

End-To-End Boosted Jet ID for Top/QCD Discrimination in a 200 PU Environment in the Phase II CMS Detector

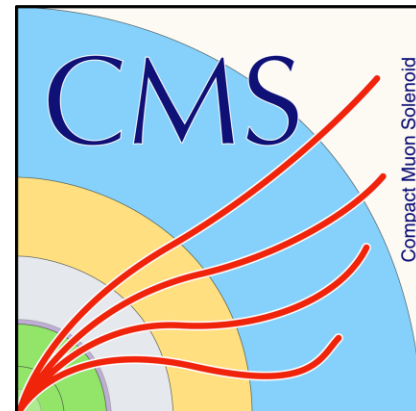
S. An^[1], M. Andrews^[1], **B. Burkle**^[2], S. Gleyzer^[3], U. Heintz^[2], J. Hogan^[4], L. Kang^[2], M. Narain^[2], E. Usai^[2]

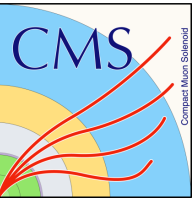
[1] **Carnegie
Mellon
University**



[3] **UF**
UNIVERSITY of
FLORIDA

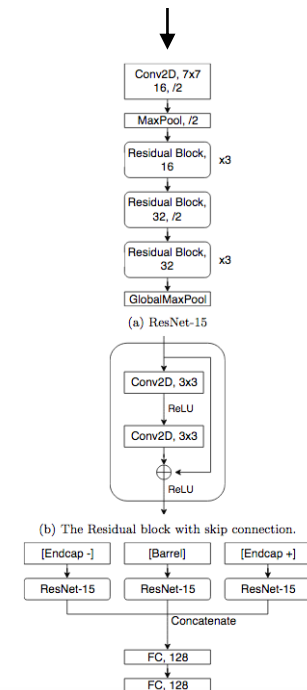
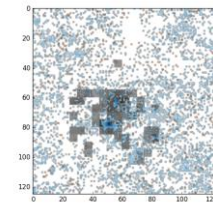
[4] 
BETHEL
UNIVERSITY

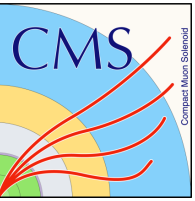
The logo for Bethel University, featuring a shield with a blue background and a yellow flame-like shape, with the year "1871" on a banner below. The text "BETHEL UNIVERSITY" is written in a serif font below the shield.



Jet Identification using End-to-End

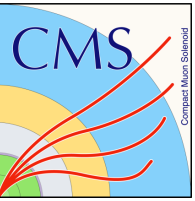
- End-to-End uses convolutional networks to identify physics objects seen in the CMS detector at the Large Hadron Collider
- Current implementation uses the following detector partitions to create a 3-channel image:
 - ECAL Showers – ECAL granularity sets pixel size of the image
 - HCAL Energy Towers – the channel pixels up-sampled to match ECAL granularity
 - Tracks – projection of tracks from the tracker onto the ECAL. Pixel size down-sampled to ECAL granularity
- For more information, see:
<https://indico.fnal.gov/event/18104/session/23/contribution/75/material/1/0.pdf>





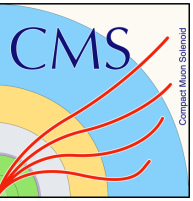
Motivation

- Given success of the End-to-End approach to particle identification, we are looking to apply to various regimes: quark/gluon discrimination, electromagnetic showers, rare Higgs decays, boosted topologies
- Top quarks exist in the final state of many physics searches which are being explored at the LHC
 - Many beyond standard model searches expect these tops to also be highly boosted
- With the upgrade to the HL-LHC in 2026, we expect over 200 simultaneous collisions per bunch crossing. This is almost 3 times as many as the peak pile-up conditions seen in this past years run (approx. 70)



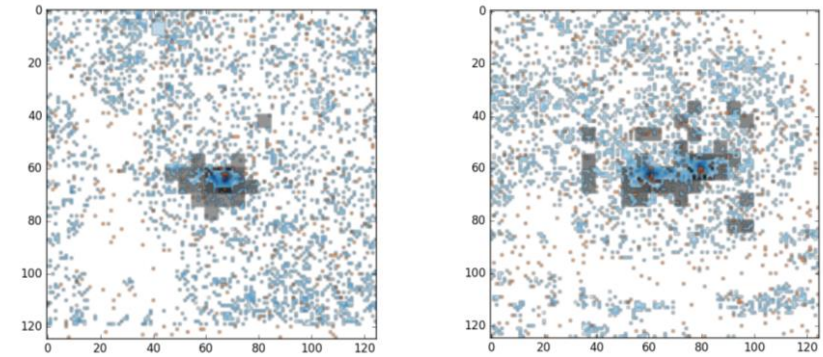
Jet Selection

- Jets were selected from the following 2017 200 PU Phase II CMS detector Monte Carlo samples
 - Top: TT_Mtt1500toInf_TuneCUETP8M1_14TeV-powheg-pythia8
 - QCD: QCD_Flat_Pt-15to7000_TuneCUETP8M1_14TeV-powheg-pythia8
- The following selection criteria were made for all jets
 - Total transverse moment of the an ak8 jet is greater than 400 GeV
 - The entire jet is contained within the barrel region of the detector
- Additionally, gen-level truth matching was used for top jet selection to only select hadronic top decays where the b and W jets were fully contained within the ak8 jet cone

