Identification of highly boosted $Z \rightarrow ee$ decays with the ATLAS detector

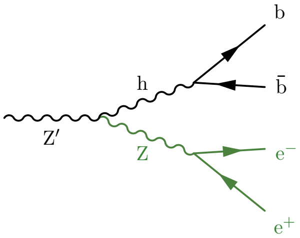
Florian Kiwit
supervised by Dominik Duda

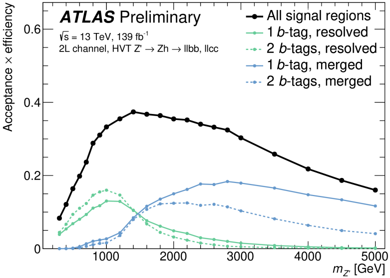
Max Planck Institute for Physics
Munich

DPG Conference Heidelberg, March 23$^{\mathrm{th}}$ 2022

Why are highly boosted Z bosons interesting?

- Many BSM theories predict new heavy vector bosons with masses at the TeV scale
  - e.g. GUT, Composite Higgs, Extra Dimensions
  - These particles can have large branching ratios to $\mathrm{h/W/Z}$ bosons
  - Heavy BSM particles will lead to high $\mathrm{p_T}$ $\mathrm{h/W/Z}$ bosons
  - Identification of boosted boson decays is crucial



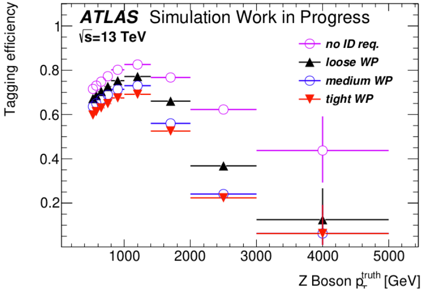


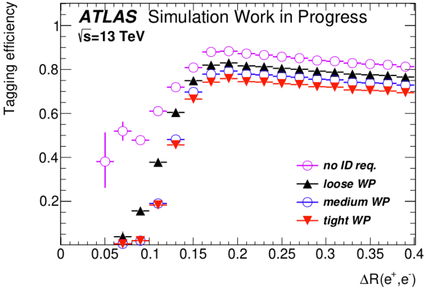
ATLAS-CONF-2020-043

The Challenge

Standard method of electron reconstruction and identification degrades with large p_T (i.e. large $m_{Z'}$)

- Due to small angular separation between the leptons from the Z decay

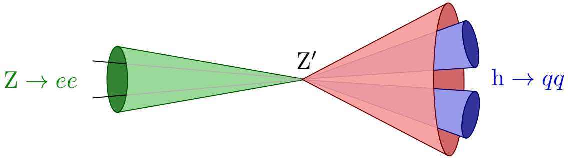




The Solution

- Reconstruct $\mathrm{Z} \to \mathrm{ee}$ as Small-Radius jet

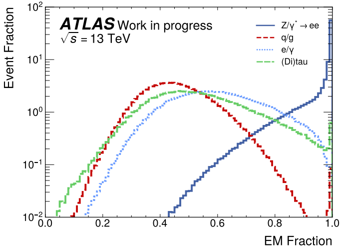
- Develop $\mathrm{Z} \to \mathrm{ee}$ tagger based on Neural Network to mitigate efficiency loss
- Training based on properties of
  - Small-Radius jet
  - Tracks (Inner Detector) } Matched to Small-R jet
  - Cluster (Calorimeter) }

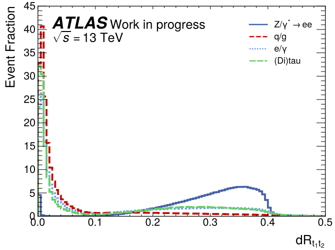


Neural Network Inputs

- MC-Simulation with Sherpa 2.2.11
  - $\mathrm{Z} \rightarrow \mathrm{ee}$ Signal
  - e/γ
  - $\mathrm{(Di)tau}$
  - $\mathrm{q/g}$

