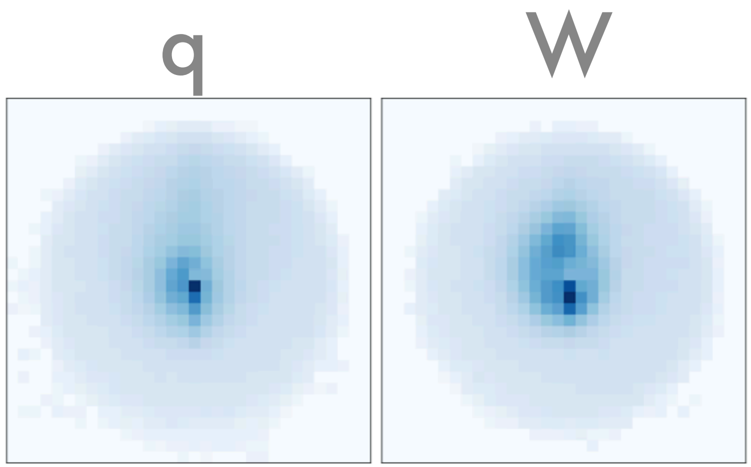
Interpretable Networks for Identifying Leptons

Daniel Whiteson, UC Irvine

Oct 2022

Motivation

Deep networks find new power for jet substructure



1603.09349

Leptons are simpler objects...

...but their backgrounds are often jets!

Questions

1. Can deep networks outperform existing physics vars for electrons and muons?

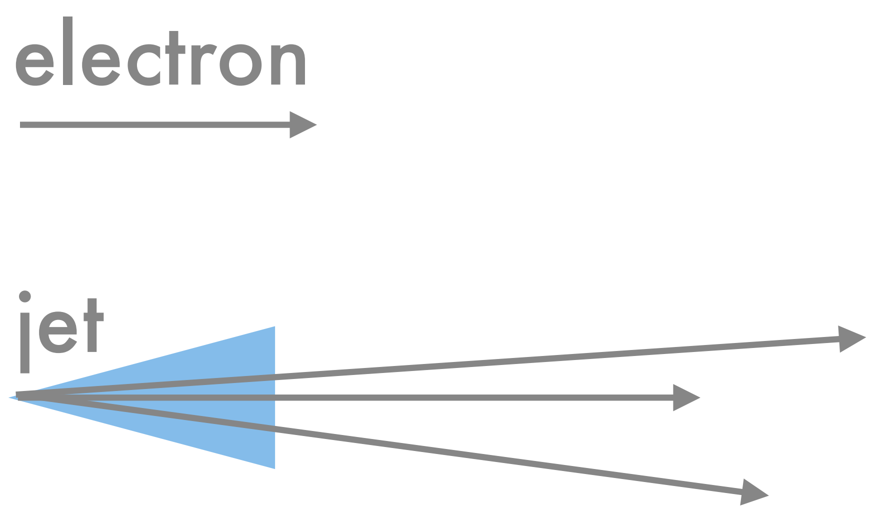
2. Can we interpret what the network has learned?

