Permutationless Event Reconstruction with Symmetry Preserving Attention Networks

Based on arXiv:2010.09206 arXiv:2106.03898, and work in progress

Public codebase: GitHub

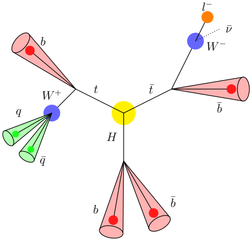
Michael Fenton, Alexander Shmakov, David Ho, Hideki Okawa, Yuji Lee, Ko-Yang Hsiao, Shih-Chieh Hsu, Daniel Whiteson, Pierre Baldi

University of California, Irvine
m.fenton@cern.ch

Top2022, Durham UK
September 6, 2022

Top Reconstruction & Attention

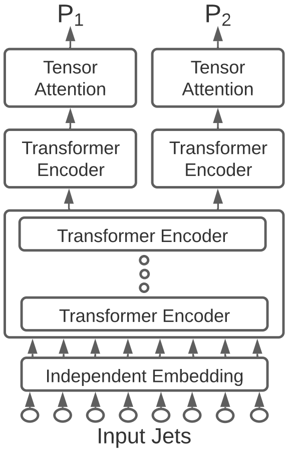
- A key stage in most top analyses is the reconstruction of the top quarks from its decay products: "jet-parton assignment"
- Typically, test every possible permutation of jets, and take the "best" one
  - KLFitter, χ^2 fits, even BDTs/DNNs



- Most state of the art NLP methods now use **attention** mechanisms Vaswani et al, superceding eg RNNs and LSTMs
  - Permutation invariant
  - Handles variable-length lists
- *Tensor Attention*: generalisation of attention to encode symmetries in particle resonstruction
  - eg $q \leftrightarrow \bar{q}$ in W decays or $t \leftrightarrow \bar{t}$ for $t\bar{t}$ events
- Allows to test every possible permutation in a single NN pass $\rightarrow$ no more time-consuming O(10-1000) per-event permutation calculations

Symmetry Preserving Attention Networks: SPA-NET

- Output: Softmax per final state particle

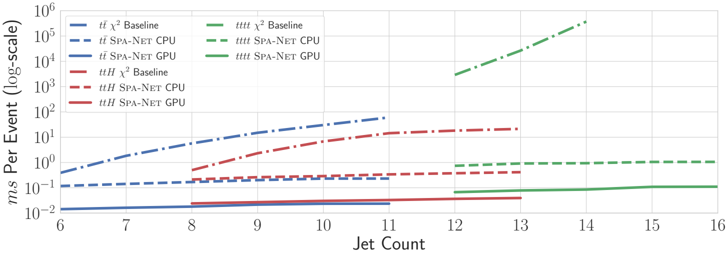


- Input: **unordered** list of jets (4vec+btag)

- ΔR matching between jets and partons for truth definition
- Loss function defined only for particles fully truth matched and masked if not; allows training on "partial events"
  - Reduces the number of generated events required for training
  - Improves performance on events in which one or more particles fall out of fiducial volume
- Can include ALL jets in every event with no CPU penalty
- Trivially can include other jet variables in training eg q/g tagger, JSS, charm tagging...

All-Hadronic Results

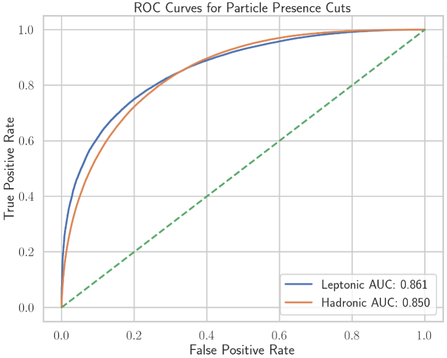
- $t\bar{t}$:
  - SPA-NET: 63.3% full event reconstruction (80.3-52.1% in NJets)
  - χ^2: 42.6% (66.4-28.1 in NJets)
- $t\bar{t}H, H \rightarrow b\bar{b}$:
  - SPA-NET: 38.3% (53.2-30.6%)
  - χ^2: 1.6% (4.0-0.4%)
- 4-top:
  - SPA-NET: 19.1% (35.0-14.9%)
  - χ^2: Intractable!



- No permutations $\rightarrow$ orders of magnitude speedup, & flat vs NJets!