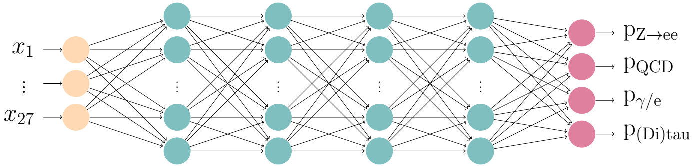
  $\left.\begin{array}{l}\mathrm{e}/\gamma \\ \mathrm{(Di)tau} \\ \mathrm{q/g}\end{array}\right\}$ Background
- Selections of $\mathrm{Z} \rightarrow \mathrm{ee}$ candidate Small-R jets:
  - $\mathrm{N_{Trks}} \geq 2$
  - $\mathrm{m_j} \in [50, 150]$ GeV
  - $\mathrm{p_{T,\ j}} > 500$ GeV
- Most important of all 27 input variables:
  - $\mathrm{EM_{Frac}}$: $\frac{\mathrm{E_{EM\ Calo}}}{\mathrm{E_{jet}}}$
  - $\mathrm{N_{Trks}}$: Number of tracks
  - $\Delta \mathrm{R_{t_1 t_2}}$: Angular separation between the two leading tracks





Neural Network Architecture

- Input distributions normalized
- Hyperparameter optimized with random and grid search
- Feed-forward neural network (PyTorch)
  - Activation: Sigmoid
  - Hidden layer: 4
  - Nodes per layer: 200
  - Output nodes: 4 (Bkg composition analysis dependent)



Neural Network Discriminant

- Discriminant quantifies 4-dim output for binary decision (Sig/Bkg)
- Performance characterized by:

$$\varepsilon = \frac{TP}{TP + FN} \quad \text{(Efficiency)}$$

$$ACC = \frac{TP}{TP + FP} \quad \text{(Accuracy)}$$

$$REJ = \frac{TP + FP}{FP} \quad \text{(Rejection Rate)}$$

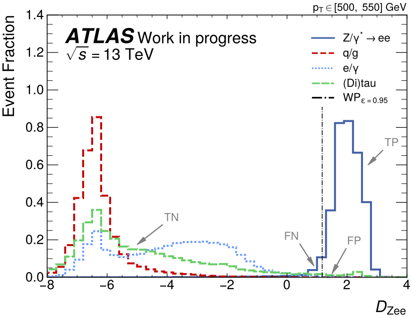
T : *True*, F : *False*, P : *Positive*, N : *Negative*

- Individual working points (WP) for $\varepsilon = 0.95$ in every $\mathrm{p_T}$ bin

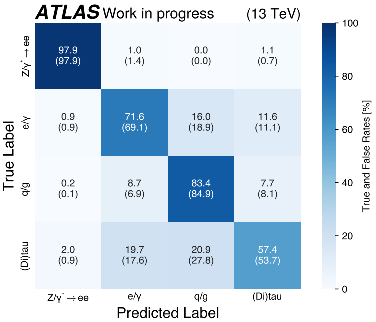
$$D_{\mathrm{Zee}} = \ln\left(\frac{(1 - f_{Zee}) \cdot p_{Zee}}{f_{QCD} \cdot p_{QCD} + f_{e/\gamma} \cdot p_{e/\gamma} + f_{\tau} \cdot p_{\tau}}\right)$$

