(Q)ParticleTransformer application in EW coupling of top quark at CEPC

Huayu Liu
liuhy288@sysu2.edu.cn

School of Science
Sun Yat-sen University

October 21, 2024

Outline

- Introduction to the measurement of top quark EW couplings
- Code testing of $t\bar{t}$ jet tagging by ParT/PNet
- Detailed Introduction of Particle Transformer
    - Input data
    - Embedding layer
    - Attention Block layer
    - Class Block layer and output
- Further ParT testing of $t\bar{t}$ event selection by adding variables
- Replace the Linear-layer with Quantum-Blcok

First Section

1.Introduction to the measurement of top quark EW couplings

Project structure

1. **Do the traditional method(cut-base)**
   Select the $t\bar{t}$ case using some traditional variables.
2. **Do the Machine learning method**
   Compared with traditional cut-base ,whether using ML to distinguish $e^+e^- \to t\bar{t}$ events in CEPC from the background with the same final state is better?
3. **Add quantum part to step2**
4. **Comparison and Summary**
   Comprehensive comparison of three methods to evaluate the screening results of machine learning and the model after adding quantum part to ttbar

$t\bar{t}$ pair production

1. **Signal process:**$e^+e^- \rightarrow \gamma/Z \rightarrow t\bar{t}$
2. **Final state:**$l^{\pm}vl^{\mp}\bar{v}b\bar{b}, l^{\pm}vq\bar{q}b\bar{b}$✓$, alljets$
   - t quark related information is largely retained in e/u or b-jet.
   - In the choosen final state, the contribution of MET is much more clear than others due to only one neutrino was produced from the process.
3. **Beam state:** 100% of the particles in a beam have a specific chirality, left-handed electrons and right-handed anti-electrons or right-handed electrons and left-handed anti-electrons
4. **Background:** Any process that have lvqqbb final states.The single top production which is difficult to distinguish ($W^* \rightarrow bt$) and others like: $\mu\bar{\mu}, q\bar{q}, \gamma/Z, WW, ZZ, ZWW, ZZZ$

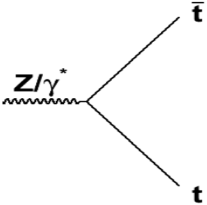


Figure: EW coupling vertax

Observables and Form Factors

- If both y/Z and tt are on-shell, EW coupling can be described by four Form Factors:

$$\Gamma_u^{Vt\bar{t}}(s) = -ie\{\gamma_u[F_{1V}^X(s) + \gamma_5 F_{1A}^X(s)] + \frac{\sigma_{uv}}{2m_t}(q_t + q_{\bar{t}})[i\frac{F_{2V}^X(s)}{2m_t} + \frac{F_{2A}^X(s)}{e}\gamma_5]\}$$

- Obtained σ_L, σ_R(left/right hand cross section) ,θ_L, θ_R(polar angle),$(A_{fb}^t)_L, (A_{fb}^t)_R$(forward backward asymmetry) from the signal process.
  They are the are obeservable which can be measured experimentally.

$$A_{FB}^t = \frac{N(cos\theta > 0) - N(cos\theta < 0)}{N(cos\theta > 0) + N(cos\theta < 0)}$$

- Four Observables($\sigma_L, \sigma_R, (A_{fb}^t)_L, (A_{fb}^t)_R$) can be shown by four Form Factors

Second Section

2.Code testing of $t\bar{t}$ jet tagging by ParT/PNet

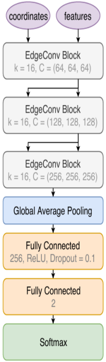
My work

- **My foucs:** Application of machine learning methods in $t\bar{t}$event selection.
- **Current tasks**:
  1.Test the application of an existing ParT/ParNet code for jet tagging at CEPC.
  2.Further ParT testing of $t\bar{t}$ event selection by adding variables.
- **Simple introduction to ParticleNet/ParticleTransformer:**

  1. **ParticleNet** is a model based on GNN architecture that focuses on processing local inter-particle information.

     Features:The interaction relationship between particles within a certain range can be captured through the relationship between points and edges.

     Advantages:Compared with traditional feature engineering and manually defined variables, ParticleNet can automatically learn high-order features of particle injection, improving classification accuracy.

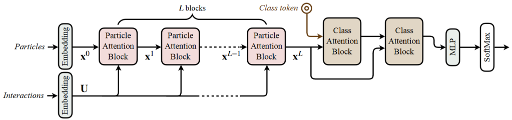


My work

2. **ParticleTransformer** is a deep learning model that combines self-attention mechanism and interaction between particles.

   Features:Ability to flexibly model long-distance dependencies between particles and capture complex relationships between particles and jets.

   Advantages:It can better handle the global dependence of particles, and is more robust in identifying jets and background noise.



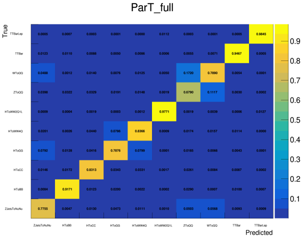
Config of testing(jet tagging)

- **Source**(without Quantum part):https://github.com/jet-universe/particle_transformer.
- **input dataset**
  10 processes containing jets: $H \to b\bar{b}$,$H \to c\bar{c}$,$H \to gg$,$H \to 2W \to 4q$,$H \to 2W \to lvq\bar{q}$,$t \to bq\bar{q}$,$t \to blv$,$W \to q\bar{q}$,$Z \to q\bar{q}$,$Z(j) \to \bar{v}$.(produced by **Shudong**)
  Each process has 100M in the training set, 20M in the test set, and 5M in the verification set.
  Source:/cefs/higgs/wangshudong/data/JetClass/Pythia
- **features:**
  kin: only kinematic variables

$$\Delta\eta, \Delta\phi, logp_T, logE, log\frac{p_T}{p_T(jet)}, log\frac{E}{E(jet)}, \Delta R$$

