Simple, Interpretable Anomaly Detectors

Layne Bradshaw with Spencer Chang & Bryan Ostdiek

Based on arXiv: 2203.01343

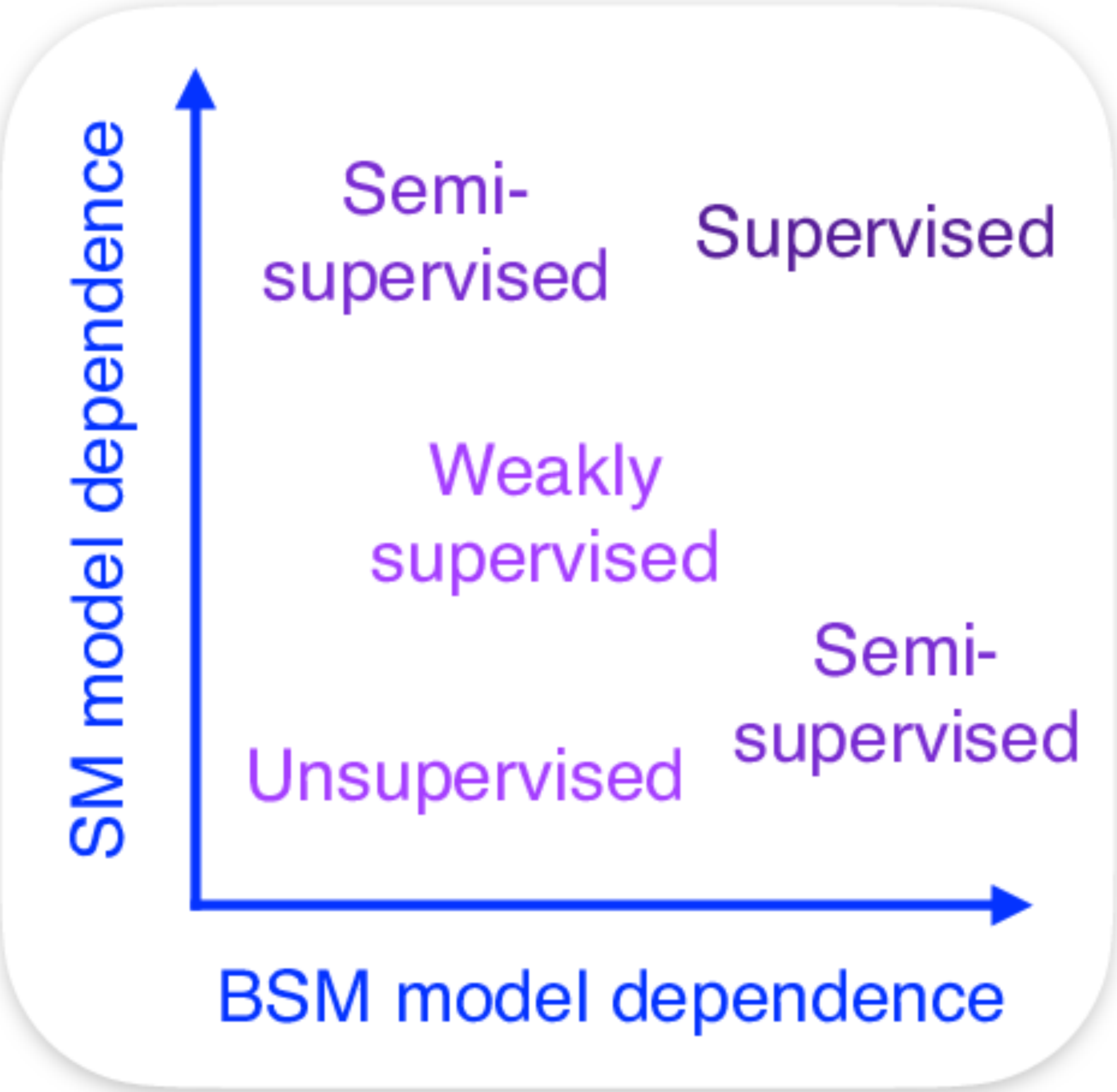
Phenomenology 2022 Symposium

Introduction

- It could be that new physics at the LHC is hiding in places we haven’t looked.

Introduction

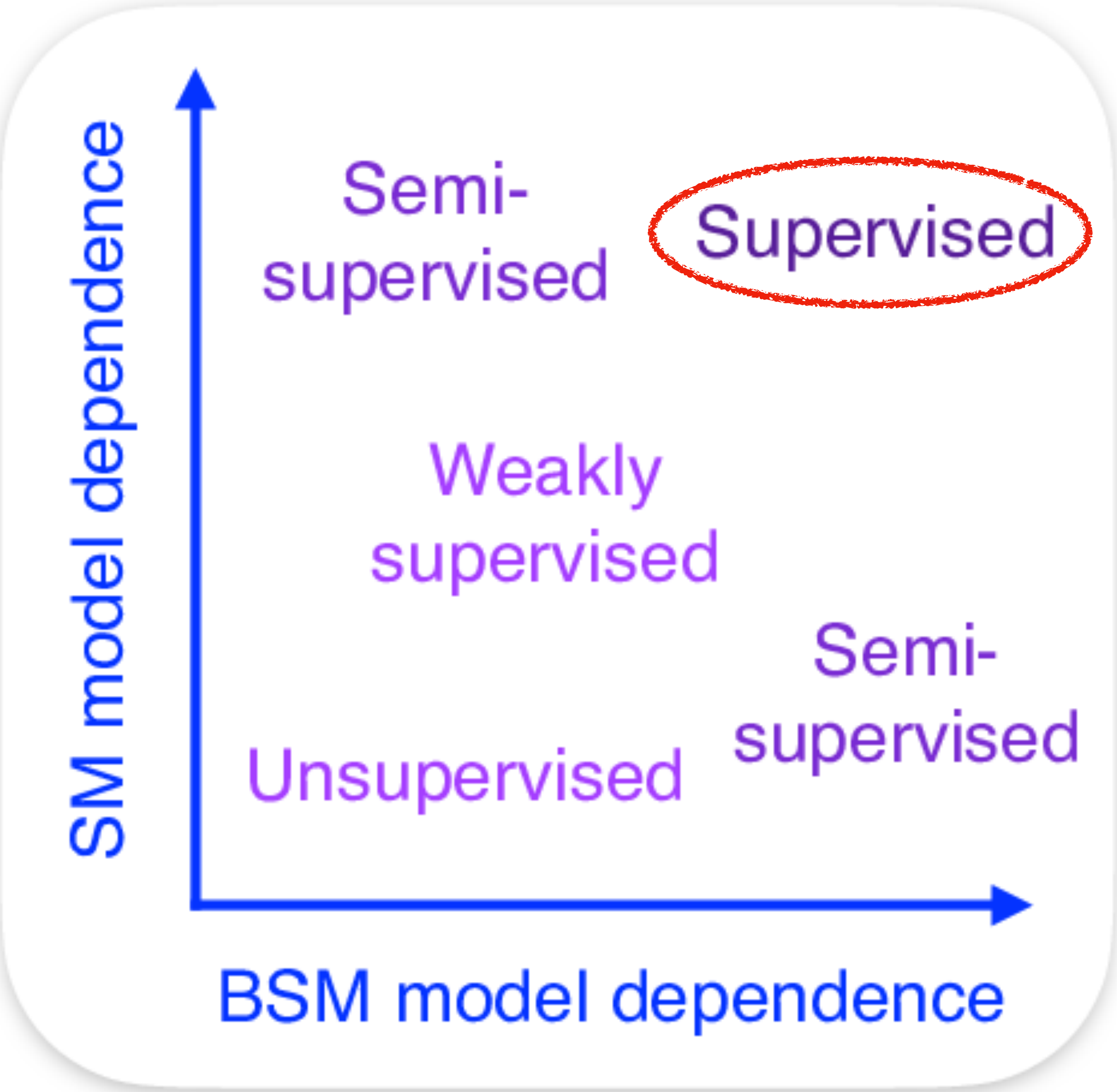
- It could be that new physics at the LHC is hiding in places we haven't looked.