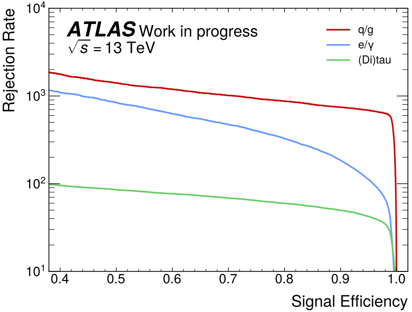
p : Probability

f : Class fraction



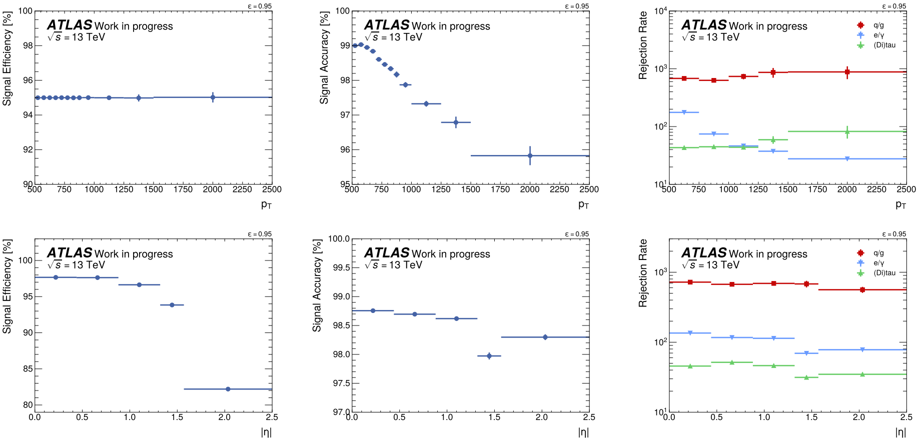
Separation Strength





- True and predicted labels for validation (train) sample
- High discrimination power between signal and backgrounds
- Significant separation between the different background classes
- Choose WP with signal efficiency of $\varepsilon = 0.95$

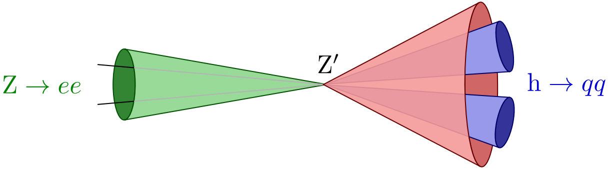
Neural Network Performance



- Accuracy decreased for very high $\mathrm{p_T}$ due to merging of the clusters
- Slightly decreased accuracy for $\eta \in [1.32, 1.57]$ due to supply wiring of the solenoid

Concluding Remarks

- $\mathrm{Z} \to \mathrm{ee}$ multiclass tagger reaches promising performance
  - Identification based on Small-R jet, Calorimeter and Inner Detector variables
  - Finalize development of tagger (Dropout, Hyperparameter optimization, Test additional inputs)
- Testing the tagger on analysis level
  - Search for W'/Z' decay as Small-Radius jet and Large-Radius jet
  - Determine updated $\sigma \times \mathrm{BR}$ at 95% CL exclusion limits and compare with standard approach



Backup

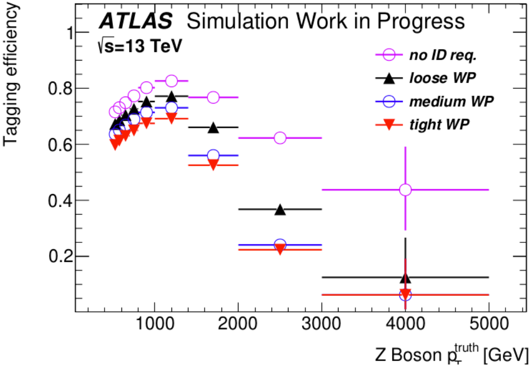
References

CMS Collaboration
Search for heavy resonances that decay into a vector boson and a Higgs boson in hadronic final states at $\sqrt{s} = 13$ TeV
The European Physical Journal C (2017)

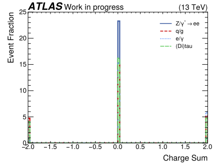
ATLAS Collaboration
Search for heavy resonances decaying into a Z boson and a Higgs boson in final states with leptons and b-jets in 139 fb^{-1} of pp collisions at $\sqrt{s} = 13 TeV$ with the ATLAS detector
ATLAS-CONF-2020-043 (2020)

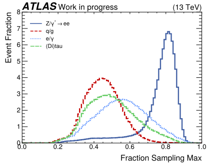
ATLAS Collaboration
ATLAS b-jet identification performance and efficiency measurement with $t\bar{t}$ events in pp collisions at $\sqrt{s} = 13$ TeV
The European Physical Journal C (2019)