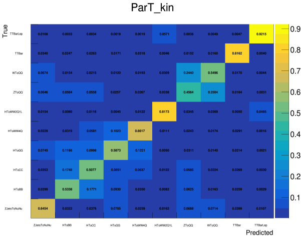
  full : kinematic variables + particle identification + trajectory displacement
- **epochs**:30,**Batch Size**:1024 , **learning rate**:0.01
- **num_heads**:8,**num_lawyer**:8,**num_classlayers**:2
- **activation function**:gelu
- **model**:ParticleNet/ParticleTransformer

Results


ParT_full

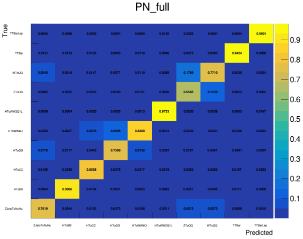
True	ZJetsToNuNu	HToBB	HToCC	HToGG	HToWW4Q	HToWW2Q1L	ZToQQ	WToQQ	TTBar	TTBarLep
TTBarLep	0.0005	0.0007	0.0003	0.0001	0.0000	0.0112	0.0003	0.0001	0.0005	0.9845
TTBar	0.0123	0.0110	0.0088	0.0050	0.0086	0.0006	0.0055	0.0071	0.9467	0.0005
WToQQ	0.0488	0.0012	0.0140	0.0076	0.0125	0.0050	0.1720	0.7890	0.0054	0.0001
ZToQQ	0.0398	0.0322	0.0329	0.0191	0.0148	0.0019	0.6790	0.1117	0.0030	0.0002
HToWW2Q1L	0.0009	0.0004	0.0019	0.0003	0.0012	0.9771	0.0019	0.0039	0.0006	0.0127
HToWW4Q	0.0201	0.0026	0.0440	0.0786	0.8366	0.0009	0.0174	0.0157	0.0114	0.0000
HToGG	0.0792	0.0128	0.0416	0.7876	0.0799	0.0001	0.0185	0.0066	0.0043	0.0001
HToCC	0.0146	0.0172	0.8313	0.0343	0.0331	0.0017	0.0261	0.0084	0.0087	0.0002
HToBB	0.0084	0.9171	0.0123	0.0200	0.0022	0.0002	0.0290	0.0007	0.0100	0.0007
ZJetsToNuNu	0.7755	0.0047	0.0130	0.0473	0.0111	0.0010	0.0503	0.0568	0.0093	0.0009

Predicted


ParT_kin

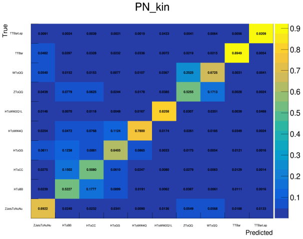
True	ZJetsToNuNu	HToBB	HToCC	HToGG	HToWW4Q	HToWW2Q1L	ZToQQ	WToQQ	TTBar	TTBarLep
TTBarLep	0.0108	0.0033	0.0034	0.0019	0.0019	0.0571	0.0036	0.0049	0.0047	0.9215
TTBar	0.0340	0.0247	0.0263	0.0171	0.0318	0.0046	0.0152	0.0160	0.8162	0.0040
WToQQ	0.0674	0.0154	0.0215	0.0120	0.0193	0.0369	0.2440	0.5496	0.0170	0.0044
ZToQQ	0.0646	0.0584	0.0558	0.0257	0.0267	0.0338	0.4564	0.2504	0.0164	0.0031
HToWW2Q1L	0.0154	0.0080	0.0116	0.0045	0.0122	0.8173	0.0345	0.0369	0.0050	0.0495
HToWW4Q	0.0229	0.0319	0.0581	0.1023	0.6617	0.0111	0.0243	0.0174	0.0291	0.0016
HToGG	0.0745	0.1166	0.0998	0.5873	0.1221	0.0050	0.0311	0.0140	0.0214	0.0021
HToCC	0.0353	0.1748	0.5077	0.0851	0.0637	0.0122	0.0585	0.0224	0.0269	0.0030
HToBB	0.0299	0.5338	0.1771	0.0930	0.0350	0.0082	0.0625	0.0163	0.0259	0.0029
ZJetsToNuNu	0.6454	0.0322	0.0376	0.0705	0.0259	0.0162	0.0688	0.0714	0.0399	0.0107

Predicted


PN_full

True	ZJetsToNuNu	HToBB	HToCC	HToGG	HToWW4Q	HToWW2Q1L	ZToQQ	WToQQ	TTBar	TTBarLep
TTBarLep	0.0005	0.0008	0.0003	0.0001	0.0000	0.0136	0.0003	0.0001	0.0004	0.9801
TTBar	0.0151	0.0152	0.0135	0.0065	0.0118	0.0008	0.0075	0.0083	0.9424	0.0008
WToQQ	0.0540	0.0015	0.0147	0.0077	0.0119	0.0055	0.1789	0.7710	0.0039	0.0001
ZToQQ	0.0406	0.0322	0.0332	0.0203	0.0137	0.0020	0.6545	0.1228	0.0023	0.0002
HToWW2Q1L	0.0009	0.0004	0.0020	0.0003	0.0013	0.9733	0.0020	0.0038	0.0005	0.0167
HToWW4Q	0.0250	0.0037	0.0578	0.0996	0.8408	0.0014	0.0229	0.0201	0.0150	0.0001
HToGG	0.0778	0.0117	0.0443	0.7598	0.0756	0.0001	0.0197	0.0067	0.0041	0.0001
HToCC	0.0159	0.0236	0.8036	0.0378	0.0317	0.0019	0.0292	0.0092	0.0107	0.0002
HToBB	0.0084	0.9066	0.0154	0.0207	0.0025	0.0003	0.0337	0.0006	0.0117	0.0008
ZJetsToNuNu	0.7619	0.0044	0.0152	0.0472	0.0106	0.0011	0.0513	0.0573	0.0089	0.0010