From 2112.03769

Introduction

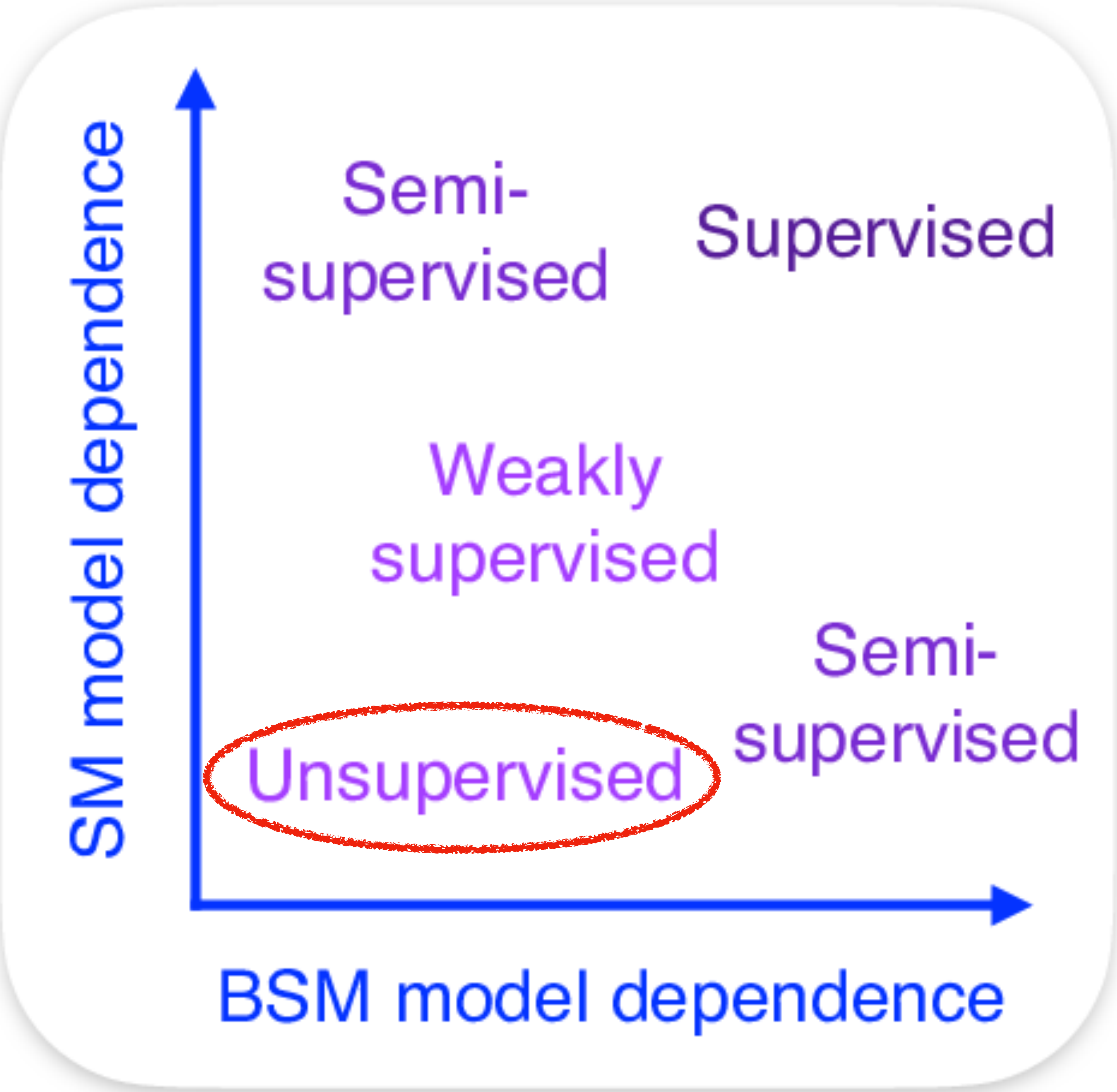
- It could be that new physics at the LHC is hiding in places we haven't looked.