Electrons

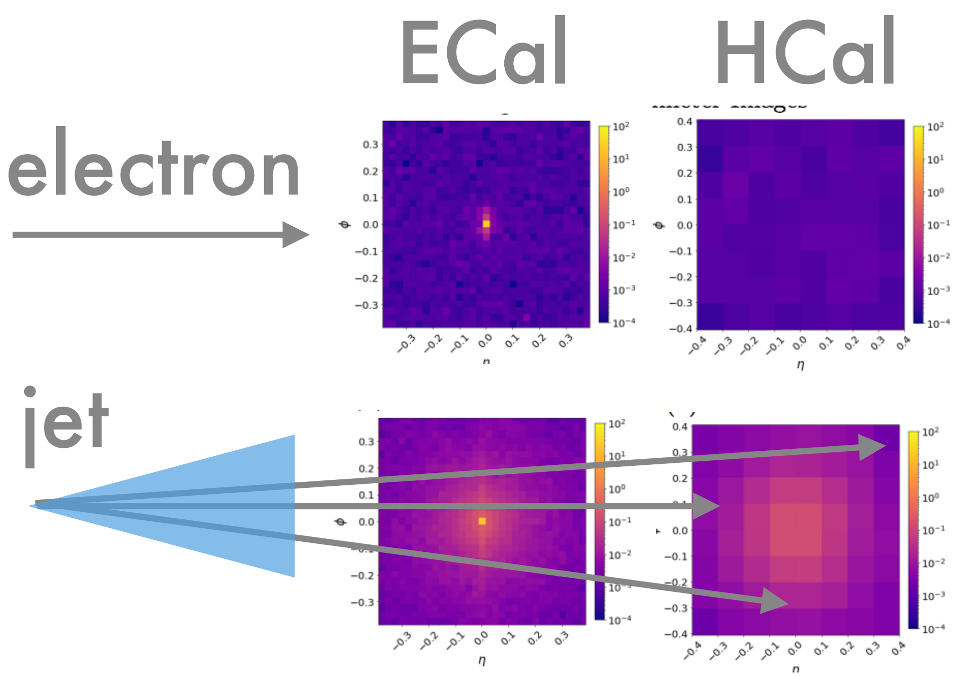
https://arxiv.org/abs/2011.01984

Learning to Identify Electrons

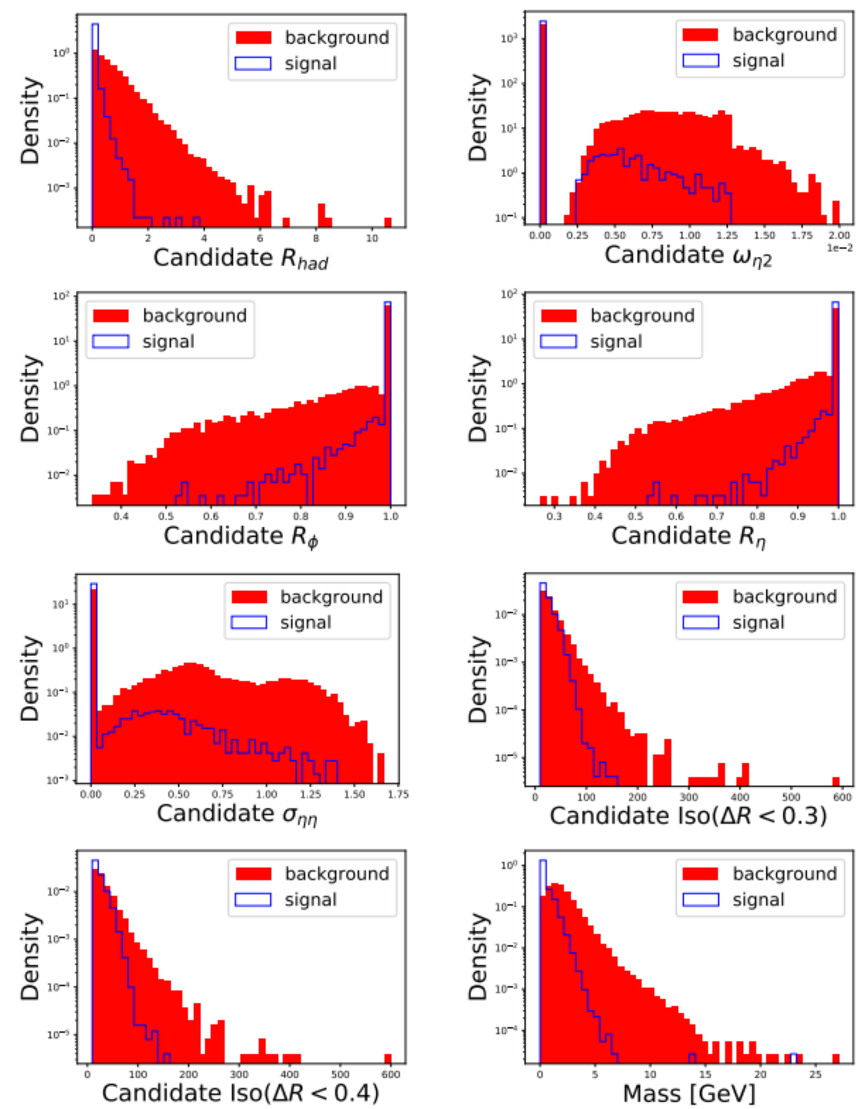
Julian Collado,[1] Jessica N. Howard,[2] Taylor Faucett,[2] Tony Tong,[2,1] Pierre Baldi,[1] and Daniel Whiteson[2]
[1]*Department of Computer Science, University of California, Irvine, CA, 92697*
[2]*Department of Physics and Astronomy, University of California, Irvine, CA 92697*
(Dated: November 5, 2020)



Electrons

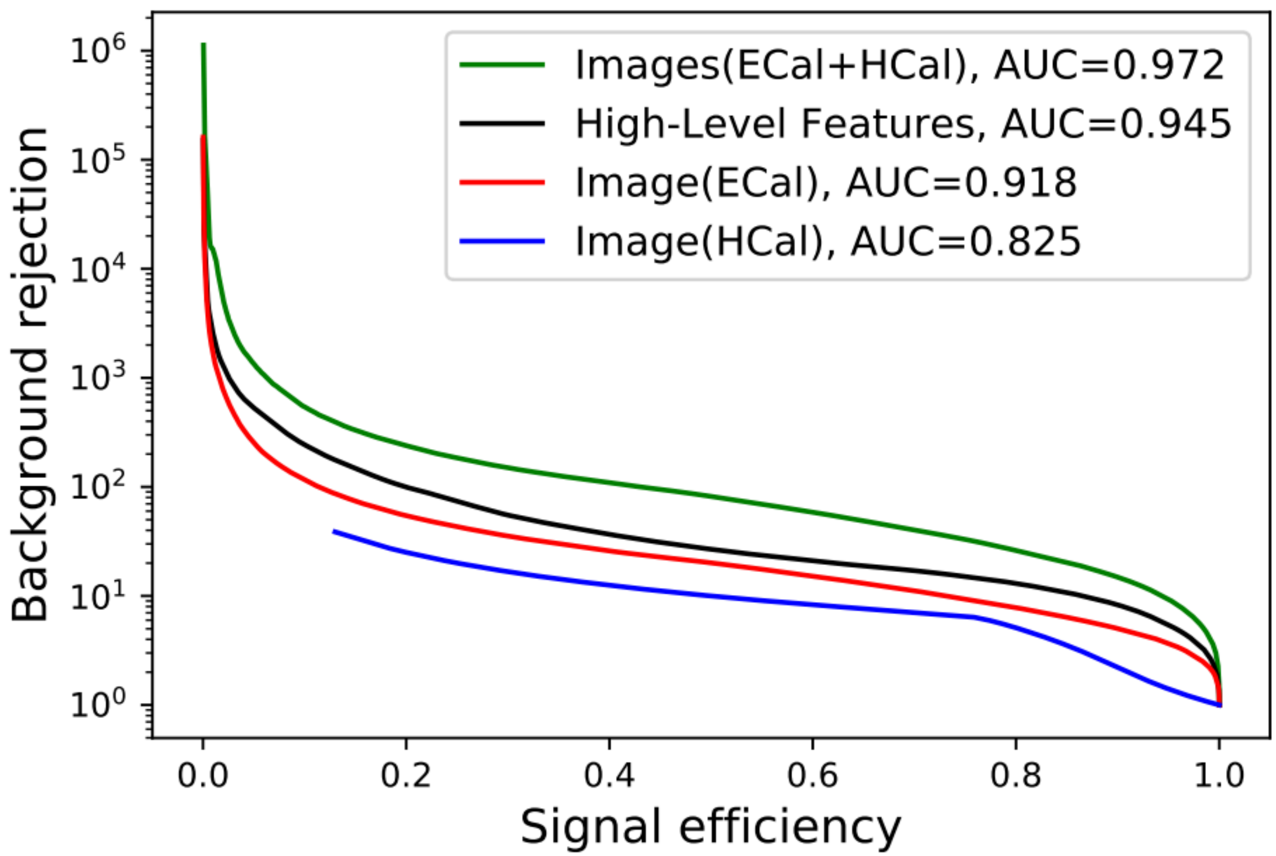


Physics features



Electrons

Performance Gap



Images outperform physicists!

Questions

1. Can deep networks outperform existing physics vars for electrons and muons?

Electrons: YES

2. Can we interpret what the network has learned?

How to interpret?

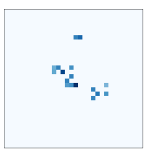
Mapping Machine-Learned Physics into a Human-Readable Space

Taylor Faucett,[1] Jesse Thaler,[2,3] and Daniel Whiteson[1]

https://arxiv.org/abs/2010.11998

What is it doing?