Predicted


PN_kin

True	ZJetsToNuNu	HToBB	HToCC	HToGG	HToWW4Q	HToWW2Q1L	ZToQQ	WToQQ	TTBar	TTBarLep
TTBarLep	0.0091	0.0024	0.0039	0.0021	0.0019	0.0433	0.0041	0.0064	0.0056	0.9209
TTBar	0.0482	0.0297	0.0328	0.0232	0.0336	0.0072	0.0219	0.0215	0.8949	0.0054
WToQQ	0.0540	0.0152	0.0153	0.0077	0.0107	0.0367	0.2525	0.6725	0.0031	0.0041
ZToQQ	0.0439	0.0779	0.0625	0.0244	0.0178	0.0385	0.5255	0.1713	0.0028	0.0024
HToWW2Q1L	0.0146	0.0070	0.0118	0.0048	0.0167	0.8258	0.0307	0.0351	0.0039	0.0469
HToWW4Q	0.0254	0.0472	0.0768	0.1124	0.7800	0.0174	0.0261	0.0185	0.0349	0.0024
HToGG	0.0611	0.1238	0.0881	0.6405	0.0865	0.0033	0.0175	0.0054	0.0121	0.0016
HToCC	0.0275	0.1502	0.5080	0.0610	0.0247	0.0080	0.0279	0.0063	0.0129	0.0014
HToBB	0.0239	0.5227	0.1777	0.0899	0.0191	0.0062	0.0387	0.0061	0.0111	0.0016
ZJetsToNuNu	0.6922	0.0240	0.0232	0.0341	0.0090	0.0136	0.0549	0.0568	0.0188	0.0133

Predicted

Results

		ParT_full	ParT_kin	PN_full	PN_kin
CEPC	$\overline{AUC}$	>0.95	>0.95	>0.95	>0.95
	$Accuracy$	0.85246	0.74387	0.83939	0.6983
JetClass dataset	$\overline{AUC}$	0.9877		0.9849	
	Accuracy	0.861		0.844	

- $\overline{AUC}$ is the average of AUC values between each two processes.

$$\overline{AUC} = \frac{\sum_{i=1}^{10} \sum_{j=1}^{11-i} AUC_{ij}}{\sum_{i=1}^{10} i}$$

- $Accuracy$ is the accuracy rate of prediction for the entire test dataset

- There is no significant difference compared to the results in the paper, and there is no significant difference between PN and ParT.
- The results using all features are better than using only kin features.

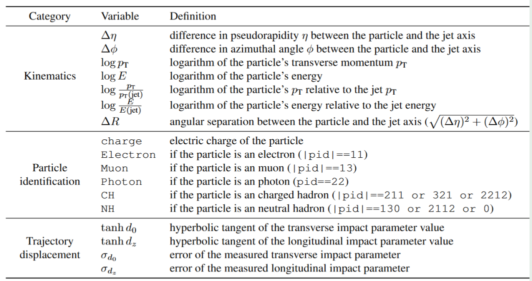
Third Section-First Subsection

- 3.Introduction of Particle Transformer
  - Input data

Input

Particle information x

- x is a tensor of size [B,F,S] describing the particles themselves.
  1. **B:** The **batch size** which means how many processes are input at one time?
  2. **F:** means number of features that how many variables describe a particle?
  3. **S:** The number of particles in the process.



Category	Variable	Definition
Kinematics	$\Delta\eta$	difference in pseudorapidity η between the particle and the jet axis
	$\Delta\phi$	difference in azimuthal angle ϕ between the particle and the jet axis
	$\log p_T$	logarithm of the particle's transverse momentum p_T
	$\log E$	logarithm of the particle's energy
	$\log \frac{p_T}{p_T(\text{jet})}$	logarithm of the particle's p_T relative to the jet p_T
	$\log \frac{E}{E(\text{jet})}$	logarithm of the particle's energy relative to the jet energy
	ΔR	angular separation between the particle and the jet axis ($\sqrt{(\Delta\eta)^2 + (\Delta\phi)^2}$)
Particle identification	`charge`	electric charge of the particle
	`Electron`	if the particle is an electron (`|pid|==11`)
	`Muon`	if the particle is an muon (`|pid|==13`)
	`Photon`	if the particle is an photon (`pid==22`)
	`CH`	if the particle is an charged hadron (`|pid|==211 or 321 or 2212`)
	`NH`	if the particle is an neutral hadron (`|pid|==130 or 2112 or 0`)
Trajectory displacement	$\tanh d_0$	hyperbolic tangent of the transverse impact parameter value
	$\tanh d_z$	hyperbolic tangent of the longitudinal impact parameter value
	σ_{d_0}	error of the measured transverse impact parameter
	σ_{d_z}	error of the measured longitudinal impact parameter

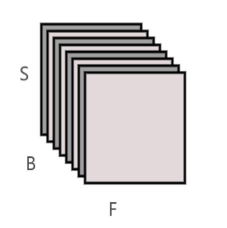


Figure: Example Features

Input

Interaction between particles U

- For one kind of feature,U is a tensor of size [B,C,S,S] which consists of the such following two parts:
- C is the number of the features added to describe the interaction .