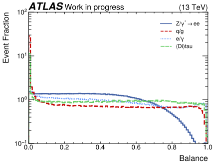
Tagging Efficiency p_T

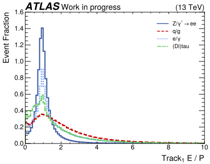


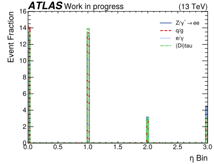
Input Distributions

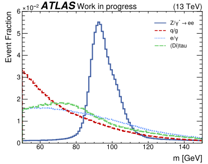


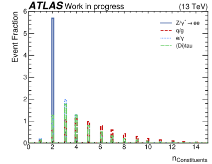


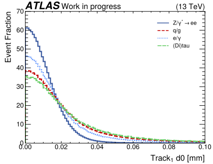


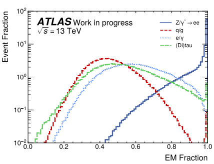


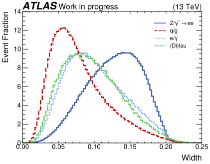


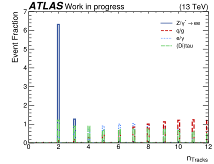


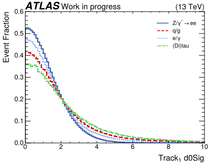




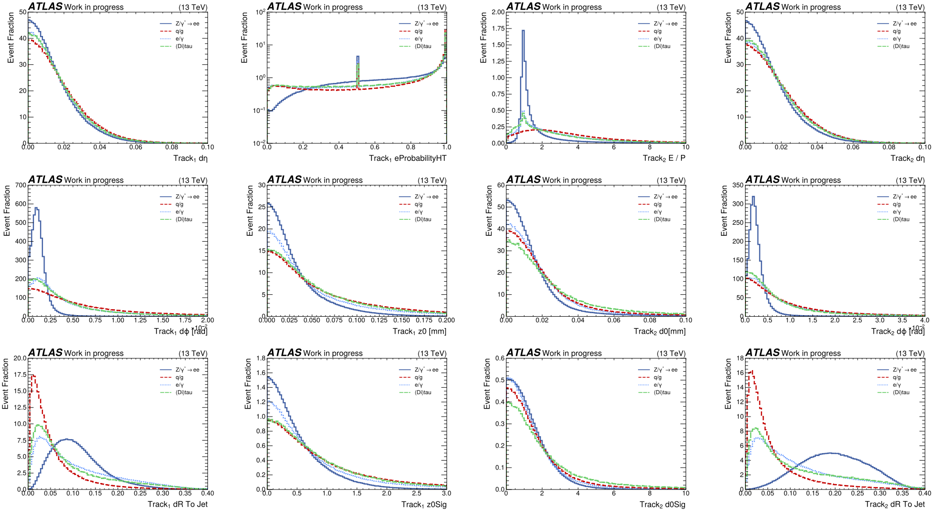




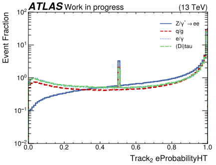


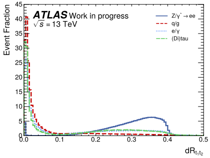


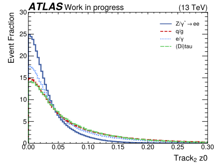
Input Distributions

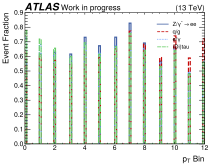


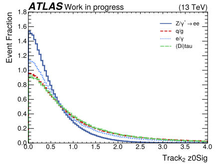
Input Distributions











ATLAS Work in progress

Class	#Events
$Z \rightarrow ee$	1,994,086
e/γ	1,663,522
QCD	2,235,635
(Di)tau	372,562