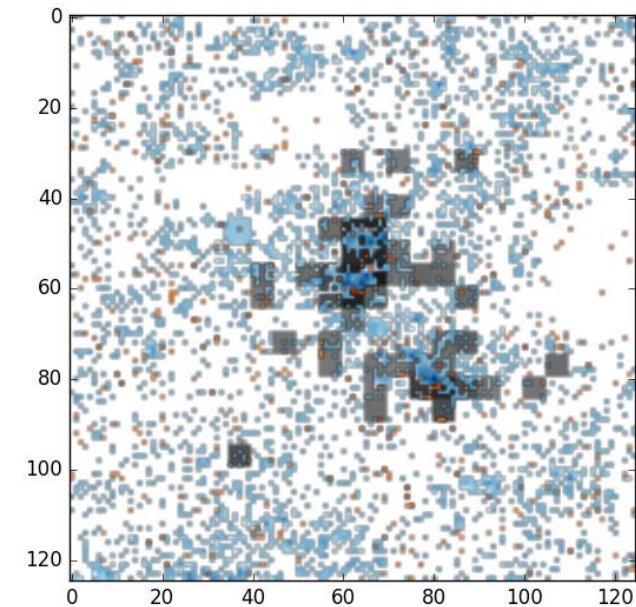


Jet Selection (cont)

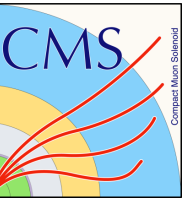
- There are two ways that QCD jets pass our selection criteria.
 - Having a single boosted jet (top left)
 - Having multiple sub-jets (top right)
- All top jets that pass our selection (bottom) contain 3 jets
- After selection, datasets consisted of the following sizes
 - Training Set: 700k jets
 - Validation Set: 50k jets
 - Testing Set: 50k jets



Jet images coming from QCD sample



Jet images coming from Top decay



Other Networks: DeepJet/DeepAK8

- DeepJet is a top jet tagger which uses a combination of 1D CNNs, RNNs, and fully connected layers to perform jet ID
- Uses particle flow candidates and secondary vertex information as network input

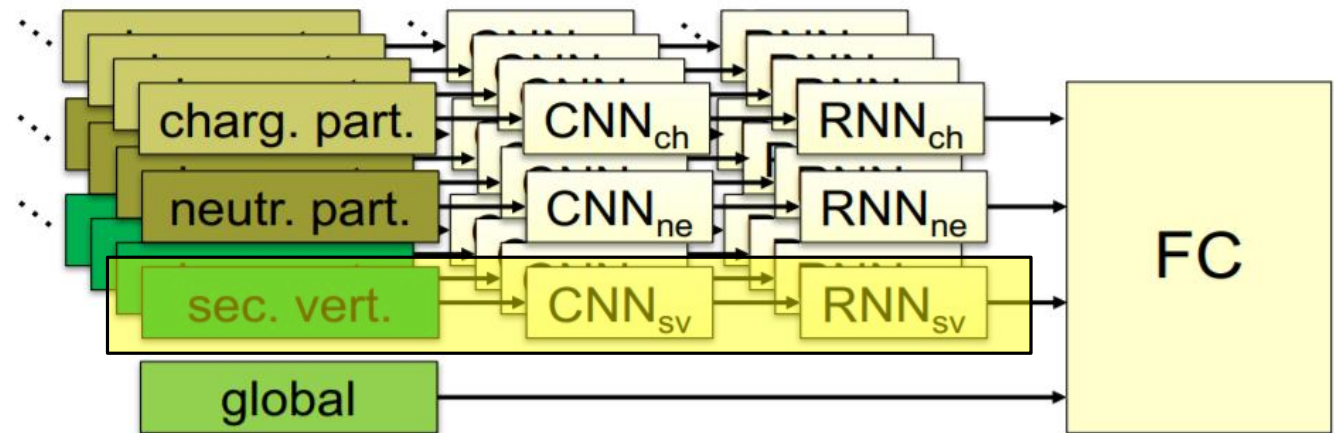
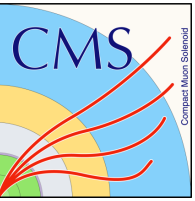
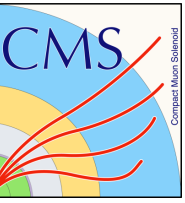