The default part which is calculated from p_T of two particles :

$$\Delta_{ij} = ln\sqrt{(\frac{1}{2}ln(1+\frac{2*p_{z_i}}{E-p_{z_i}}) - \frac{1}{2}ln(1+\frac{2*p_{z_j}}{E-p_{z_j}}))^2 + (arctan(\frac{p_{y_i}}{p_{x_i}})^2 - arctan(\frac{p_{y_j}}{p_{x_j}})^2)},$$

$$z_{ij} = \frac{min(p_{T_i}, p_{T_j})}{p_{T_i}+p_{T_j}},$$

$$k_{Tij} = min(p_{T_i}, p_{T_j}) * \delta,$$

$$m_{ij}^2 = (E_i + E_j)^2 - (p_{T_i} + p_{T_j})^2$$

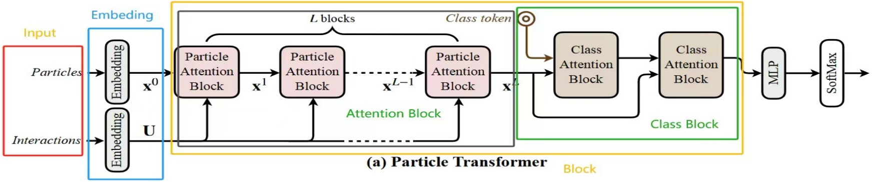
It can be concatted by artificially added extra matrix uu(size； [C_u,S,S]) which is empty by default.

$$C = C_u + 4$$

Third Section-Second Subsection

- 3.Introduction of Particle Transformer
  - Input data
  - Embedding layer

Embeding layer



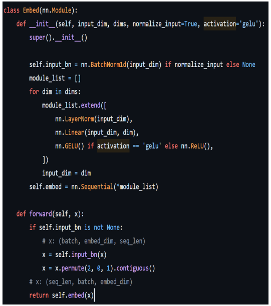
(a) Particle Transformer

- Through several MLP(Multilayer Perceptron), the feature dimensions of x and U are sequentially raised to the dimensions specified by the *dims* array.

  $x : [B, F, S] \xrightarrow{MLP} [B, 128, S] \xrightarrow{MLP} [B, 512, S] \xrightarrow{MLP}$
  $[B, 128, S] \xrightarrow{Output}$

  $U : [B, C, S, S] \xrightarrow{MLP} [B, 64, S, S] \xrightarrow{MLP} [B, 64, S, S] \xrightarrow{MLP}$
  $[B, 8, S, S] \xrightarrow{Output}$

* All numbers are parameters that can be changed.



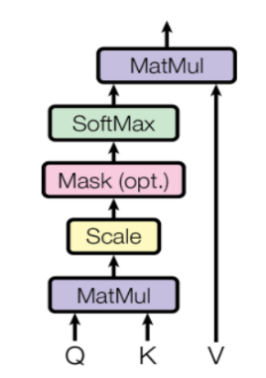
```
class Embed(nn.Module):
    def __init__(self, input_dim, dims, normalize_input=True, activation='gelu'):
        super().__init__()

        self.input_bn = nn.BatchNorm1d(input_dim) if normalize_input else None
        module_list = []
        for dim in dims:
            module_list.extend([
                nn.LayerNorm(input_dim),
                nn.Linear(input_dim, dim),
                nn.GELU() if activation == 'gelu' else nn.ReLU(),
            ])
            input_dim = dim
        self.embed = nn.Sequential(*module_list)

    def forward(self, x):
        if self.input_bn is not None:
            # x: (batch, embed_dim, seq_len)
            x = self.input_bn(x)
            x = x.permute(2, 0, 1).contiguous()
        # x: (seq_len, batch, embed_dim)
        return self.embed(x)
```

Third Section-Third Subsection

- 3.Introduction of Particle Transformer
  - Input data
  - Embedding layer
  - Attention Block layer

Attention Block

Self-Attention (Single Head)



- Calculation of Q(Query), K(Key), V(Value) tensor:

$$\begin{cases} Q = W_q x_n + b_q & \text{size: [B, } N_1\text{, } S\text{]} \\ K = W_k x_n + b_k & \text{size: [B, } N_1\text{, } S\text{]} \\ V = W_v x_n + b_v & \text{size: [B, } N_2\text{, } S\text{]} \end{cases}$$

- Output of *MatMul*:

$$\text{SoftMax}\left(\frac{QK^T}{\sqrt{N_1}}\right) V$$

Size: [B, S, S] * [B, N_2, S] = [B, N_2, S]

- Pass Linear layer:

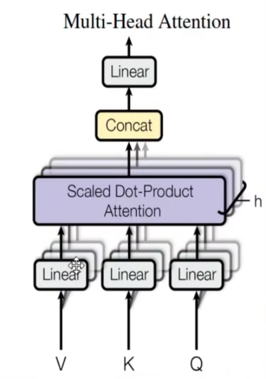
$$x_{n+1} = \text{Linear}\left(\text{SoftMax}\left(\frac{QK^T}{\sqrt{N_1}}\right) V\right)$$

to change dim from N_2 to 128.

Attention Block (head_num=8)

Multi-Head

- Divide x into 8 parts on the feature dimension: $x_{n1}, x_{n2}, \dots, x_{n8}$
- Size of x_{ni}: $[B, \frac{128}{8}, N]$



- Calculation of Q (Query), K (Key), V (Value) tensors:

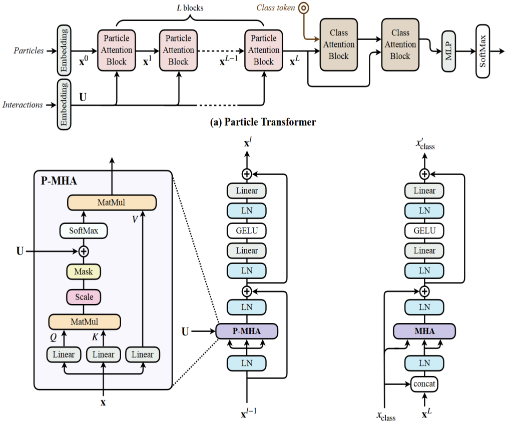
$$\begin{cases} Q_i = W_{qi}x_{ni} + b_{qi} & \text{Size: } [B, N_1, S, 8] \\ K_i = W_{ki}x_{ni} + b_{ki} & \text{Size: } [B, N_1, S, 8] \\ V_i = W_{vi}x_{ni} + b_{vi} & \text{Size: } [B, N_2, S, 8] \end{cases} \tag{1}$$

