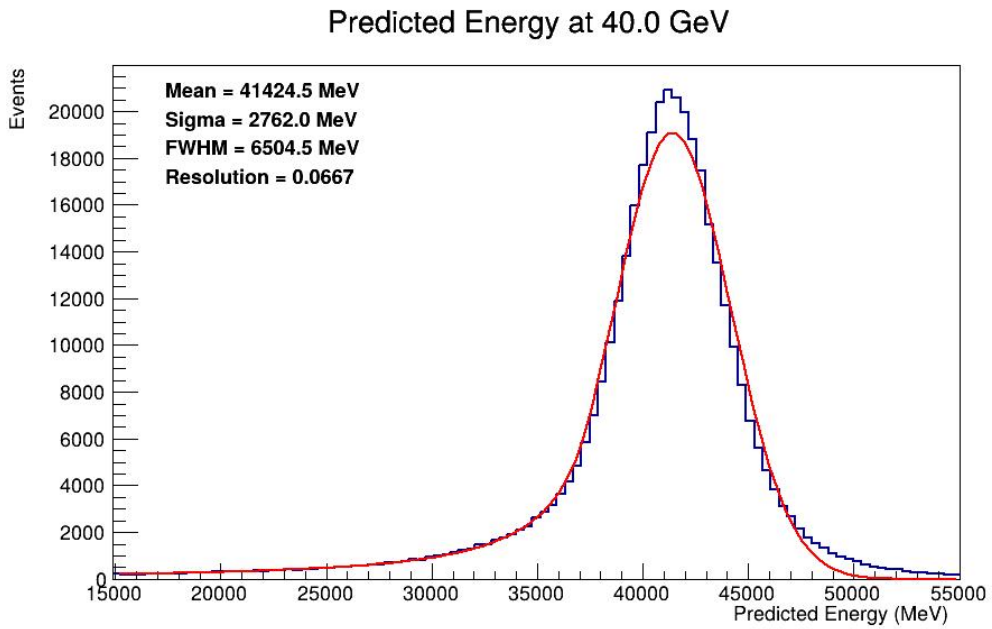
back up

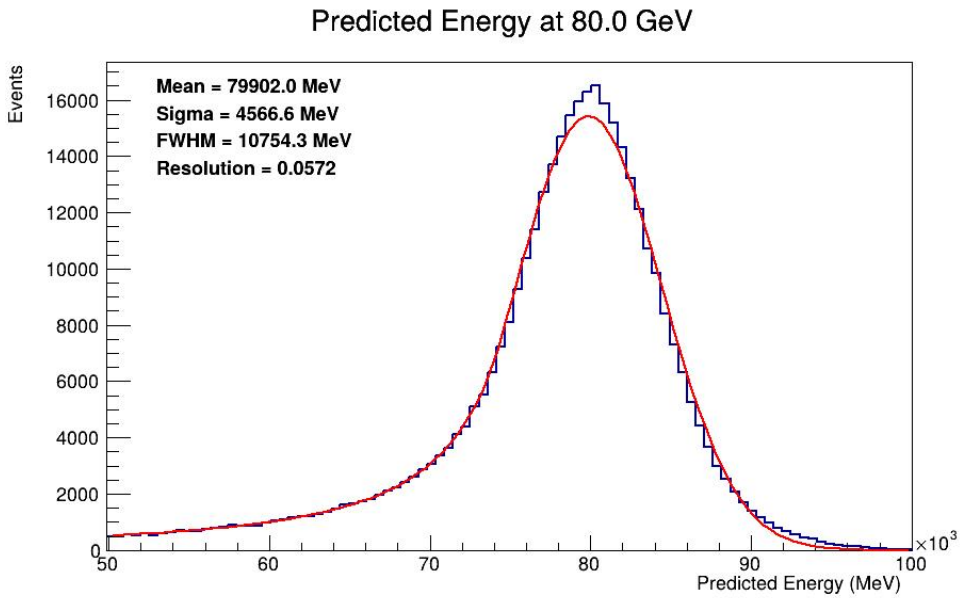
GELU

$$\mathrm{GELU}(x) \approx 0.5x\left(1+\tanh\left[\sqrt{\frac{2}{\pi}}\left(x+0.044715x^3\right)\right]\right)$$