From 2112.03769

Introduction

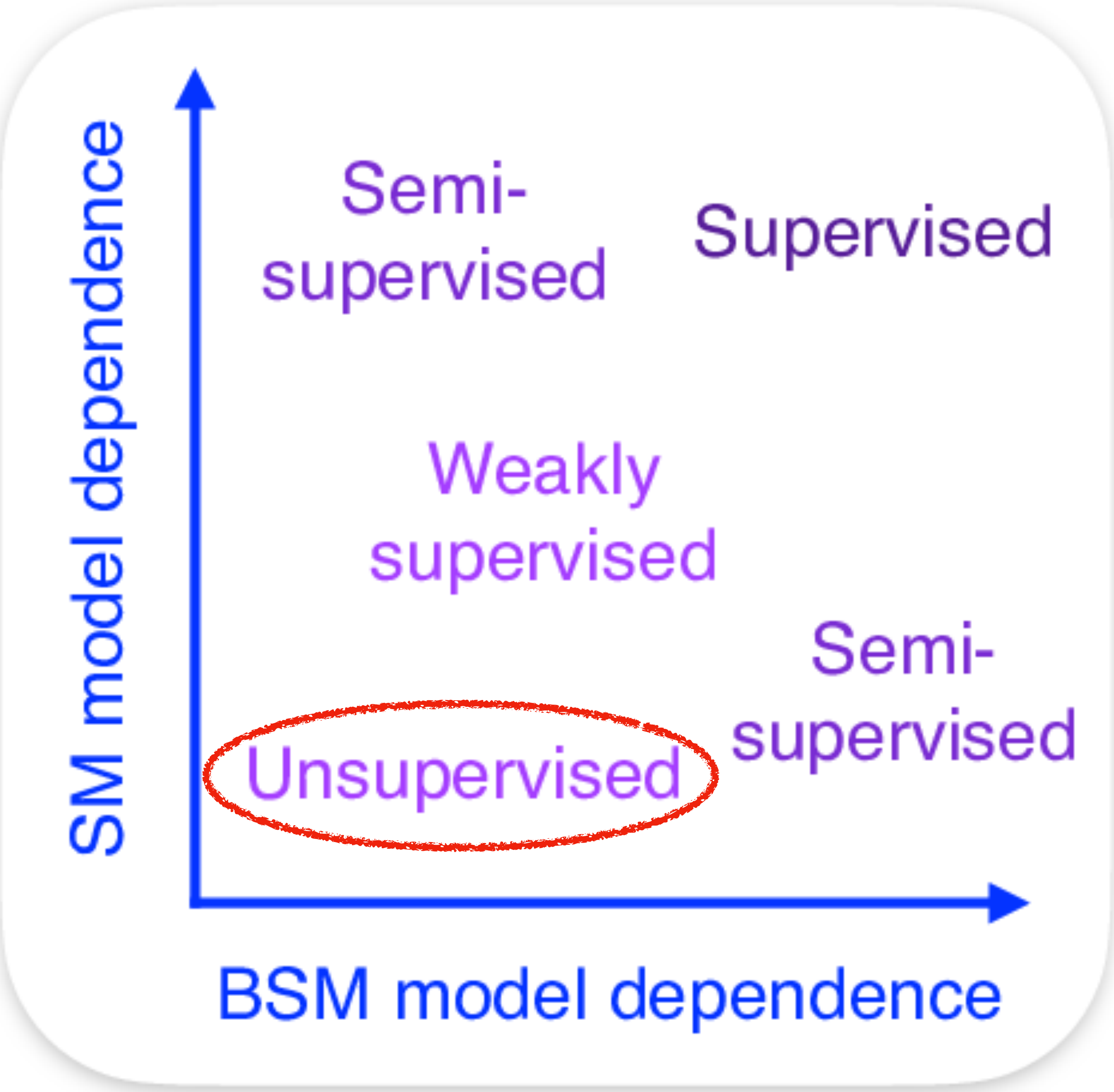
- It could be that new physics at the LHC is hiding in places we haven't looked.
- Want to design broader, model agnostic searches.



From 2112.03769

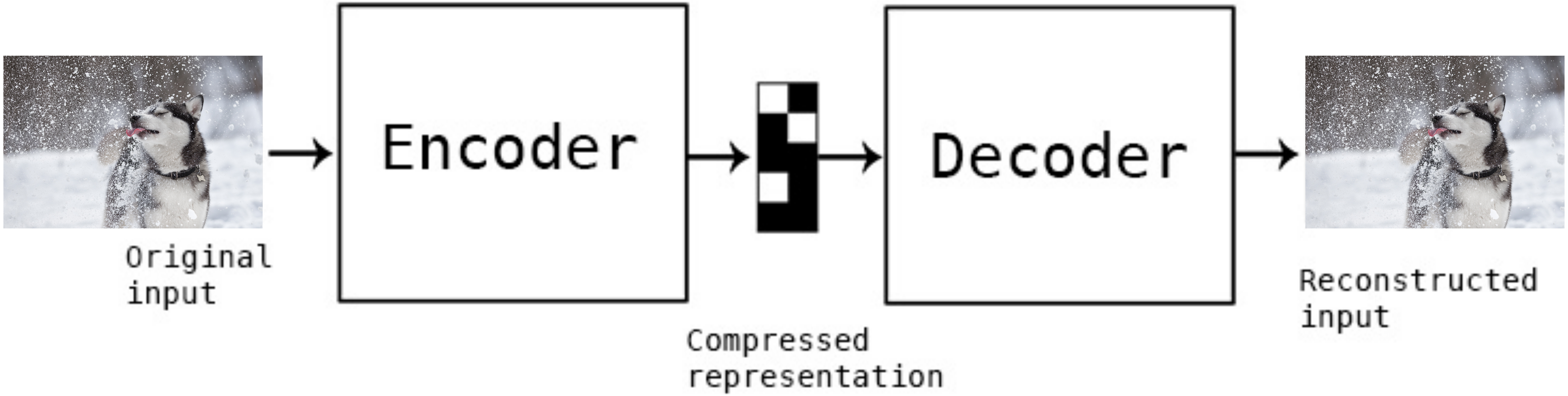
Introduction

- It could be that new physics at the LHC is hiding in places we haven’t looked.
- Want to design broader, model agnostic searches.
- Anomaly detection is a popular unsupervised method.