- Output of *Concat*:

$$\sum_{i=1}^{8} \text{SoftMax}\left(\frac{Q_i K_i^T}{\sqrt{N_1}}\right) V_i \oplus$$

Size: $[B, 8 \times V, S]$

Particle Attention Block



(a) Particle Transformer

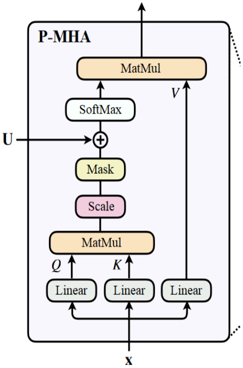
(b) Particle Attention Block

(c) Class Attention Block

- Here is the complete structure of ParticleTransformer,It can be divided into three modules except the embedding module:
  - Particle Attention Block
  - Class Block
  - Score Calculation
- Here x_{clsaa} is equal to Class token .size[B,128,1]

Particle Attention Block

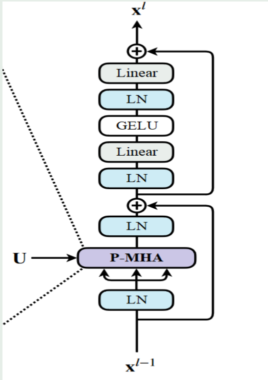
- The 8 in size of U martix [B,8,S,S] is required to be equal to head_num.Each head will be allocated a U_i(Size:(B,S,S).



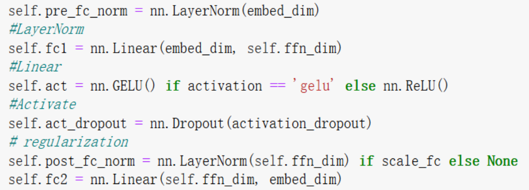
* Muti-Head Attention(MHA) Output of $Concat: \sum_{i=1}^{8} SoftMax(\frac{Q_i K_i^T}{\sqrt{N_1}}) V_i \oplus$
* Particle-Muti-Head Attention(P-MHA) Output of $Concat: \sum_{i=1}^{8} SoftMax(\frac{Q_i K_i^T}{\sqrt{N_1}} + U_i) V_i \oplus$
* The size of each part is as same as MHA.

Particle Attention Block

Residual Connection Part



1. **First Residual Connection:**Set the output of P-MHA as $\tilde{x_n}$,and put $x_n + \tilde{x_n}$ into Residual Connection Part.
2. **Second Residual Connection:**After LN(layer norm),Linear layer,Activate function,LN,Linear get $\widetilde{x_n + \tilde{x_n}}$.



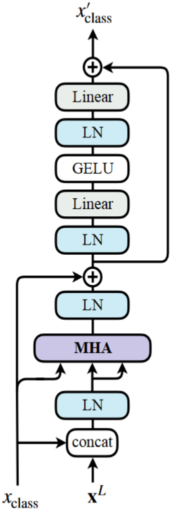
```
self.pre_fc_norm = nn.LayerNorm(embed_dim)
#LayerNorm
self.fc1 = nn.Linear(embed_dim, self.ffn_dim)
#Linear
self.act = nn.GELU() if activation == 'gelu' else nn.ReLU()
#Activate
self.act_dropout = nn.Dropout(activation_dropout)
# regularization
self.post_fc_norm = nn.LayerNorm(self.ffn_dim) if scale_fc else None
self.fc2 = nn.Linear(self.ffn_dim, embed_dim)
```

3. **Final output of Block** :$x_{n+1} = x_n + \tilde{x_n} + \widetilde{x_n + \tilde{x_n}}$

Third Section-Forth Subsection

- 3.Introduction of Particle Transformer
  - Input data
  - Embedding layer
  - Attention Block layer
  - Class Block layer and output

Class Block



Compare with MHA, the differences of Class Block are;

1. Set $z = (x^{class}, x)$ its size [B,128,S+1]

$$\begin{cases} Q_i = W_{qi} x_{ni}^{class} + b_{qi} \quad Q \quad \text{size:}[B, N_1, S, 8] \\ K_i = W_{ki} z_{ni} + b_{ki} \quad K \quad \text{size:}[B, N_1, S, 8] \\ V_i = W_{vi} z_{ni} + b_{vi} \quad V \quad \text{size:}[B, N_2, S, 8] \end{cases} \tag{2}$$

2. At the first Residual Connection Part replace x_n to x_{class}
3. Output Size: [B,128,1]

After a MLP (Linear layer) the size is from [B,128,1] to [B,1,1] which is equivalent to [B,1].

Then pass the last part an activate it becomes the score of each jet/process

Input Data

4.Further ParT testing of $t\bar{t}$ event selection by adding variables

Forth Section

Data: $t\bar{t}$ and SingleTop event level data (sim & reco) from **MUSTAPHA**.

Input Features

Category	Variable	Definition
Particles	$\delta\eta$	difference in pseudorapidity η between the particle and the overall-event axis
	$\delta\phi$	difference in azimuthal angle ϕ between the particle and the overall-event axis
	$\log p_T$	logarithm of the particle’ s transverse momentum p_T
	$\log E$	logarithm of the particle’ s energy E
	$\log \frac{p_T}{p_T(\text{total})}$	logarithm of the particle’ s relative to overall-event P_T(total)
	$\log \frac{E}{E(\text{total})}$	logarithm of the particle’ s energy relative to the total energy E(total)
	δR	angular separation between the particle and the overall-event axis $\sqrt{(\delta\eta)^2 + (\delta\phi)^2}$
Event	nJet	number of jets
	ntaus	number of τ
	nElec	number of electron
	nMuon	number of μ
	nGamma	number of photon
	Emax	the max energy of one object
	MET	Missing Transverse Energy
	TotalEnergyT	The totalenergy of all components in the events

- Particle features are added by default, while the event features are selected based on significant differences in the distributions of the two types of events as new added features.
- Since the input can only be processed along the particle count dimension, the event features are expanded to match the particle count dimension by replication and concatenated with the particle features for input.