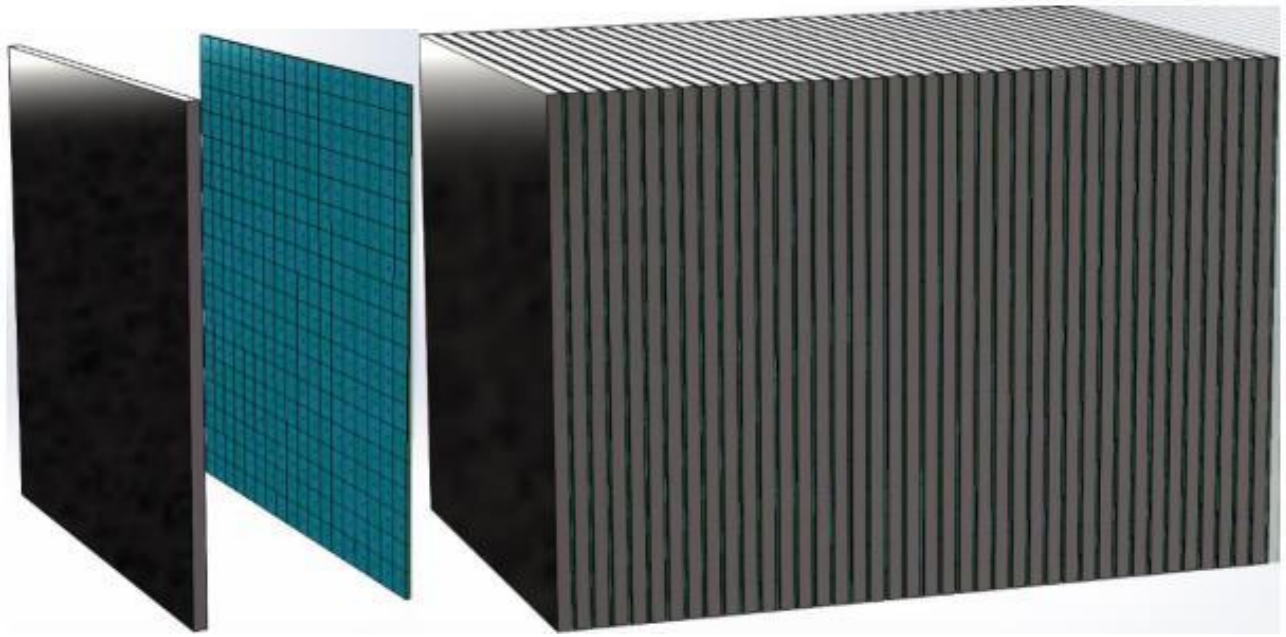


Energy points





0.25cut



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

源：pion，高斯源

范围：1-120GeV

数量：train=6.5e5，test=1.5e5

cell_erengy < 0.25,pass

cell_erengy > 0.25, Include

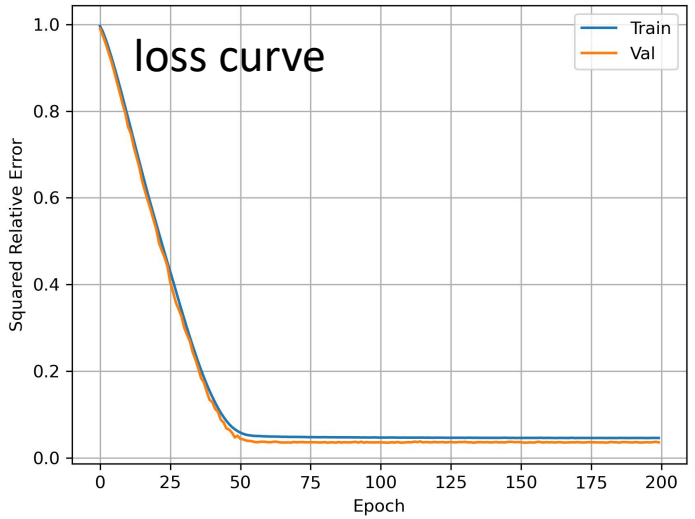
$E_{layer}[40]$

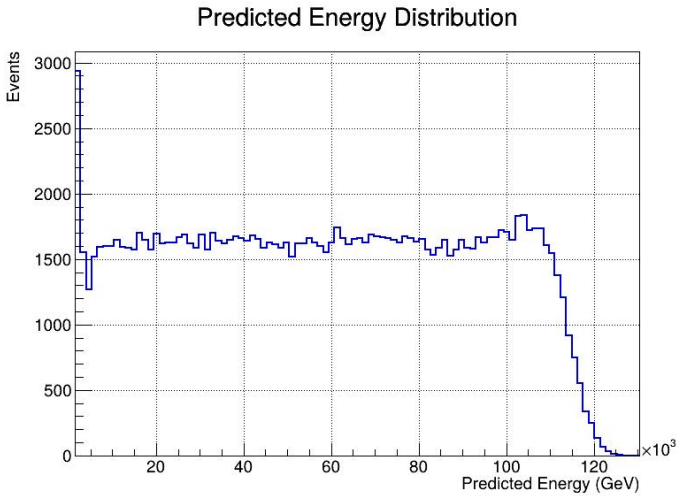
$N_{layer}^{hits}[40]$