WIP: Lepton+Jets Final States

- One-hot-encode different objects (jets, leptons, etc)
- Add event level variables (MET etc)
- $t\bar{t}$:
  - SPA-NET: 75.6% full event reconstruction (85.5-59.8% vs NJets)
  - KLFitter: 52.1% (77.2-23.7%)
  - Permutation DNN: 64.9% (80.3-48.8%) ▸ Erdmann et al
- $t\bar{t}H, H \to b\bar{b}$
  - SPA-NET: 54.2% in 6j, 42.6% in 7j
  - KLFitter: 31.4% in 6j, 17.7% in 7j

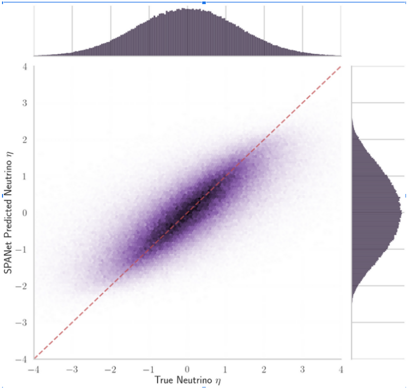


- AUC of 85% for "presence" output; ie, does the network think this particle has a true reconstruction in this event? $\to$ can be used to cut events that are not reconstructable
- No reco efficiency dependence found on top mass, parton shower, BSM (Z')

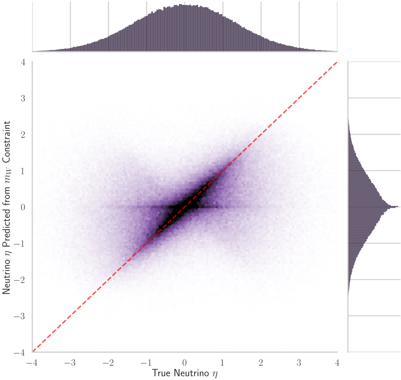
New Features: Regression of Kinematics (WIP)

- Train SPANET to reconstruct $t\bar{t}$ and simultaneously train an output head to predict neutrino η

SPA-NET regression



m_W constraint



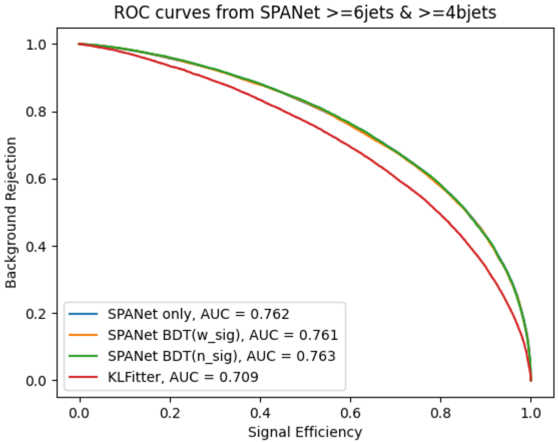
- Much more diagonal than traditional "pseudo-top" using m_W constraint

---

The code for these features is not public yet, but can be provided upon request

New Features: Direct Signal vs Background Discrimination (WIP)

- Train SPANET to reconstruct $t\bar{t}H, H \to b\bar{b}$, and simultaneously train an output head to separate $t\bar{t}H$ and $t\bar{t} + b\bar{b}$



- equivalent performance to training BDT on HL variables calculated using reconstructed system
- Both outperform BDT trained on same variables using KLFitter reconstruction

The code for these features is not public yet, but can be provided upon request 🙂

The SPA-NET Package : ▸ GITHUB

- We have released a user friendly package to implement SPA-NET for arbitrary final states (NOT limited to top!)
- Simple config file:

```
[SOURCE]
mass = log_normalize
pt = log_normalize
eta = normalize
phi = normalize
btag = none

[EVENT]
particles = (t1, t2)
permutations = [(t1, t2)]

[t1]
jets = (q1, q2, b)
permutations = [(q1, q2)]

[t2]
jets = (q1, q2, b)
permutations = [(q1, q2)]
```

Input features, normalisations

Target topology

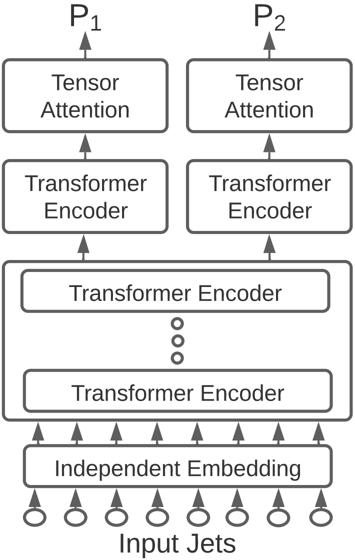
Symmetry between particles

Particle decays and symmetries

- Easy to add new input variables, change symmetry assumptions, etc
- Just need to convert data into right format and then run train.py 🙂

Summary

- If you reconstruct multi-object final states in your analysis, you should use SPA-NET 😉
- Full event reconstruction is possible in arbitrarily complex final states - even all-hadronic 4-top!
- No strong dependence on training data used $\rightarrow$ do not expect uncertainties to be enhanced by using in measurements
- New features in development to improve reconstruction of missing components, signal vs background discrimination, cutting out unreconstructable events...
- Full details in arXiv:2010.09206 arXiv:2106.03898 (and more to follow)
- Code at GitHub
- Please come find me at the breaks if you'd like to discuss 🙂