Over-fitting Result

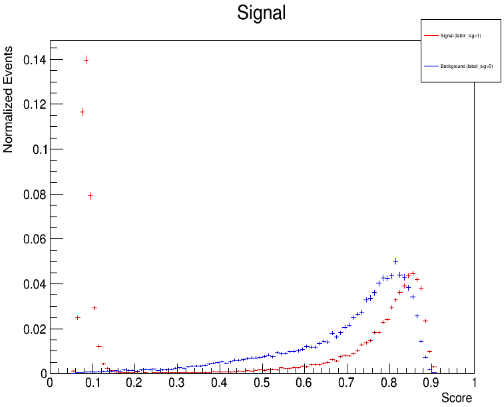
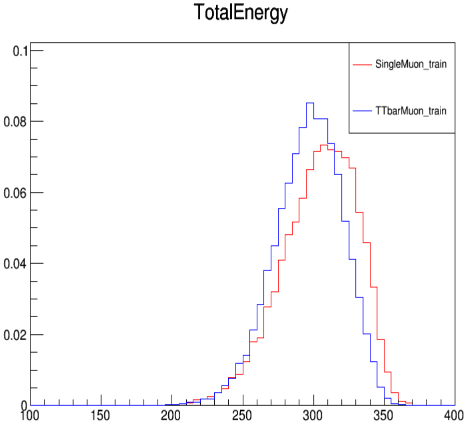


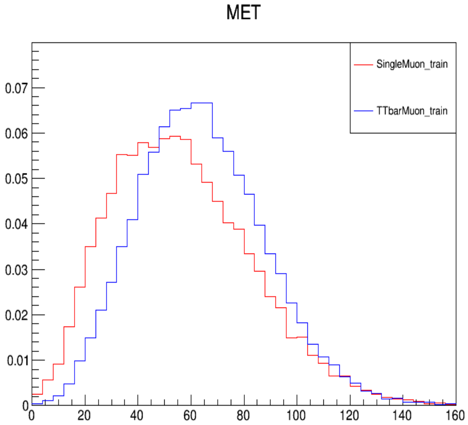
Figure: Signal Score

- Last time, we saw a strange peak on the right side in the test set results when the epoch was set to 15.
- The reason is that there is no electron-channel sample in the SingleTop test set.
- So I do the another ParT model testing on the only muon channel samples.

Features Distribution(Muon Channel)

- Total Energy and Missing Transverse Energy





Features Distribution(Muon Channel)

- NGamma and E_{max}

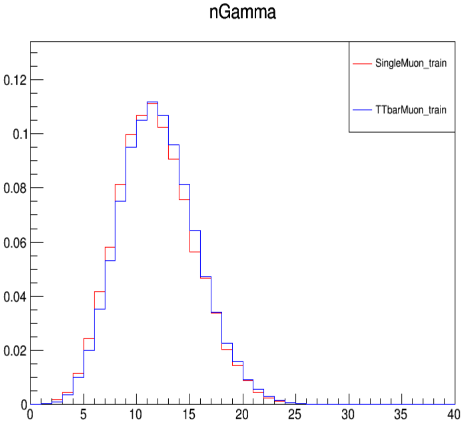


Figure: NGamma

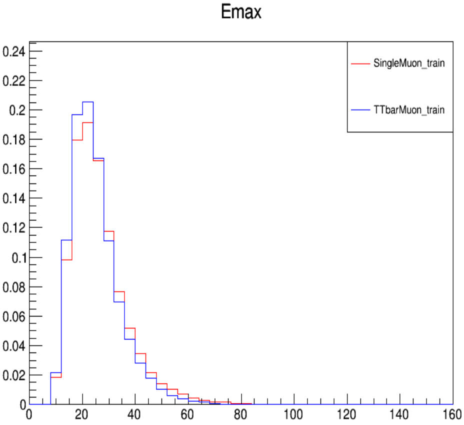


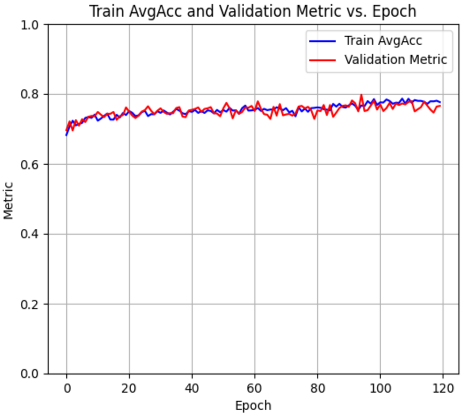
Figure: E_{max}

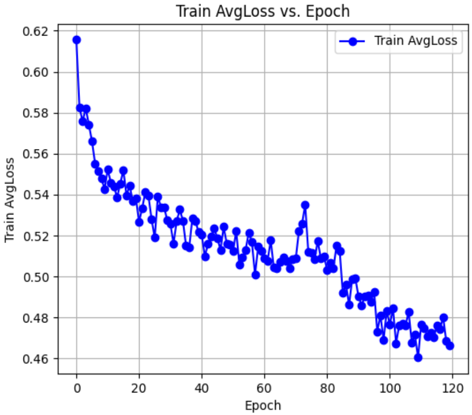
Lepton Channel

- Lepton is from lepton decay of W bosons and the branching ratios of the electron and muon channels are essentially the same theoretically.
- Becasue of