$\frac{E}{N}[40]$

$E_{HCAL}[1]$

0.25cut

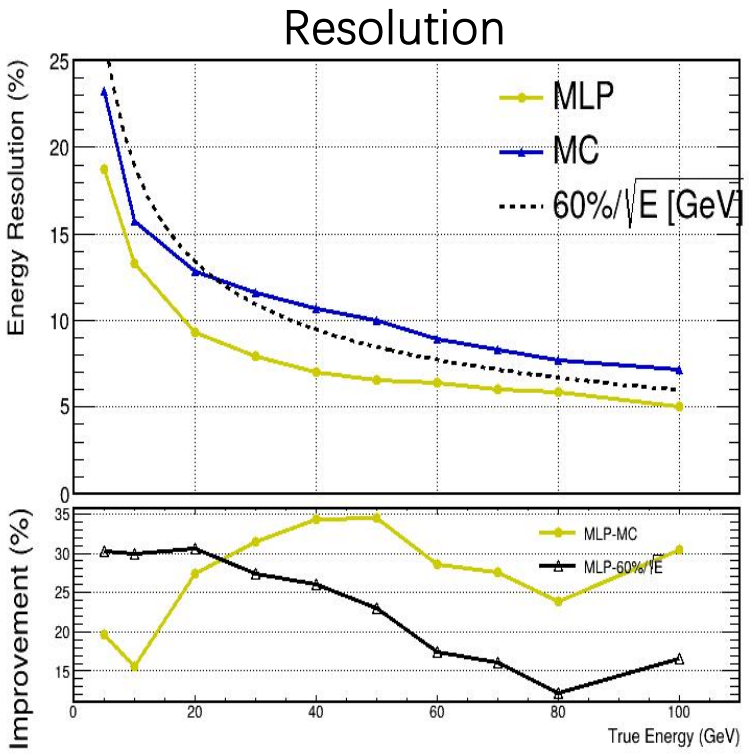

loss curve
Train
Val
Squared Relative Error
Epoch
1.0
0.8
0.6
0.4
0.2
0.0
0
25
50
75
100
125
150
175
200

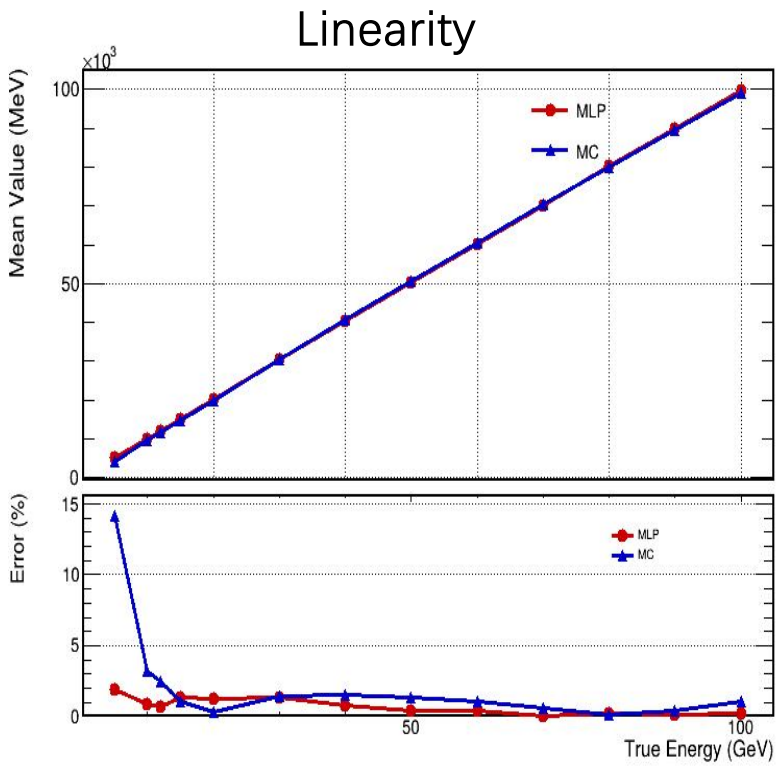

Predicted Energy Distribution
Events
Predicted Energy (GeV)
3000
2500
2000
1500
1000
500
0
20
40
60
80
100
120
×10³