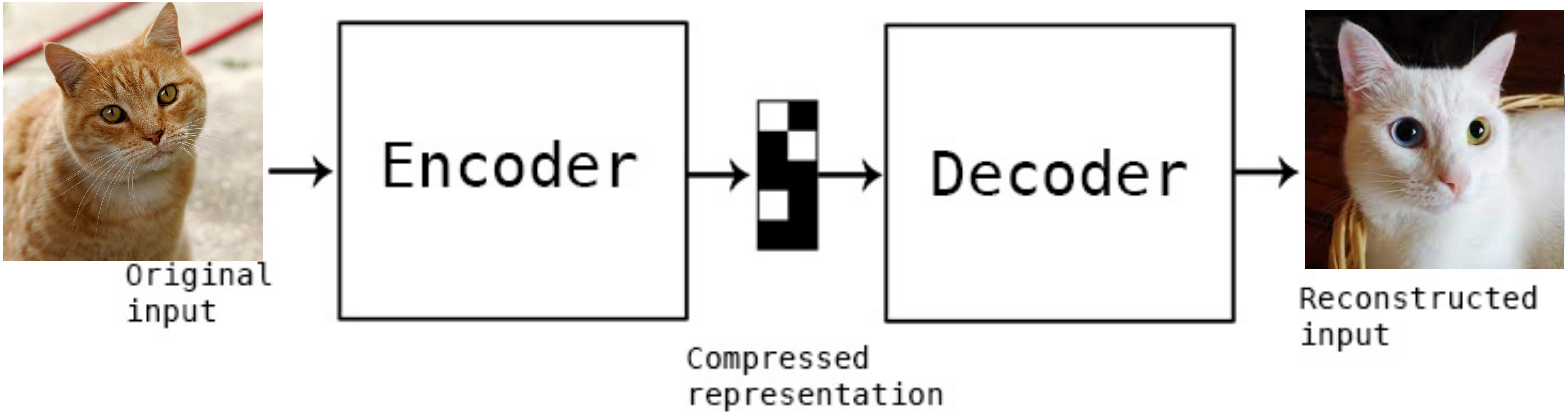
From 2112.03769

Anomaly Detection with Convolutional Autoencoders



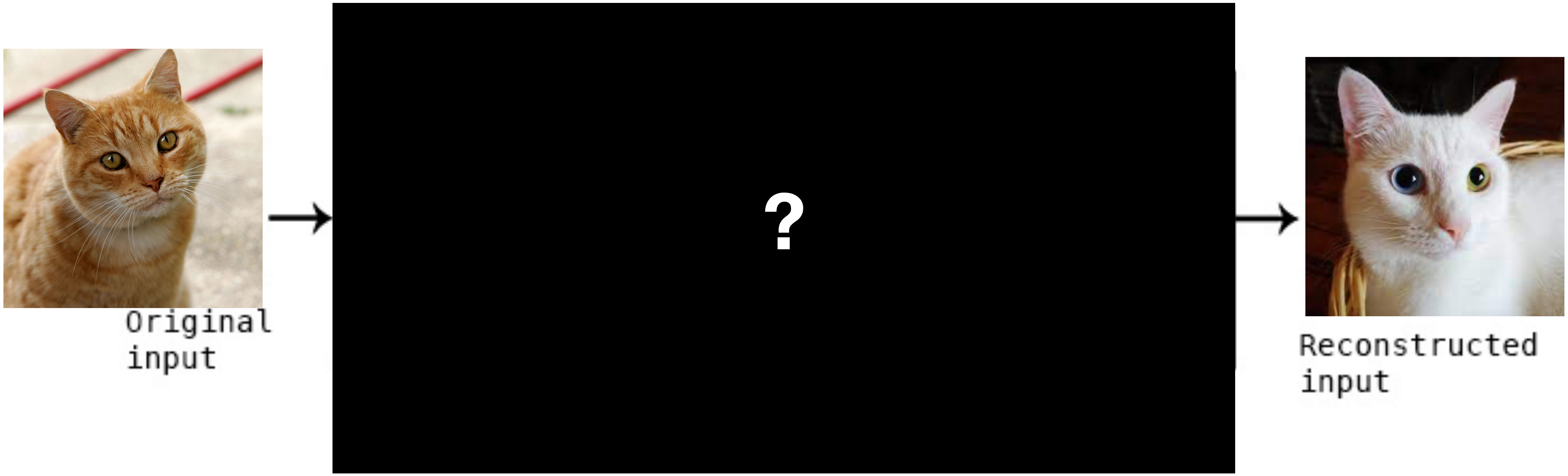
https://blog.keras.io/building-autoencoders-in-keras.html

Anomaly Detection with Convolutional Autoencoders



https://blog.keras.io/building-autoencoders-in-keras.html

Anomaly Detection with Convolutional Autoencoders



https://blog.keras.io/building-autoencoders-in-keras.html

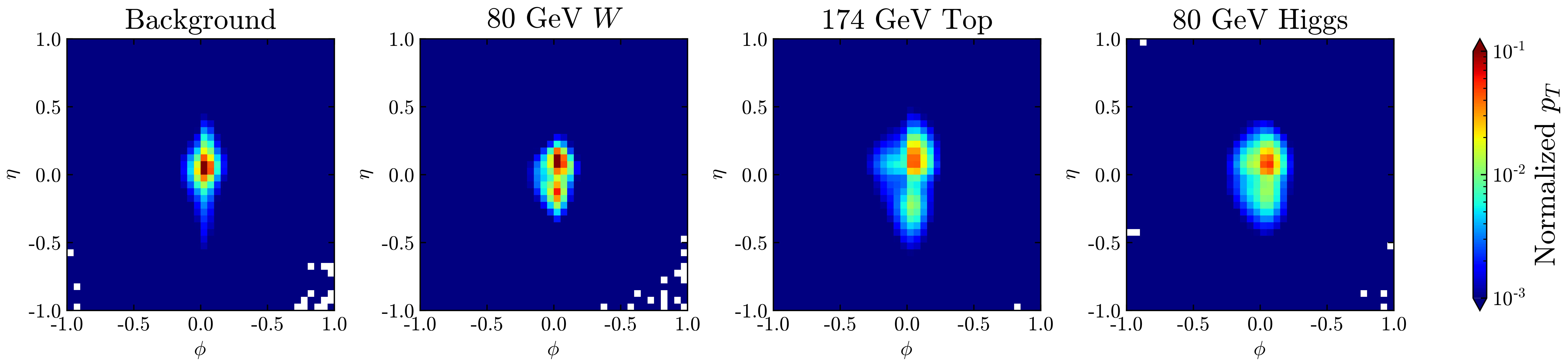
Simulated Dataset
From arXiv: 2007.01850

- Background: $pp \to jj$
- W-like signals: $pp \to W' \to WZ, W \to jj, Z \to \nu\bar{\nu}$ with $m_{W'} = 1.2 \text{ TeV}, m_W \in \{59,80,120,174\} \text{ GeV}$
- Top-like signals: $pp \to Z' \to t\bar{t}$ with $m_{Z'} = 1.3 \text{ TeV}, m_t \in \{80,174\} \text{ GeV}$
- Higgs-like signals: $pp \to HH, H \to hh, h \to jj$ with $m_H = 174$ GeV, $m_h \in \{20,80\} \text{ GeV}$