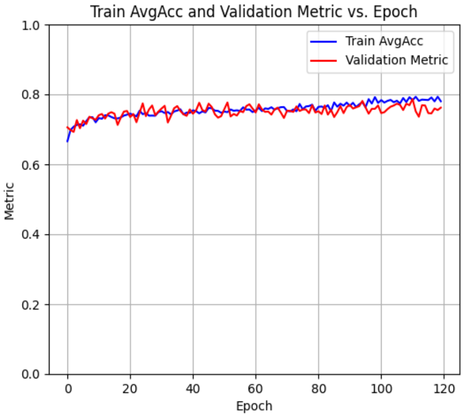
Epoch Set

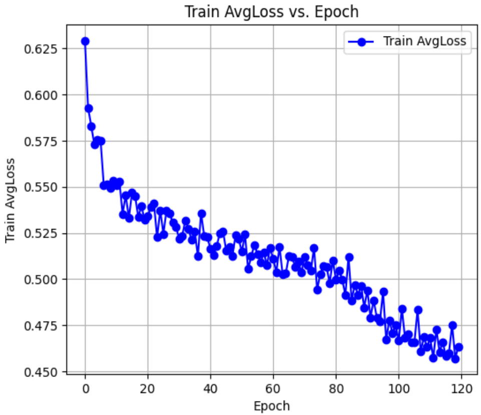
- Plot the Average Accuracy(means Correct classification rate) and loss(cross entropy los)to find out the best epoch set.





Epoch Set





- For the kin-model,the suitable epoch is 69.
- For the full-model the suitable epoch is 61.

Results

- Signal score:
- The number in each bin has been regularized by removing the total number of itself.

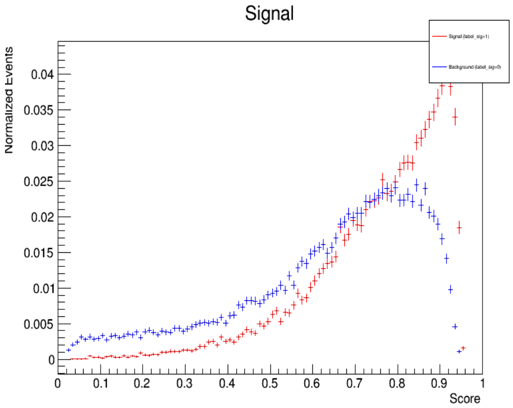


Figure: Just Particle features

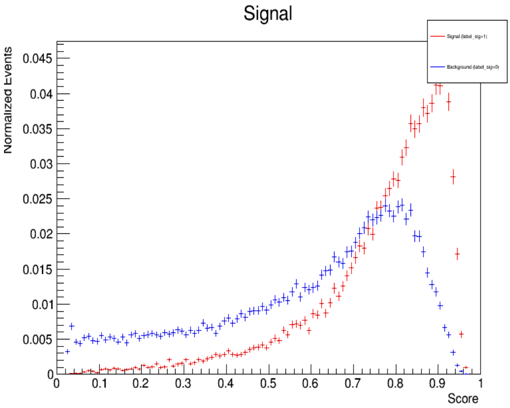


Figure: Particle+Event features

Results

- Migration Matrix

	Sig	Bkg
Sig	0.92755	0.07245
Bkg	0.78007	0.21992

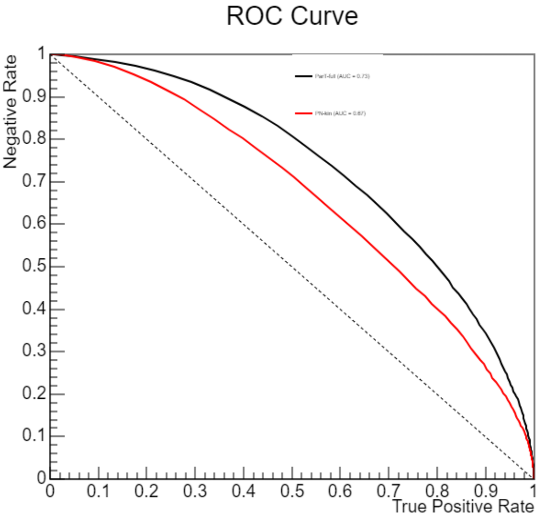
Just Particle features

	Sig	Bkg
Sig	0.92140	0.078600
Bkg	0.69981	0.300189

Particle+Event features

- Suppose there are only these two events,Although the newly added variables do have an effect, they effectively serve as a stricter cut. Generally speaking, the cross-section for TTbar is about 6 to 7 times that of SingleTop. Even though the new model's selection results in fewer events being identified as signals, the proportion of SingleTop background has been significantly reduced.
  - **Just Particle features**: signal rate$= \frac{0.927548}{0.927548+0.15*0.78007} = 0.88798$
  - **Particle+Event features**: signal rate$= \frac{0.9214}{0.9214+0.15*0.69981} = 0.89772$

ROC


ROC Curve
Negative Rate
True Positive Rate
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1
ParT-full (AUC = 0.73)
PN-kin (AUC = 0.67)

Results and forward

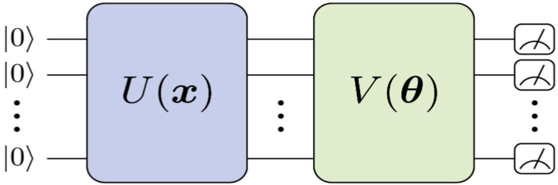
- Due to the mass difference between the electron and muon, I believe the two channels should be discussed separately. We only need to achieve a higher signal branching ratio in one of the channels.
- Now the test of Muon Channel without value cutting is done.
- The signal branching ratio is 0.88798 of kin-group and 0.89772 of full-group.
- From the distributions of TotalEnergy and MET ,there is some differences between SingleTop and $t\bar{t}$ so I think we can do a cutting before input.(For exampke Muon Channel:$250GeV < E_{Total} < 340GeV, MET < 100GeV$).
- The signal and background do not show significant differences in at least these two variables within the filtered interval, which has a negative effect on model training. Moreover, after applying the cuts, we can still retain a sufficient number of events.
- In my opinion ,I think electron channel is better because less mass means higher momentum means better.(This week I will do this by the new data)

Fifth Section

5. Replace the Linear-layer with Quantum-Blcok

Variational Quantum Classifier

- U(x):A series of quantum gates applied to the input x which is used to embed classical data into the quantum circuit.
- V(θ):A series of quantum gates with trainable parameters, along with control gates, is used to construct new output variables.