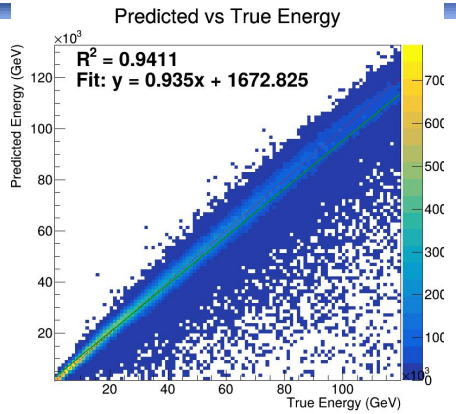
训练结果

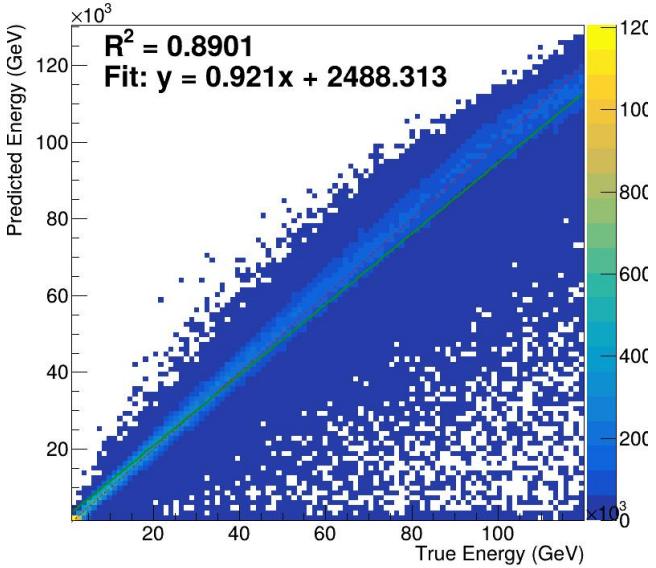
Energy_cut = 0.25 MeV

Event_cut(layer): N_hit>37
Max_hit<5







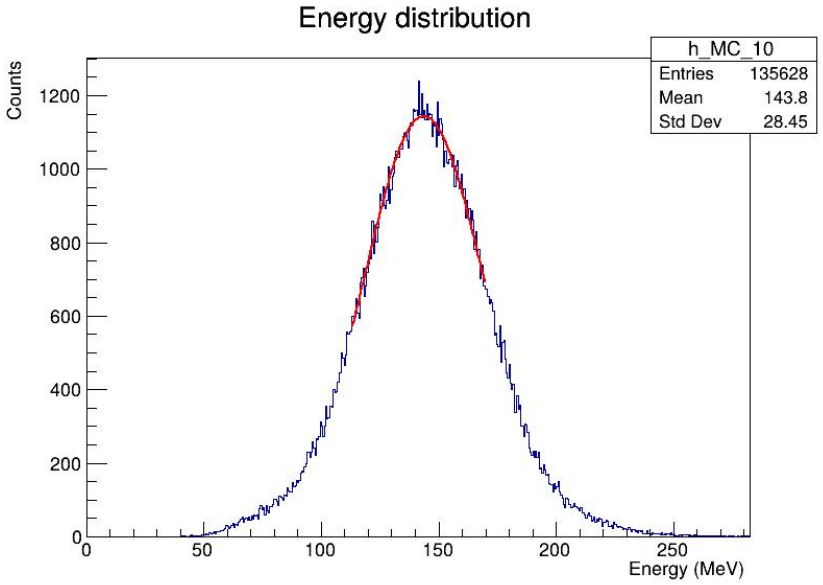


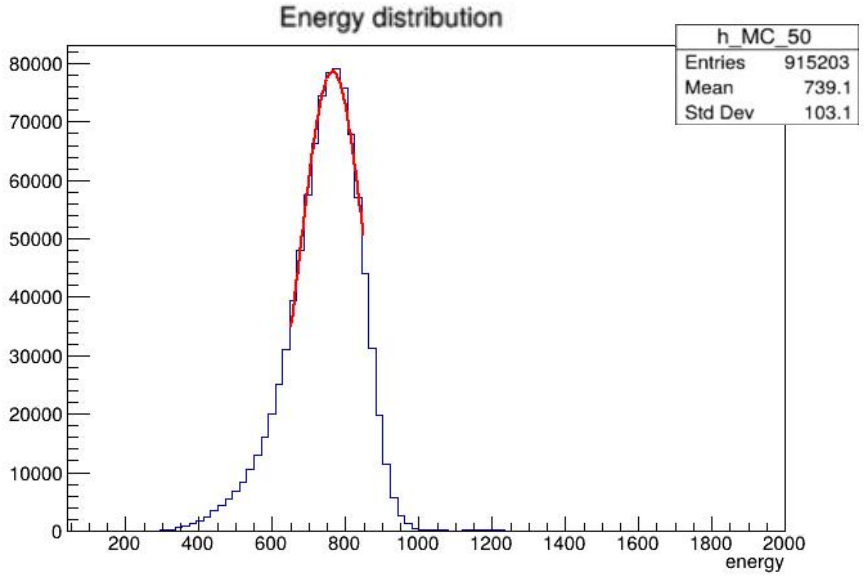
- 分辨率提升显著，平均26%
- 整体能量线性变好，平均0.74%
- 预测结果相关度良好，整体0.921

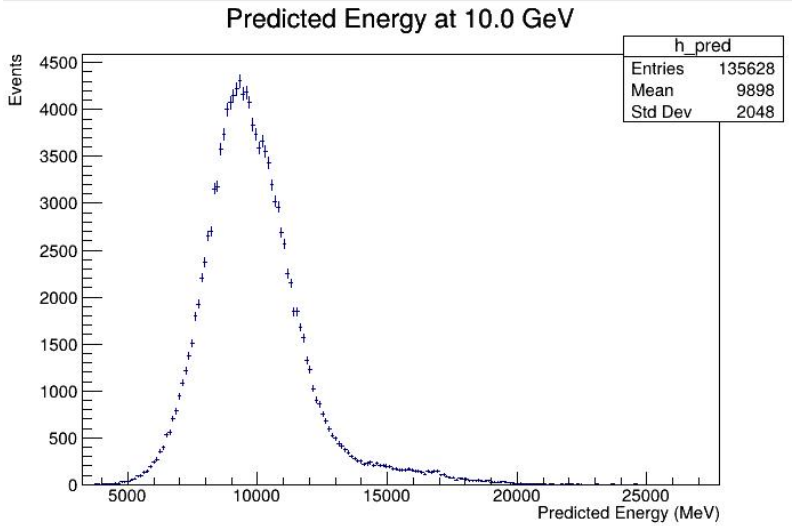
$$Imp = \frac{|R_{MC} - R_{MLP}|}{R_{MC}}$$

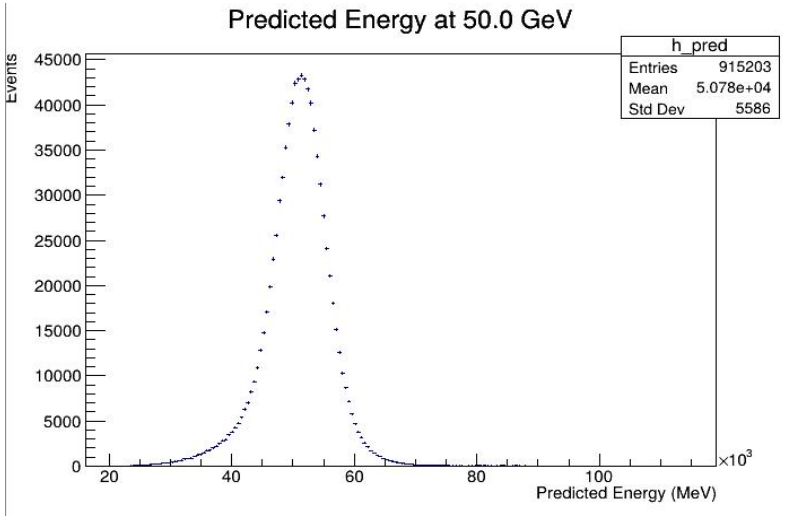
$$Error = \frac{|Mean - True|}{True}$$

data_beam


Energy distribution
h_MC_10
Entries 135628
Mean 143.8
Std Dev 28.45
Counts
Energy (MeV)


Energy distribution
h_MC_50
Entries 915203
Mean 739.1
Std Dev 103.1
energy


Predicted Energy at 10.0 GeV
h_pred
Entries 135628
Mean 9898
Std Dev 2048
Events
Predicted Energy (MeV)


Predicted Energy at 50.0 GeV
h_pred
Entries 915203
Mean 5.078e+04
Std Dev 5586
Events
×10³
Predicted Energy (MeV)

back up

event	energy(GeV)	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					