Simulated Dataset

From arXiv: 2007.01850

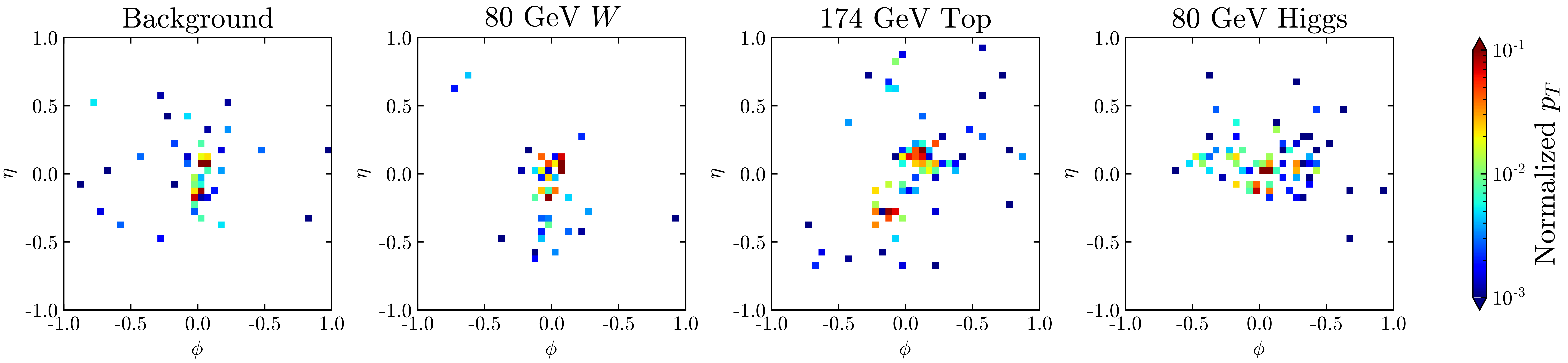
- Background: $pp \to jj$
- W-like signals: $pp \to W' \to WZ, W \to jj, Z \to \nu\bar{\nu}$ with $m_{W'} = 1.2\ \text{TeV}, m_W \in \{59,80,120,174\}\ \text{GeV}$
- Top-like signals: $pp \to Z' \to t\bar{t}$ with $m_{Z'} = 1.3\ \text{TeV}, m_t \in \{80,174\}\ \text{GeV}$
- Higgs-like signals: $pp \to HH, H \to hh, h \to jj$ with $m_H = 174$ GeV, $m_h \in \{20,80\}\ \text{GeV}$



Simulated Dataset

From arXiv: 2007.01850

- Background: $pp \rightarrow jj$
- W-like signals: $pp \rightarrow W' \rightarrow WZ, W \rightarrow jj, Z \rightarrow \nu\bar{\nu}$ with $m_{W'} = 1.2\ \text{TeV}, m_W \in \{59{,}80{,}120{,}174\}\ \text{GeV}$
- Top-like signals: $pp \rightarrow Z' \rightarrow t\bar{t}$ with $m_{Z'} = 1.3\ \text{TeV}, m_t \in \{80{,}174\}\ \text{GeV}$
- Higgs-like signals: $pp \rightarrow HH, H \rightarrow hh, h \rightarrow jj$ with $m_H = 174$ GeV, $m_h \in \{20{,}80\}\ \text{GeV}$



Iterative Matching Procedure

Method Inspired by Faucett, Thaler, Whiteson [2010.11998]

Train a Neural Network, NN_n, on a set of inputs, X_n

Iterative Matching Procedure

Method Inspired by Faucett, Thaler, Whiteson [2010.11998]

Train a Neural Network, NN_n, on a set of inputs, X_n

Quantify how well NN_n matches the Autoencoder

Iterative Matching Procedure

Method Inspired by Faucett, Thaler, Whiteson [2010.11998]

Train a Neural Network, NN_n, on a set of inputs, X_n

Quantify how well NN_n matches the Autoencoder

Find the “next best” observable

Iterative Matching Procedure

Method Inspired by Faucett, Thaler, Whiteson [2010.11998]

Train a Neural Network, NN_n, on a set of inputs, X_n

Quantify how well NN_n matches the Autoencoder

Find the “next best” observable

Add the “next best” observable to X_n

Iterative Matching Procedure

Method Inspired by Faucett, Thaler, Whiteson [2010.11998]