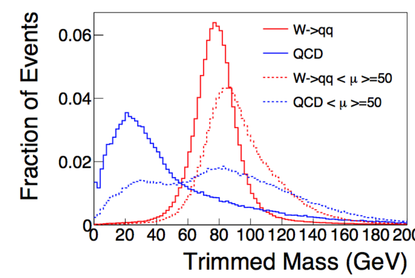
Our low-level (LL) data are often high-dim



LL

expert physics knowl.

HL



Can't interpret LL data

But HL doesn't always capture the information

Yet we prefer HL

If HL data includes all necessary information...

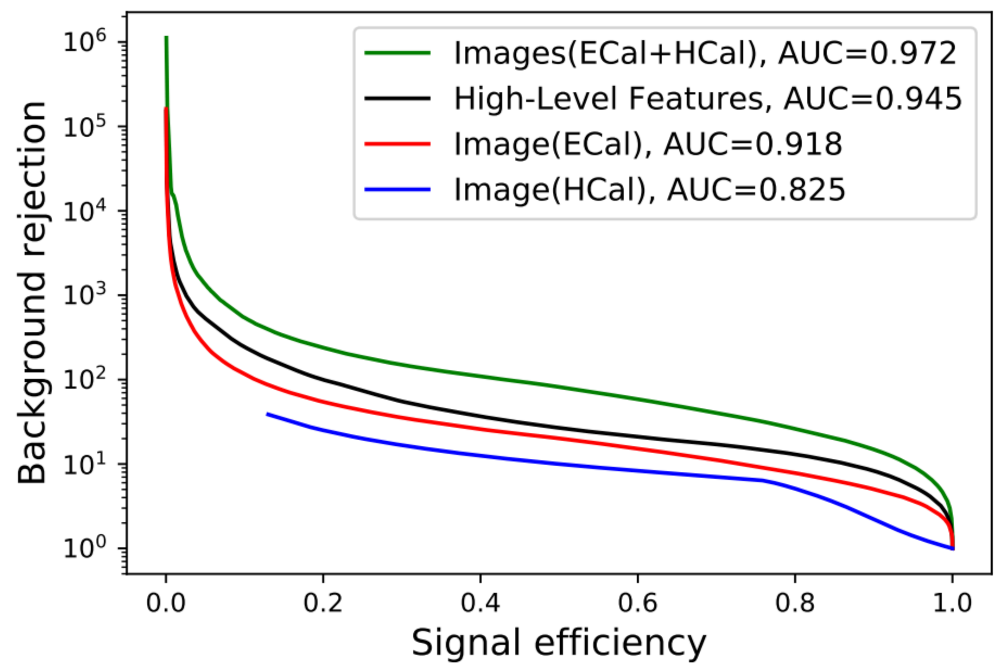
- It is easier to understand
- Its modeling can be verified
- Uncertainties can be sensibly defined
- It is more compact and efficient

- LL -> HL is physics, so we like it.

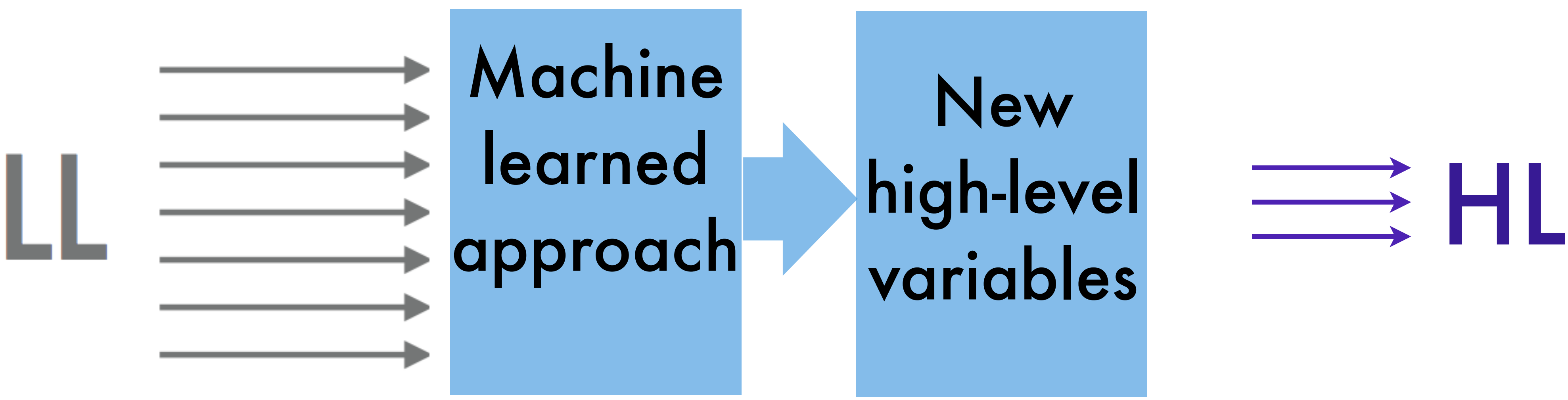
Our question

How has the DNN found its solution?
What can we learn from it?

<u>**Residual knowledge:**</u>
Is there a new HL variable?
Can it reveal physics?



Learning from ML

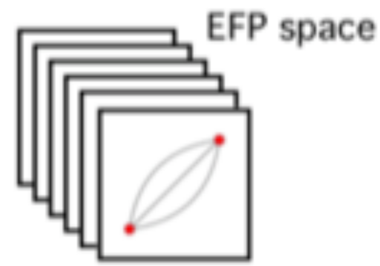


Use LL analysis as a probe, not a final product.

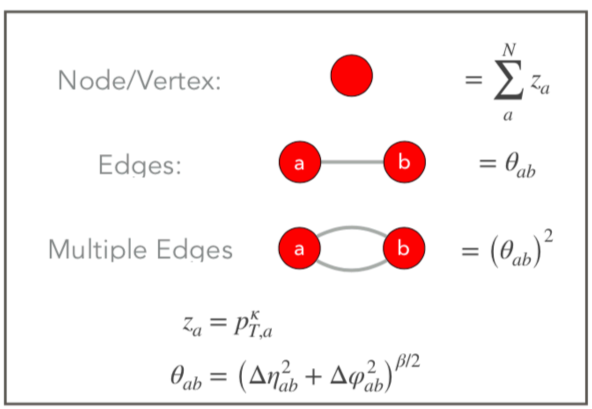
How?