ATLAS Work in progress

Input parameter	Definition
dR tt	Radial distance between tracks with highest and second highest p_T
Balance	$\frac{p_{T,\text{Trk1}} - p_{T,\text{Trk2}}}{p_{T,\text{Trk1}} + p_{T,\text{Trk2}}}$
Charge	Sum of charge of the two tracks with the highest p_T
EM Fraction	Fraction of energy deposited in the EM calorimeter compared to the total energy of the jet
Eta Bin	Observable that describes in which eta interval the dielectron candidate jet is located
Frac Sampling Max	The ratio of a partial jet energy (contained in the calorimeter layer with the most energy of that jet) to the total jet energy
Mass	Mass of the dielectron candidate jet
N Constituents	Number of calorimeter clusters within the dielectron candidate jet
N Trks	Number of tracks
Width	Width of the dielectron candidate jet
pT Bin	Observable that describes in which p_T interval the dielectron candidate jet is located
Trk1 (Trk2) d0	Transverse impact parameter of the leading (subleading) track
Trk1 (Trk2) d0sig	Significance of the transverse impact parameter of the leading (subleading) track
Trk1 (Trk2) dEta	$\Delta\eta$ between track 1 (2) and closest calorimeter cluster
Trk1 (Trk2) dPhi	$\Delta\phi$ between track 1 (2) and closest calorimeter cluster
Trk1 (Trk2) dRToJet	Radial distance between track with highest (second highest) p_T to the axis of the jet
Trk1 (Trk2) eProbability HT	Electron probability based on transition radiation in the TRT for the leading (subleading) track in the fat electron candidate jet
Trk1 (Trk2) EOverP	$\frac{E_\text{cluster}}{p_{T,\text{Trk1}}} \left(\frac{E_\text{cluster}}{p_{T,\text{Trk2}}}\right)$
Trk1 (Trk2) z0	Longitudinal impact parameter of the leading (subleading) track
Trk1 (Trk2) z0sig	Significance of the longitudinal impact parameter of the leading (subleading) track

Hyper Parameter Optimization

- 144 trainings with random configuration from search space
- Efficiency, accuracy and background rejection plotted in dependence of η and p_T for 20 trainings with highest accuracies

ATLAS Work in progress

Parameter	**Range**	**Mode**
lr	[0.0001, 0.1]	log
Hidden Size	[50, 500]	int
Midlayer	[1, 5]	int
Batch Size	[500, 5000]	int
Activation	ReLU Sigmoid LeakyReLU	item

20 Best Configurations

Latest Set of Hyperparameters:

- 200 Nodes
- 4 midllayer
- Sigmoid activation
- lr = 0.005
- Batchsize: 2048

ATLAS Work in progress

Learning R	Nodes	Layer	Batchsize	Activation	Accuracy	Stable
$8 \cdot 10^{-4}$	325	5	778	Sigmoid	75.24	✗
$5 \cdot 10^{-3}$	356	5	1221	LeakyReLU	75.20	✓
$1 \cdot 10^{-3}$	178	4	2591	Sigmoid	75.19	✓
$3 \cdot 10^{-3}$	214	4	4000	ReLU	75.18	✗
$7 \cdot 10^{-3}$	202	3	2918	ReLU	75.15	✗
$1 \cdot 10^{-3}$	433	5	4565	Sigmoid	75.15	✓
$8 \cdot 10^{-3}$	244	4	1875	LeakyReLU	75.14	✓
$6 \cdot 10^{-3}$	355	3	2324	Sigmoid	75.12	✗
$4 \cdot 10^{-3}$	252	5	4850	ReLU	75.12	✗
$4 \cdot 10^{-3}$	393	3	4286	ReLU	75.12	✗
$3 \cdot 10^{-3}$	199	2	2621	Sigmoid	75.11	✓
$2 \cdot 10^{-3}$	237	5	4405	ReLU	75.11	✗
$3 \cdot 10^{-3}$	138	5	2589	ReLU	75.10	✗
$5 \cdot 10^{-3}$	417	4	1083	Sigmoid	75.09	✗
$2 \cdot 10^{-3}$	460	2	2518	Sigmoid	75.09	✓
$7 \cdot 10^{-3}$	111	5	1907	Sigmoid	75.08	✓
$3 \cdot 10^{-3}$	228	4	2673	Sigmoid	75.07	✗
$5 \cdot 10^{-3}$	369	3	733	ReLU	75.07	✗
$2 \cdot 10^{-3}$	170	2	4603	LeakyReLU	75.06	✓
$2 \cdot 10^{-2}$	251	5	1917	LeakyReLU	75.05	✓