Train a Neural Network, NN_n, on a set of inputs, X_n

Quantify how well NN_n matches the Autoencoder

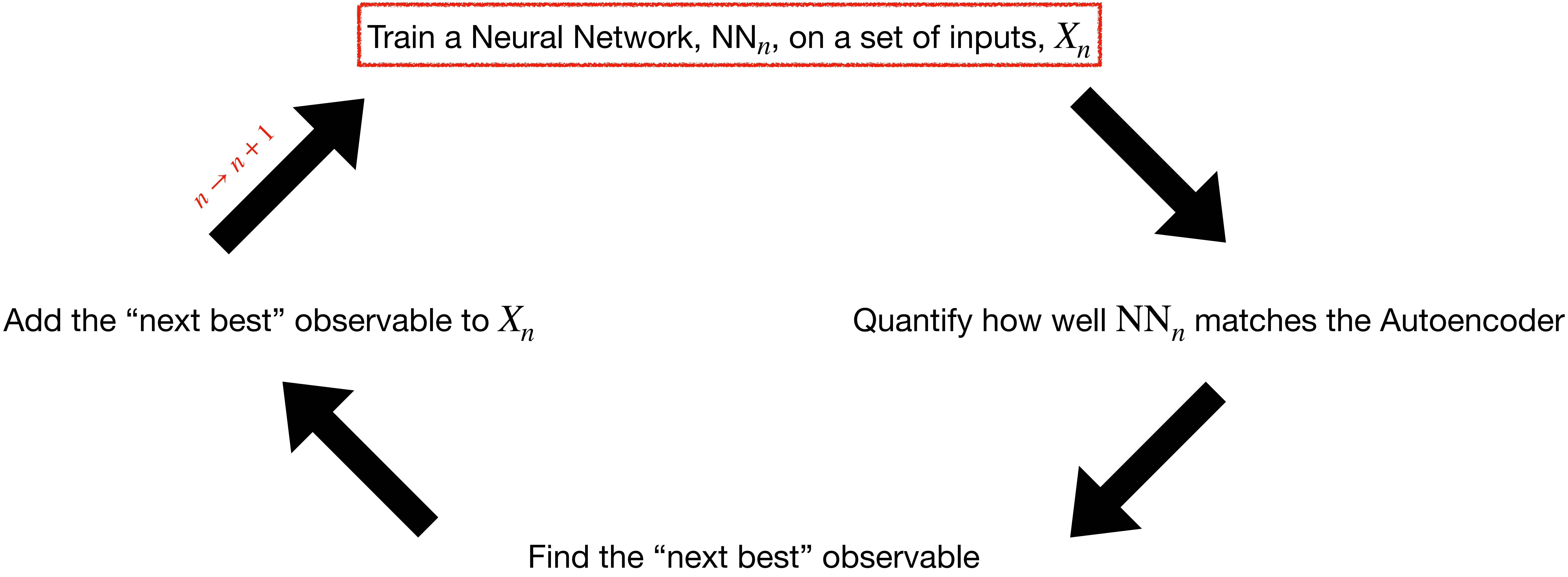
Find the “next best” observable

Add the “next best” observable to X_n

$n \to n+1$

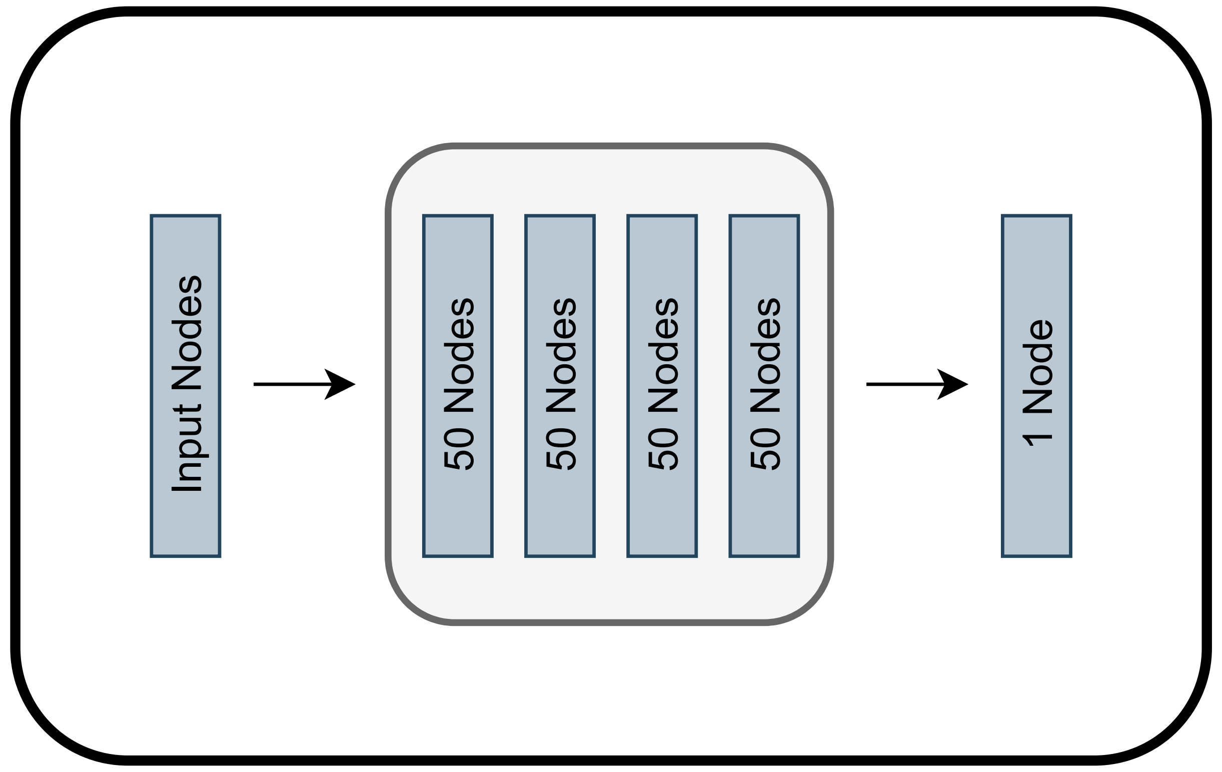
Iterative Matching Procedure

Method Inspired by Faucett, Thaler, Whiteson [2010.11998]