I. Define space of interpretable observables

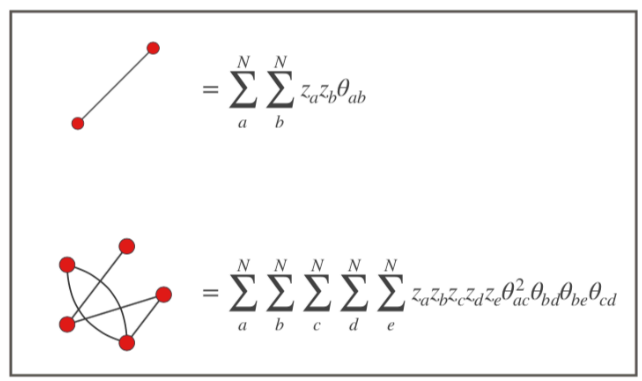
- provides context
- defines problem
- does NN live in this space?
- Can it be compactly represented?
- Yes or No are both interesting!



Graph components



Examples



How?

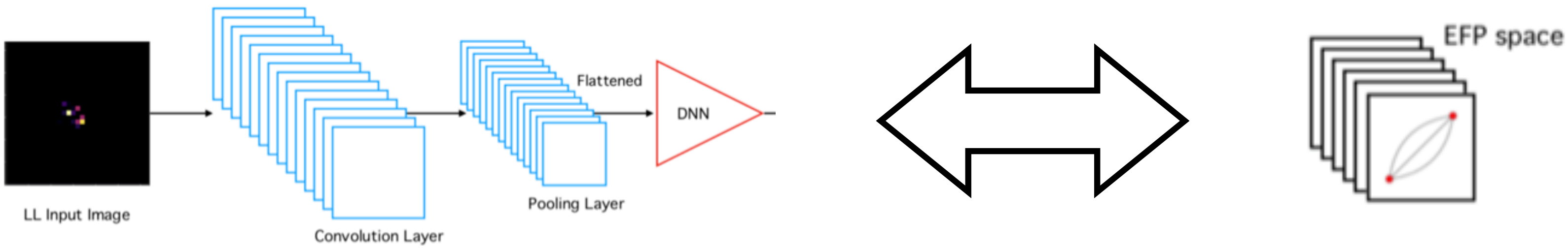
<u>I. Define space of interpretable observables</u>

- provides context
- defines problem
- does NN live in this space?
- Can it be compactly represented?
- Yes or No are both interesting!

<u>II. Define mapping metric</u>

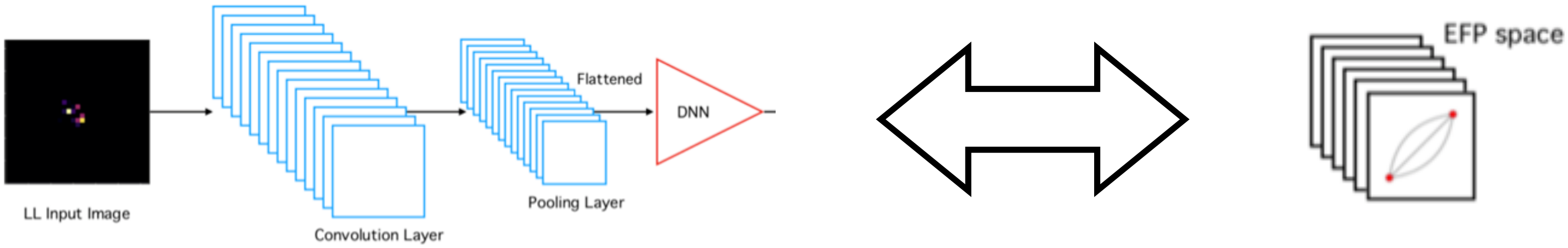
- how do you compare two solutions?
- can’t use functional identity or linear correlation

Mapping



How to map from deep network
into our space of interpretable observables?

Mapping



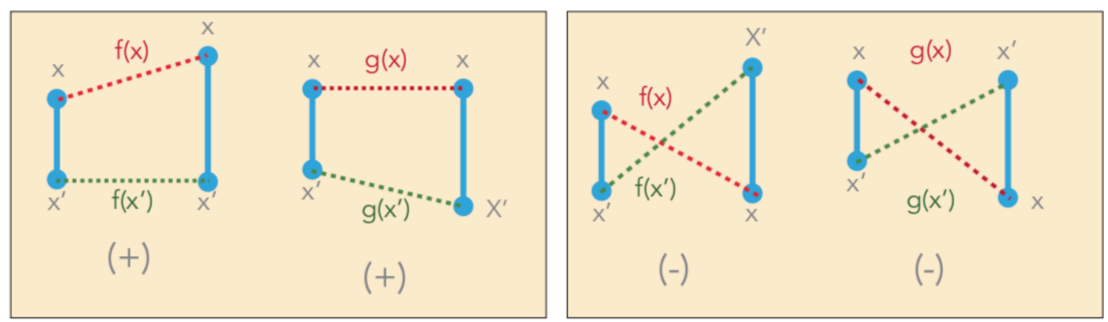
Function sameness

Complete equivalence
not the idea

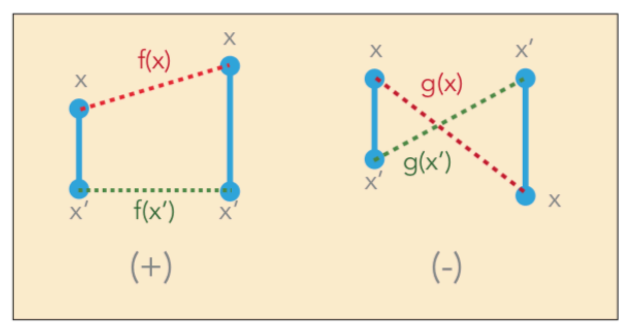
Any 1:1 transformation
of function has no impact
in our context