Variable Importance

Mutual Information:
Measure of the mutual dependence between two random variables (Each input and output)

Procedure:

1. Train 10x
2. Determine μ_{Acc} and σ_{Acc}
3. Drop input with lowest importance (Mutual Info)
4. (Repeat 1-3)

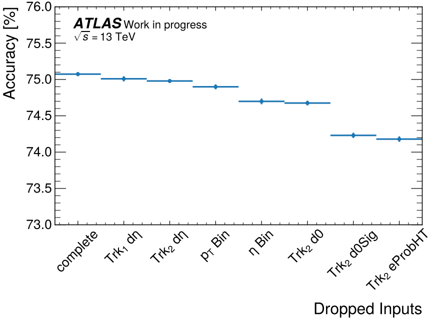
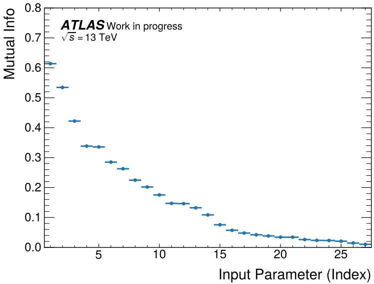


Figure: Left to right: Inputs missing additionally

Latest Set of Input Parameters

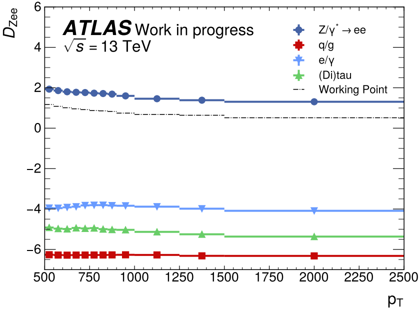
- $\mathrm{Trk}_1\ \mathrm{d}\eta$ and $\mathrm{Trk}_2\ \mathrm{d}\eta$ removed
- $\mathrm{p_T}$ Bin and η Bin needed for sample weighting
- Number of inputs reduced to 27

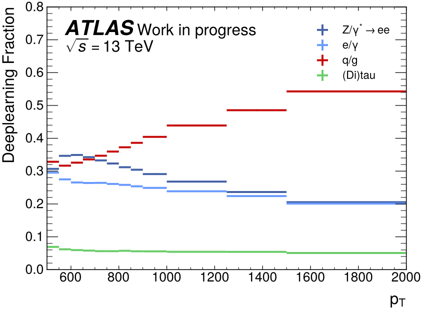


ATLAS Work in progress

Index	Parameter	Importance
1	EM Fraction	0.614
2	N Trks	0.535
3	FracSamplingMax	0.422
4	N Constituents	0.339
5	dR tt	0.336
6	Trk2 dRToJet	0.285
7	Balance	0.263
8	Mass	0.225
9	Trk1 dPhi	0.202
10	Trk1 dRToJet	0.175
11	Trk2 dPhi	0.147
12	Width	0.146
13	Trk2 E/P	0.133
14	Trk1 E/P	0.109
15	ChargeSum	0.076
16	Trk1 d0	0.057
17	Trk1 z0	0.048
18	Trk1 z0Sig	0.042
19	Trk2 z0	0.038
20	Trk2 z0Sig	0.034
21	Trk1 d0Sig	0.034
22	Trk1 eProbabilityHT	0.026
23	Trk2 d0Sig	0.024
24	Trk2 eProbabilityHT	0.023
25	Trk2 d0	0.021
26	Eta Bin	0.015
27	pT Bin	0.010

Deeplearning Score





Loss and Accuracy Curves

- Overfitting for Tau (smallest sample)
- Only small increase of val acc with reduced lr
- Network of epoch with best val acc saved for further analysis