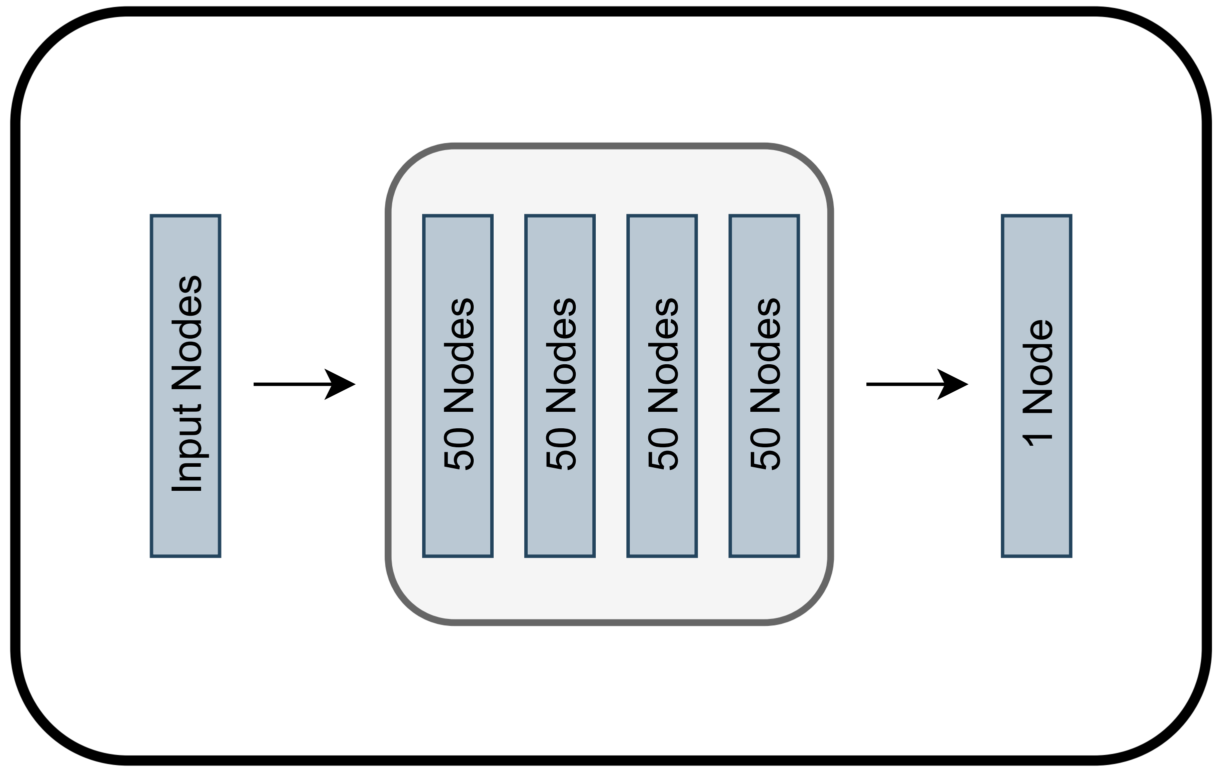
Network Architectures

High-Level Neural Network



Network Architectures

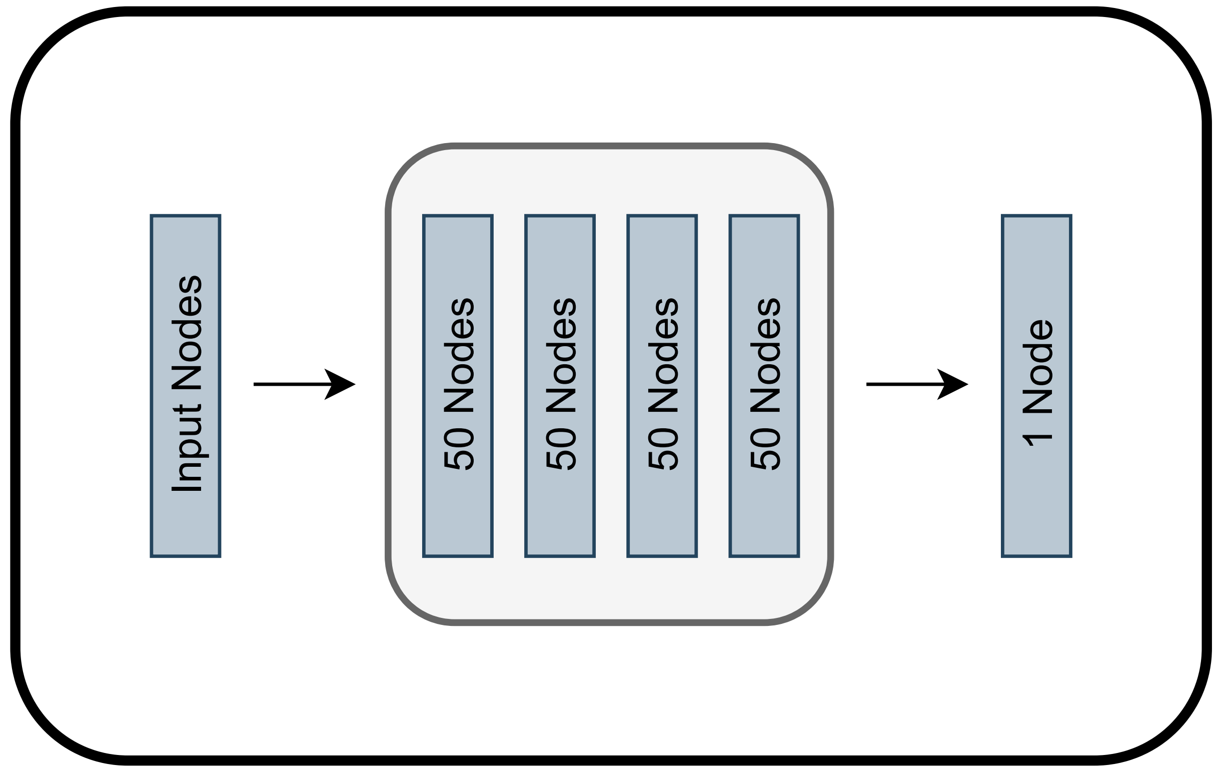
High-Level Neural Network



- Designed to regress the AE's anomaly score

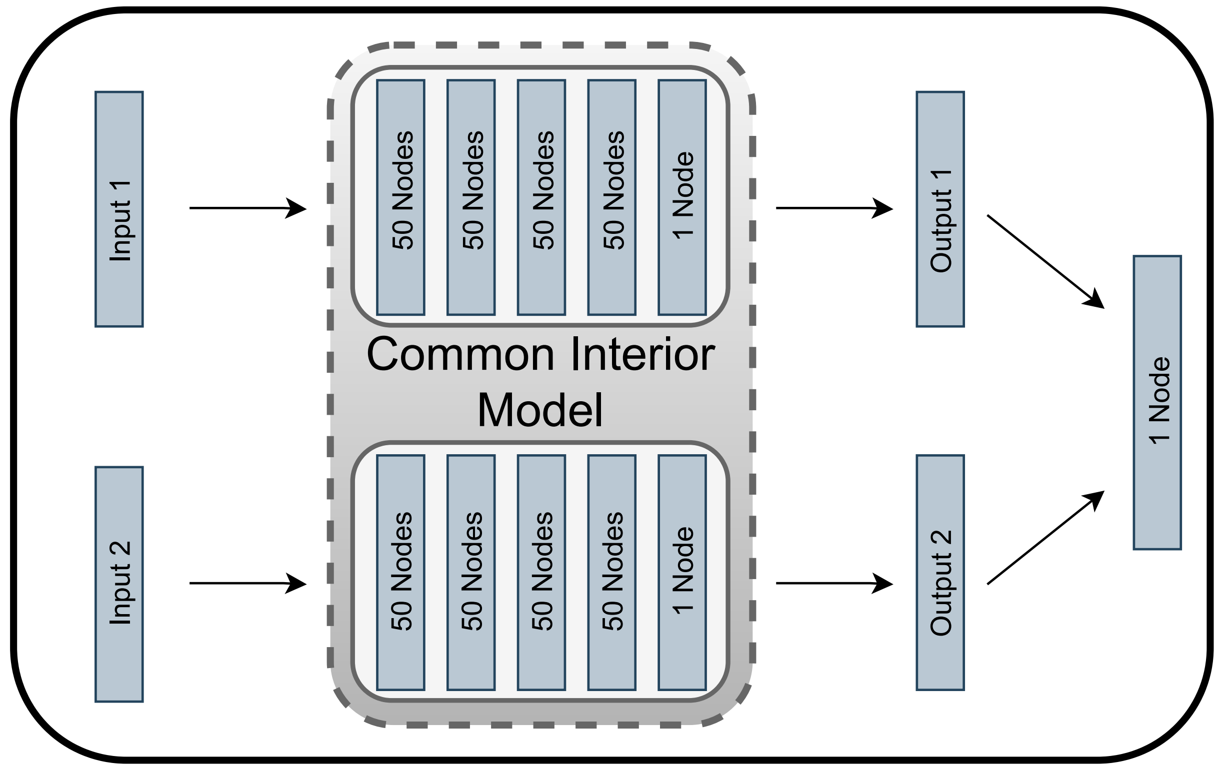
Network Architectures

High-Level Neural Network



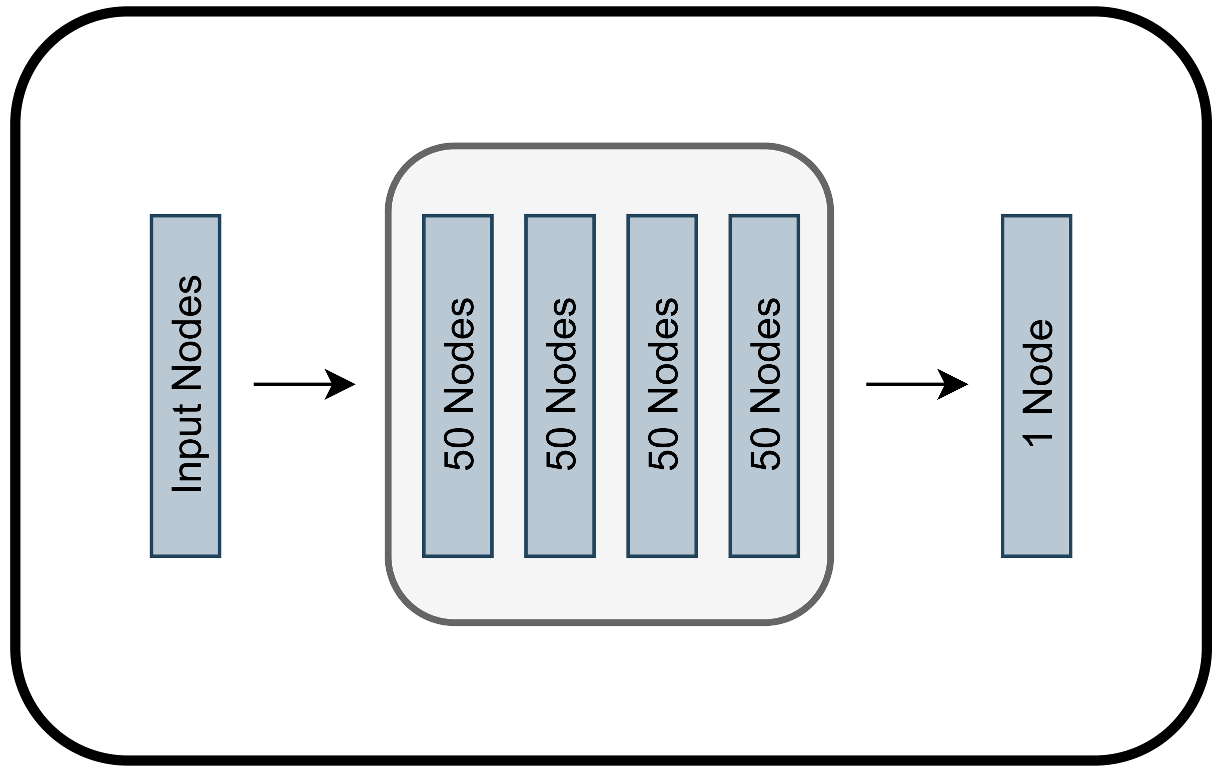
- Designed to regress the AE's anomaly score

Paired Neural Network