Only care about the
ordering of points
not the actual function
values

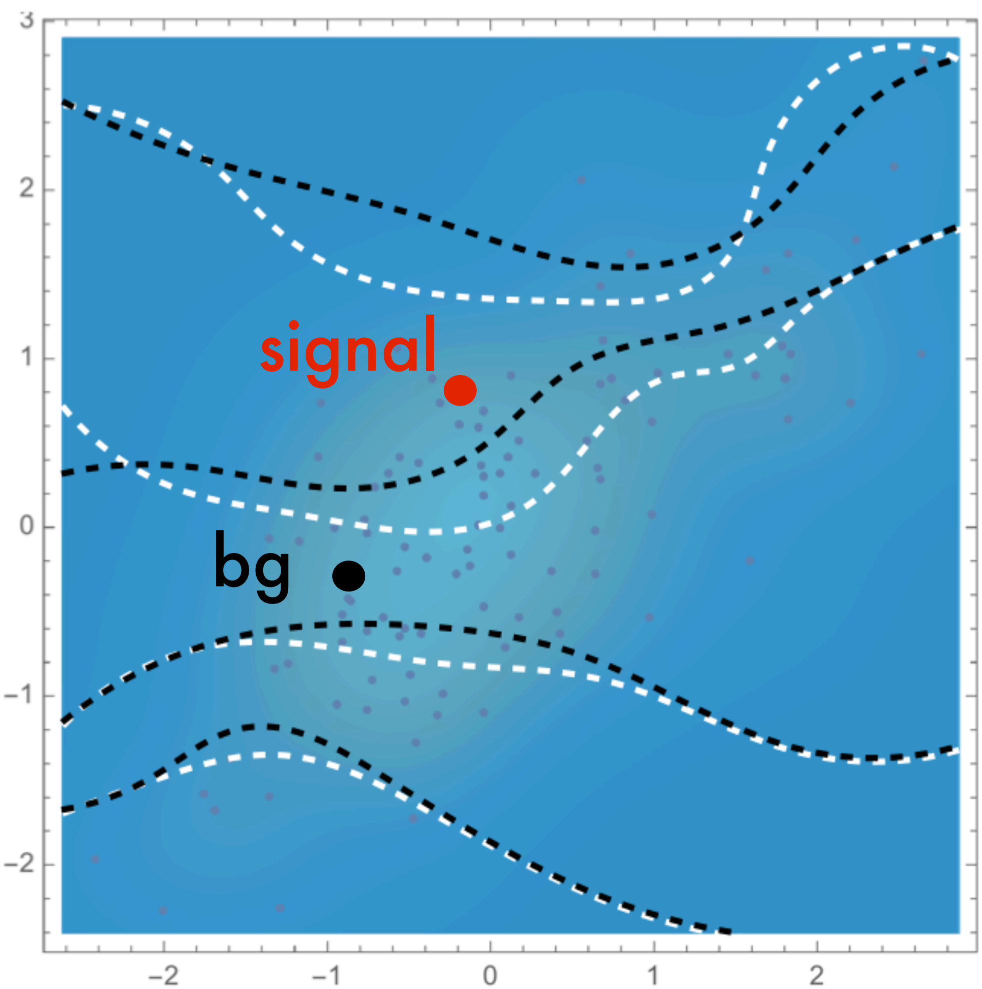
Similar Orderings



Dissimilar Orderings



Discriminant ordering



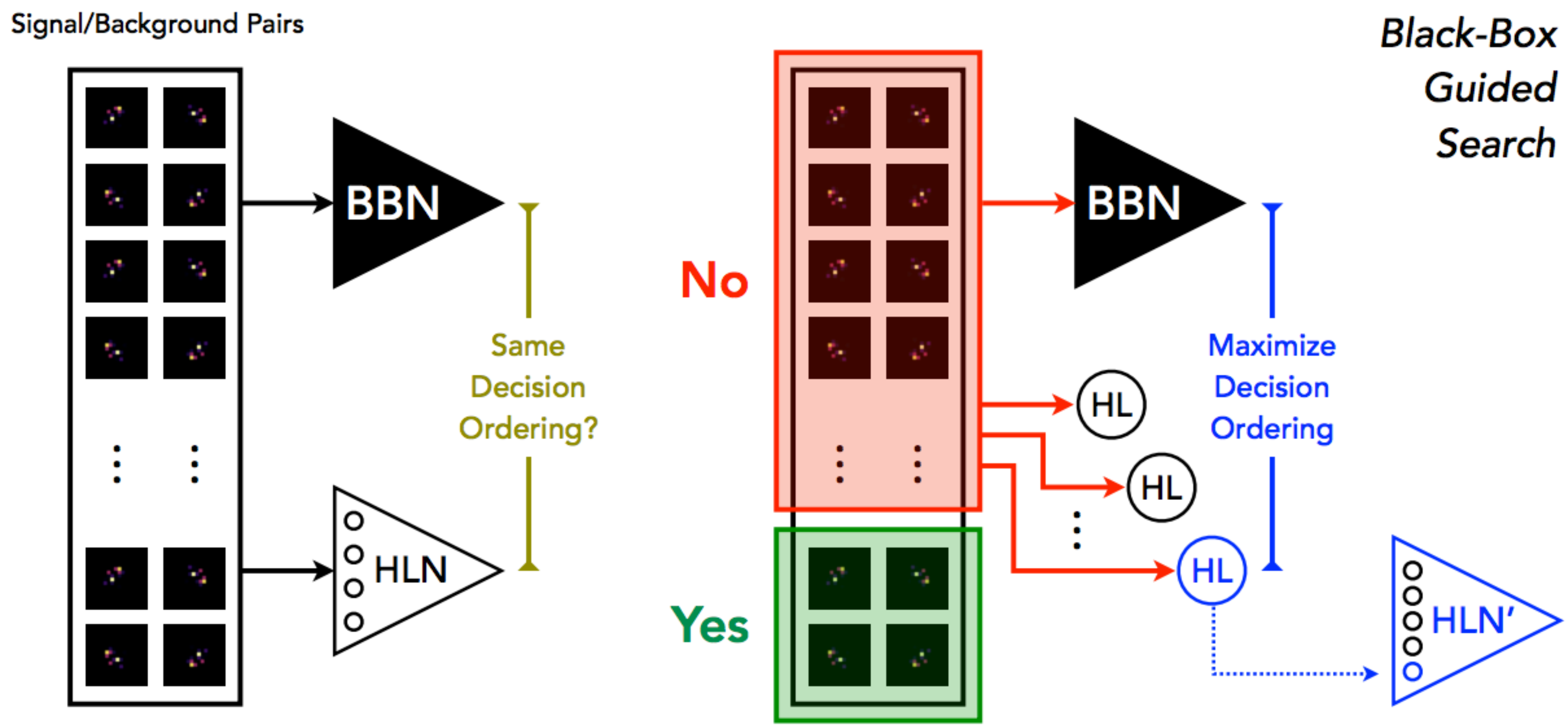
Evaluate how often they give a bg-sig pair the same ordering.