Backup

Full Results: all-had $t\bar{t}$

- AllHad ttbar

		Event	SPA-NET Efficiency		χ^2 Efficiency	
	N_{jets}	Fraction	Event	Top Quark	Event	Top Quark
All Events	$== 6$	0.245	0.643	0.696	0.461	0.523
	$== 7$	0.282	0.601	0.667	0.408	0.476
	≥ 8	0.320	0.528	0.613	0.313	0.395
	Inclusive	**0.848**	**0.586**	**0.653**	**0.387**	**0.457**
Complete Events	$== 6$	0.074	0.803	0.837	0.664	0.696
	$== 7$	0.105	0.667	0.754	0.457	0.556
	≥ 8	0.145	0.521	0.662	0.281	0.429
	Inclusive	**0.325**	**0.633**	**0.732**	**0.426**	**0.532**

Full Results: all-had $t\bar{t}H, H \rightarrow b\bar{b}$

	N_{jets}	Event Fraction	SPA-NET Efficiency Event	SPA-NET Efficiency Higgs	SPA-NET Efficiency Top	χ^2 Efficiency Event	χ^2 Efficiency Higgs	χ^2 Efficiency Top
All Events	$== 8$	0.261	0.370	0.497	0.540	0.056	0.193	0.092
	$== 9$	0.313	0.343	0.492	0.514	0.053	0.160	0.102
	≥ 10	0.313	0.294	0.472	0.473	0.031	0.150	0.056
	Inclusive	**0.972**	**0.330**	**0.485**	**0.502**	**0.045**	**0.164**	**0.081**
Complete Events	$== 8$	0.042	0.532	0.657	0.663	0.040	0.220	0.135
	$== 9$	0.070	0.422	0.601	0.596	0.019	0.152	0.079
	≥ 10	0.115	0.306	0.545	0.523	0.004	0.126	0.073
	Inclusive	**0.228**	**0.383**	**0.583**	**0.572**	**0.016**	**0.153**	**0.087**

Full Results: all-had 4-top

	N_{jets}	Event Fraction	SPA-NET Efficiency Event	SPA-NET Efficiency Top Quark
All Events	$== 12$	0.219	0.276	0.484
	$== 13$	0.304	0.247	0.474
	≥ 14	0.450	0.198	0.450
	Inclusive	**0.974**	**0.231**	**0.464**
Complete Events	$== 12$	0.005	0.350	0.617
	$== 13$	0.016	0.249	0.567
	≥ 14	0.044	0.149	0.504
	Inclusive	**0.066**	**0.191**	**0.529**

Full Results: lepton+jets $t\bar{t}$ (WIP)

Jet Selection	Full Event Efficiency			
	All	W_{had}	t_{had}	t_{lep}
= 4 Jets	84.64	90.64	86.15	90.05
= 5 Jets	72.61	80.16	82.39	87.01
= 6 Jets	62.91	71.59	79.03	84.16
= 7 Jets	55.87	65.66	75.53	82.06
= 8 Jets	50.10	60.25	73.63	79.44
= 9 Jets	46.39	55.65	70.41	79.20
= 10 Jets	42.72	54.64	67.88	75.50

Full Results: lepton+jets $t\bar{t}H, H \rightarrow b\bar{b}$ (WIP)

Jet Selection	Reconstruction Efficiency (Full Events)				
	All	W_{had}	b_{had}	b_{lep}	Higgs
$= 6$ Jets	54.12	79.88	68.63	73.10	62.75
$= 7$ Jets	42.60	69.55	63.29	69.60	56.16
$= 8$ Jets	35.29	62.17	60.82	68.18	52.52
$= 9$ Jets	29.63	56.08	56.83	66.79	48.92
$= 10$ Jets	23.86	50.38	54.37	65.87	45.63
$= 11$ Jets	22.35	43.18	52.65	69.32	44.32
$= 12$ Jets	26.03	45.21	46.58	71.23	46.58

χ^2 method

$$\chi^2 = \frac{(m_{bqq} - m_{b'q'q'})^2}{\sigma^2_{\Delta m_{bqq}}} + \frac{(m_{qq} - m_W)^2}{\sigma^2_W} + \frac{(m_{q'q'} - m_W)^2}{\sigma^2_W} \tag{1}$$

- Exhaustively evaluate χ^2 for each permutation of the event and take the minimum as the best reconstruction

Permutation DNN

Results from ▸ Erdmann et al

Jet selection	Algorithm	Reconstruction efficiency			
		all	W_{had}	b_{had}	b_{lep}
≥ 4 jets	DNN	80.2%	85.0%	82.2%	89.8%
	KLFitter (m_t^{fixed})	72.7%	79.7%	76.2%	83.3%
	KLFitter	64.5%	76.5%	68.5%	76.3%
≥ 5 jets	DNN	66.6%	75.8%	76.7%	83.6%
	KLFitter (m_t^{fixed})	36.5%	61.1%	59.4%	59.5%
	KLFitter	27.6%	56.2%	48.1%	51.5%
≥ 6 jets	DNN	57.1%	68.3%	72.7%	79.3%
	KLFitter (m_t^{fixed})	20.7%	47.8%	47.5%	46.6%
	KLFitter	14.1%	42.1%	36.0%	38.8%

- Both DNN and KLFitter systematically lower in our dataset than in this dataset (NB: results presented as $\geq N$ jets so not always directly comparable)
- Trends consistent