Image taken from Markus Stoye's presentation from 2018 Boosted Objects Workshop



Other Networks: BEST

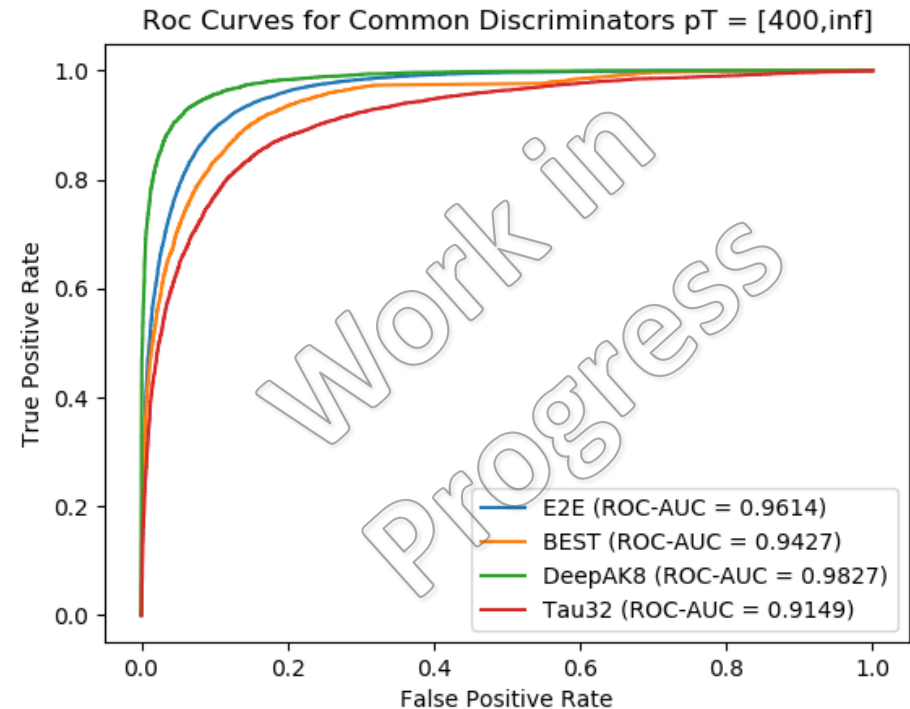
- BEST is neural network specialized in performing boosted jet ID
 - Discriminates between jets originating from t, W, Z, H, b, and QCD background
- Input variables put an emphasis on using the jet mass and the angular distributions of sub-jets
 - Like DeepAK8, BEST is based off of high level variables and includes secondary vertex information as an input

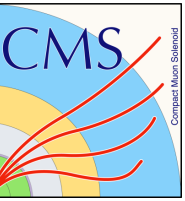


End-to-End vs Other Algorithms

Note: These benchmark tests are very rough; the algorithms were not trained equivalently.

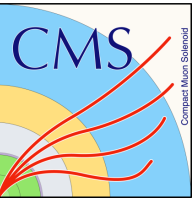
- DeepAK8 and BEST algorithms are both trained using the full set of particle flow variables
 - End-to-End does not yet have the full tracker information
 - Adding this is expected to give End-to-End a strong boost in discriminating power
- DeepJet was trained on 100 million jets compared to 700 thousand jets which we trained on for this study





Moving Forward

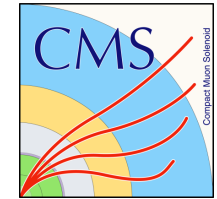
- Implement tracker rec hits into detector image
 - Will let our network learn information about the jet's primary and secondary vertices
 - Secondary vertex information is one of the strongest high level variables used when identifying b-jets
- Dataset Improvements
 - Increase the size of the training dataset
 - Increase the total number of jets and use a flatter p_T distribution for top jets
 - Train network using more QCD samples
 - QCD dijet samples, b-enriched QCD samples, more ak8 jets which contain multiple sub-jets, etc.



Conclusion

- End-to-End is a state of the art approach to particle ID for electrons, photons, and jets (q/g discrimination)
- Preliminary studies show that End-to-End is already comparable to other top algorithms for top/QCD discrimination
 - After introducing additional vertexing and tracking information, we are optimistic that End-to-End performance will improve even more for these specialized cases

Backup



p_T – binned AUC scores

	400 - 600	600 - 800	800 - 1000	1000 - 1200	1200 - inf	All
E2E	0.9521	0.9614	0.9583	0.9681	0.9639	0.9614
Deep AK8	0.9850	0.9766	0.9811	0.9802	0.9571	0.9827
Best	0.9563	0.9452	0.9229	0.9020	0.9010	0.9427
Tau32	0.9290	0.9052	0.8732	0.8749	0.8107	0.9149