$$\mathrm{DO}(x,x') = \Theta\Big(\big(f(x)-f(x')\big)\big(g(x)-g(x')\big)\Big)$$

Sample the space.

$$\mathrm{ADO} = \int \mathrm{d}x\,\mathrm{d}x'\,p_{\mathrm{sig}}(x)\,p_{\mathrm{bkg}}(x')\,\mathrm{DO}(x,x').$$

Finding the HL



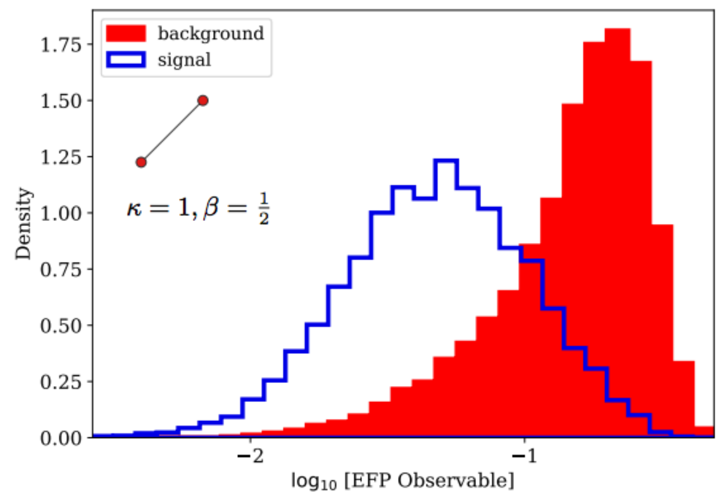
Use decision ordering to isolate disagreement and select new HL feature

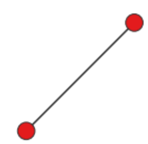
Closing the gap

Scan for electron NN finds a new feature

Created to study quarks and gluons.

Helps separate electrons and jets!



 $= \sum_{a,b=1}^{N} z_a z_b \theta_{ab}^{\frac{1}{2}}$

Closing the gap

Base		Additions (κ, β)		(AUC)
7HL				0.945
7HL	$+M_{\text{jet}}$			0.956
7HL		$(1, \frac{1}{2})$		0.970
7HL	$+M_{\text{jet}}$	$(1, 1)$	$(1, \frac{1}{2})$	0.971
7HL		$(2, -)$		0.970
7HL	$+M_{\text{jet}}$	$(2, 1)$	$(2, -)$	0.971
CNN				0.972

Muons

Learning to Isolate Muons

Julian Collado,[1] Kevin Bauer,[2] Edmund Witkowski,[2] Taylor Faucett,[2] Daniel Whiteson,[2] and Pierre Baldi[1]

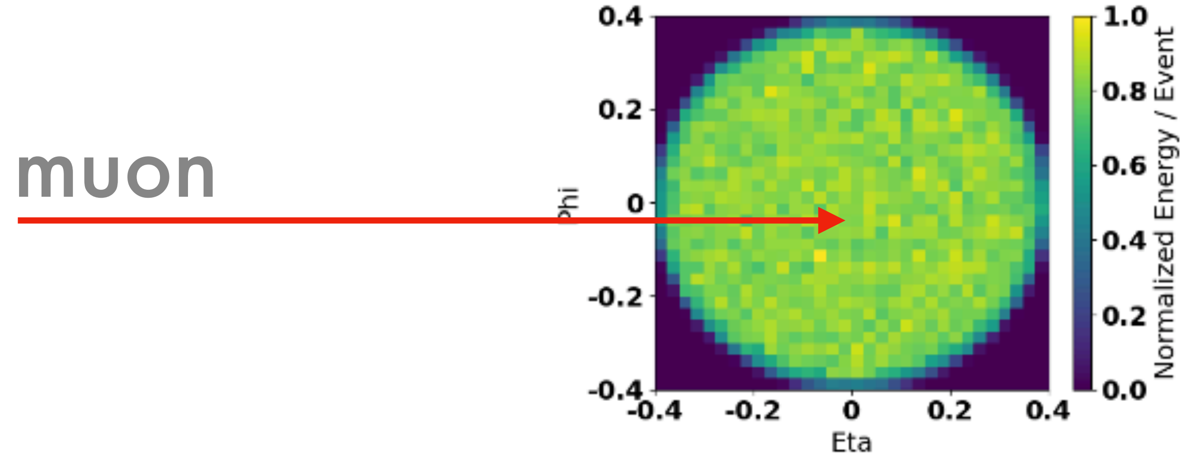
[1] *Department of Computer Science, University of California, Irvine, CA, 92697*

[2] *Department of Physics and Astronomy, University of California, Irvine, CA 92697*

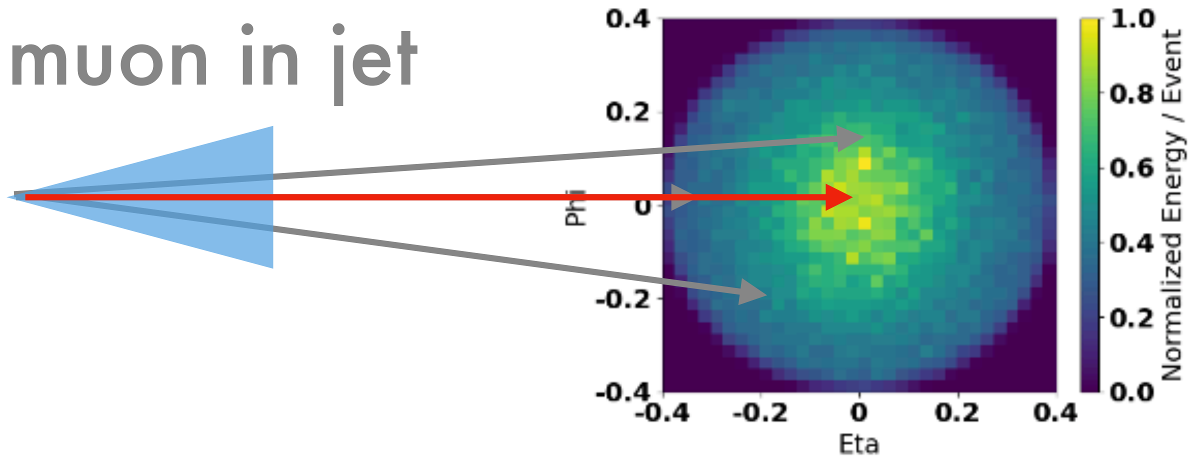
(Dated: February 2, 2021)