The VQC takes n input features and outputs no more than n output variables; the output variables are the final expectations of each qubit.

Quantum Gate

A state of qubit can be shown as $|\psi >= cos\theta|0 > +e^{i\phi}sin\theta|1 >$ in bloch sphere.And It can be operated by series quantum gate.

- H:Transform qubits from a pure state to a superposition state.

$$H|0 >= \frac{\sqrt{2}}{2}|0 > +\frac{\sqrt{2}}{2}|1 >, H|1 >= \frac{\sqrt{2}}{2}|0 > -\frac{\sqrt{2}}{2}|1 >$$

- Control Gate(CX):Used to allow one qubit to control another qubit, when combined with the H gate, it can create an entangled state.

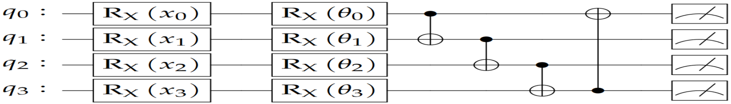
$$CX(a_{00}|00 > +a_{01}|01 > +a_{10}|10 > +a_{11}|11 >) = a_{00}|00 > +a_{01}|01 > +a_{10}|11 > +a_{11}|10 >$$

- Universal Rotation Gate (U): Rotate qubit.θ, ϕ, λ are the angles on x, y, and z axes.

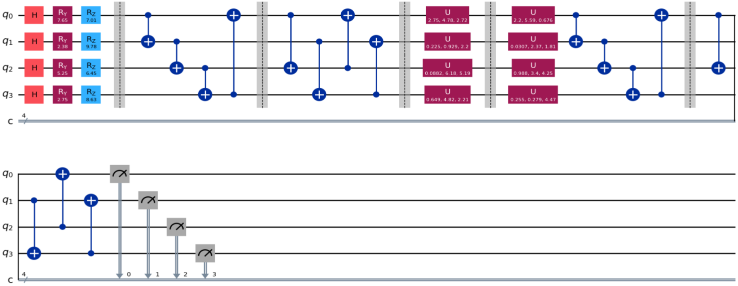
$$U(\theta, \phi, \lambda) = \begin{pmatrix} cos(\frac{\theta}{2}) & e^{-i\lambda}sin(\frac{\theta}{2}) \\ e^{i\phi}sin(\frac{\theta}{2}) & e^{i(\phi+\lambda)}cos(\frac{\theta}{2}) \end{pmatrix}$$

Circuit

- In the paper arXiv:2405.10284v1,its circuit is designed as:

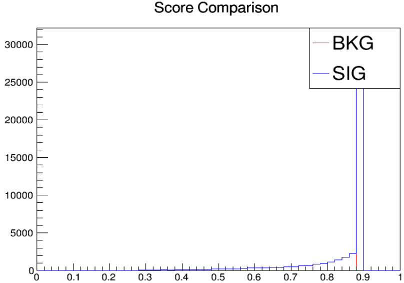


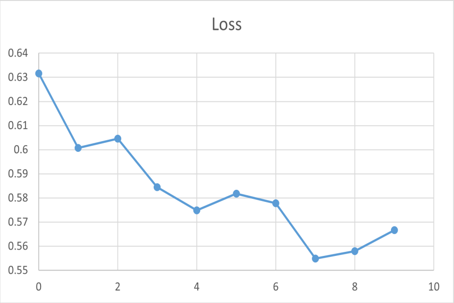
- Inspired by an RNN quantum circuit, the circuit I am using is as follows:



Result

- ROC score:0.3....





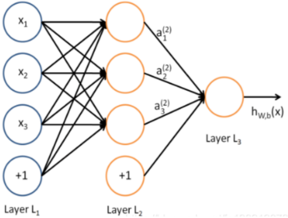
- Due to the simulation of quantum circuits on classical computers, the required resources grow exponentially with the number of input variables, resulting in the following limitations:
  - batch_size: $4/8/16$ (can not be much big)
  - input size: 8(muti-head embed size)

Summary

1. Use ParT/ParNet to do jet tagging on CEPC MC data. It is verified that ParNet and ParT can be used for subsequent research.
   - The best b-jet identification using traditional methods only has an accuracy of 0.7.
   - It is verified that the two models of PN/ParT are equally effective when applied to CEPC data.
2. Apply the model for $t\bar{t}$ preliminary selection.
3. **Future work plan**:
   - Try to couple charges and add them to the U matrix to describe the electromagnetic interaction between particles.
   - Generate more types of backgrounds ($\mu\bar{\mu}, q\bar{q}, \gamma/Z, WW, ZZ, ZWW, ZZZ$) to verify model optimization results.
   - The current data size is approximately in the range of 50k, and we are trying to obtain more data to increase the stability of the model.

Backup

- Code source link: weaver.ParticleTransformer
- Ref paper:
  1. arXiv:2202.03772v3 [hep-ph] 29 Jan 2024
  2. arXiv:1706.03762v7 [cs.CL] 2 Aug 2023
  3. arXiv:1307.8102v1 [hep-ex] 30 Jul 2013
- Activate function:$GELU(x) \approx 0.5x[1 + tanh(\sqrt{\frac{2}{\pi}}(x + 0.047715x^3))]$
- MLP/linear:


x_1
x_2
x_3
+1
+1
$a_1^{(2)}$
$a_2^{(2)}$
$a_3^{(2)}$
$h_{W,b}(x)$
Layer L_1
Layer L_2
Layer L_3