2102.02278

Muons



(a) Mean Prompt Muon

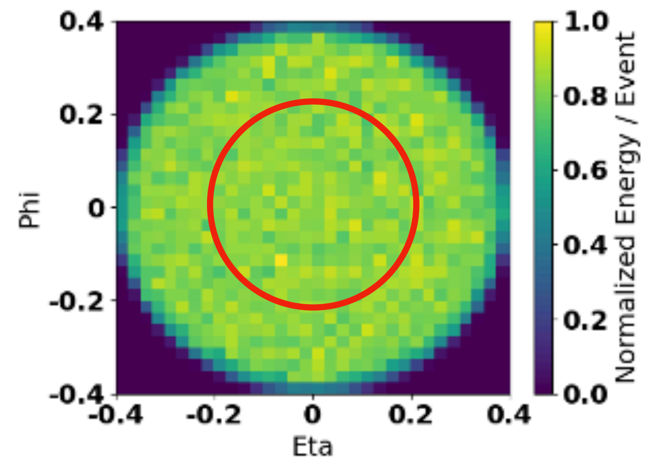


(b) Mean Non-prompt Muon

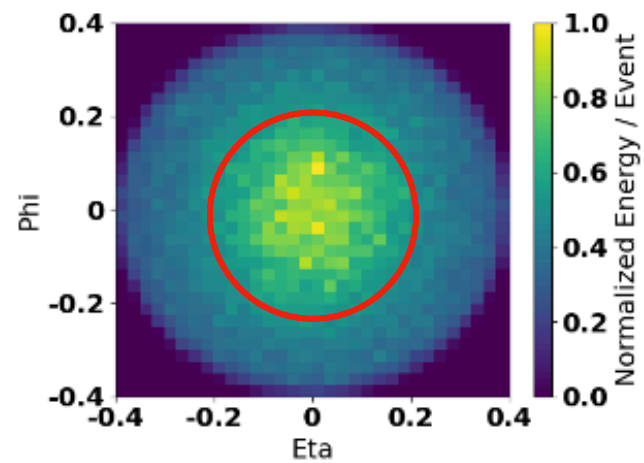
Isolation cones

Standard approach: isolation cone

$$I_\mu(R_0) = \sum_{i, R<R_0} \frac{p_{\mathrm{T}}^{\text{cell } i}}{p_{\mathrm{T}}^{\text{muon}}}$$

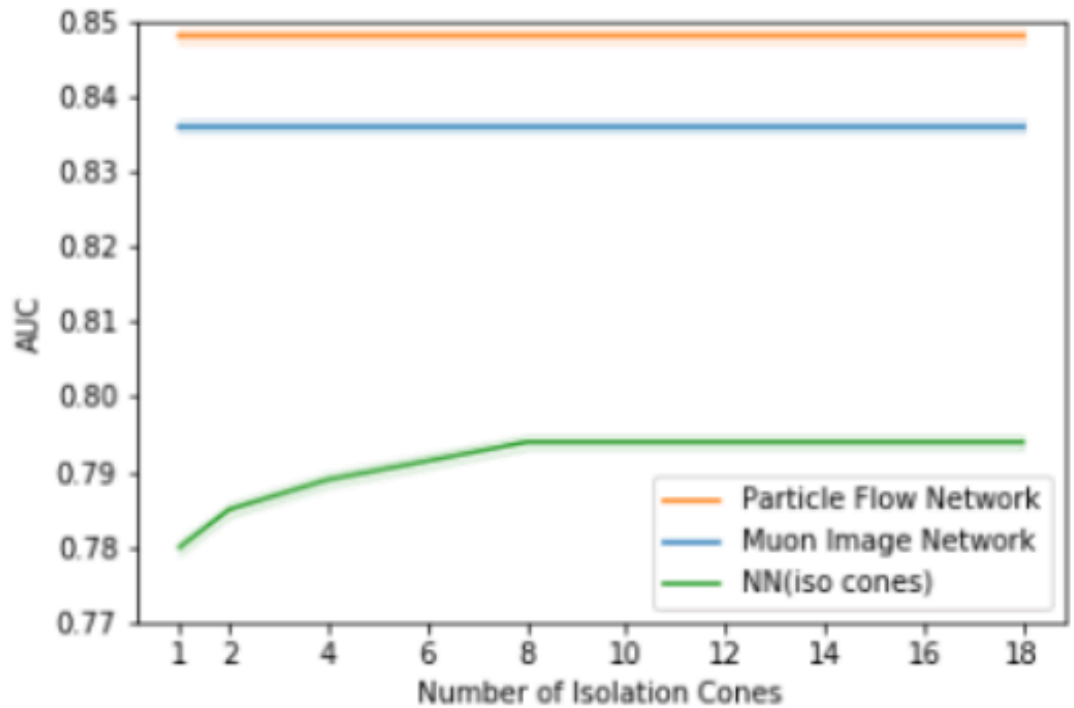


(a) Mean Prompt Muon



(b) Mean Non-prompt Muon

Results



Calo-cell networks

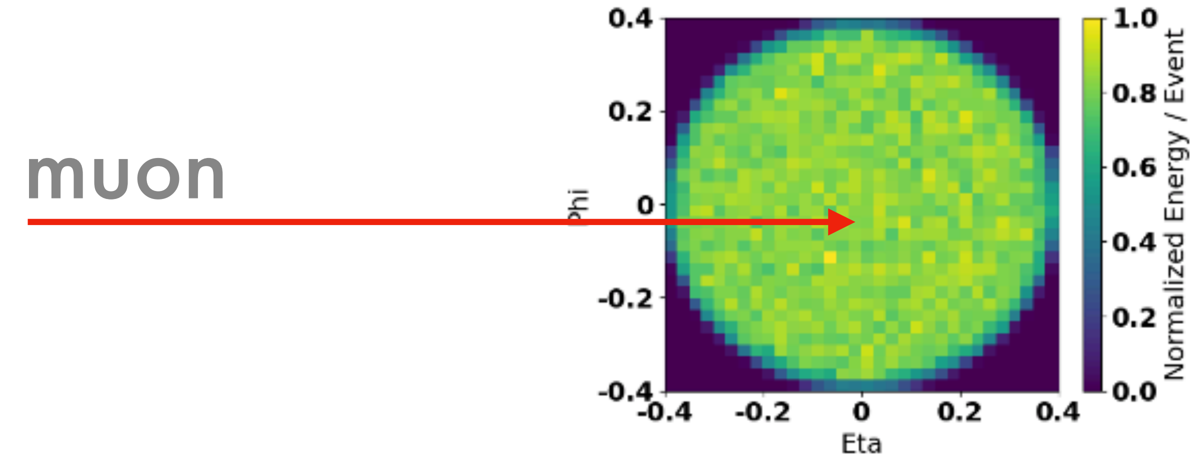
Iso networks

More iso cones improves performance
Isolation cannot match calo-cell networks

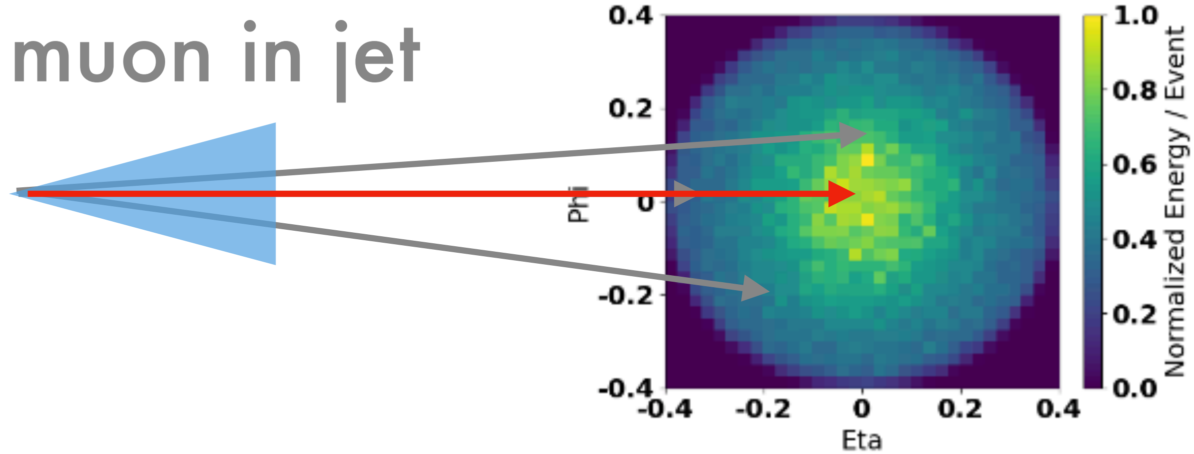
Muons

Could there be non-radial information relevant?

Jets have complex structures!



(a) Mean Prompt Muon

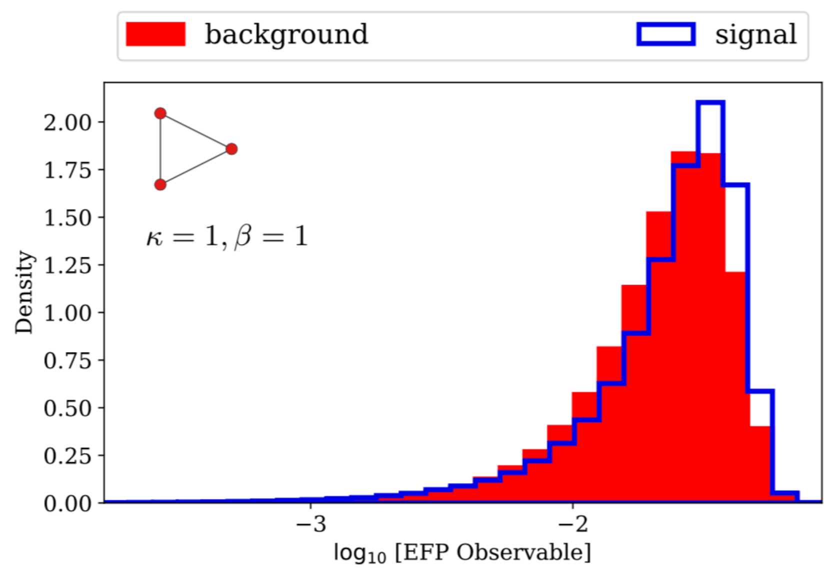


(b) Mean Non-prompt Muon

Useful observable

This observable helps!

$$= \sum_{a,b,c=1}^{N} z_a z_b z_c \theta_{ab} \theta_{bc} \theta_{ca}$$



An open gap

Adding one EFP helps

Many EFPs don't close the gap!

Method	AUC	ADO[CNN]
Single Iso Cone	0.780	0.865
8 Iso	0.794	0.878
8 Iso + $\sum p_T$ + 1 IRC-safe EFPs	0.813	0.897
8 Iso + $\sum p_T$ + 4 IRC-safe EFPs	0.821	0.908
8 Iso + $\sum p_T$ + 10 IRC-unsafe EFPs	0.827	0.923
Calo image CNN	0.836	1
Calo cell Energy-Flow Net	0.843	0.946
Calo cell Particle-Flow Net	0.848	0.948

They don't know where the muon is
can't calculate angle relative to muon.
Needs a new class of EFP.

Conclusions

Deep networks can identify gaps
  where low-level data contains unused info

Mapping strategies can interpret
  capture performance in interpretable obs.

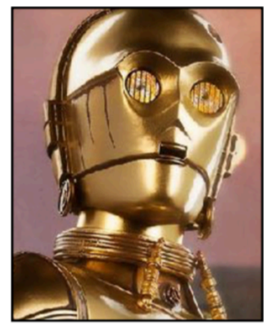
Collaborators

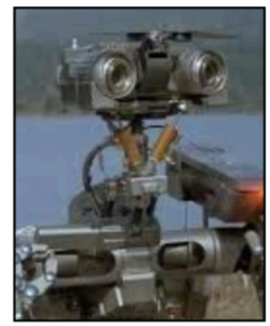
UCI Department of Physics & Astronomy

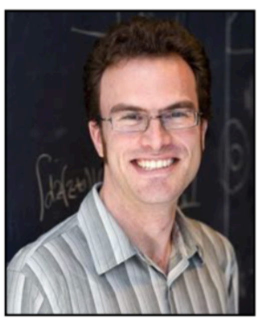
Taylor Faucett

Daniel Whiteson

The Machines







Jesse Thaler

MIT
DEPARTMENT OF PHYSICS