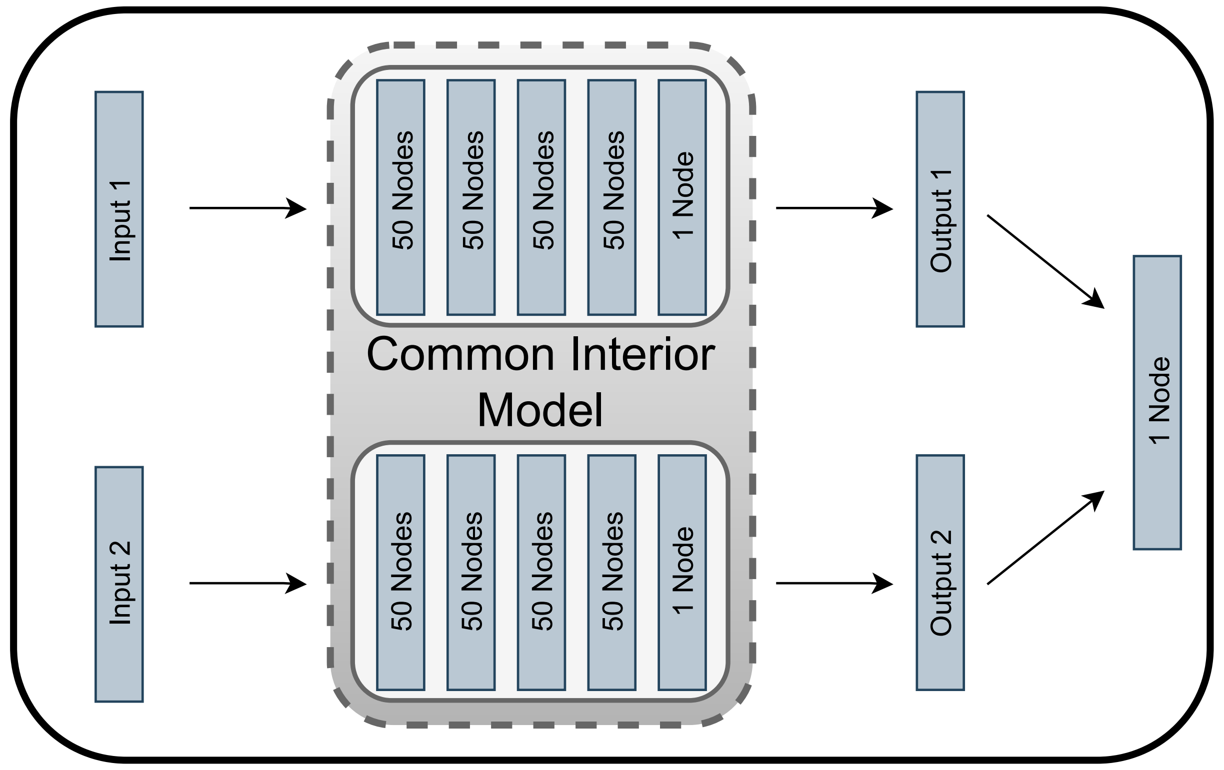
Network Architectures

High-Level Neural Network



- Designed to regress the AE's anomaly score

Paired Neural Network



- Designed to learn which of a pair of events the AE deems to be more anomalous

Iterative Matching Procedure

Method inspired by 2010.11998

Train a Neural Network, NN_n, on a set of inputs, X_n

$n \to n+1$

Add the “next best” observable to X_n

$\boxed{\text{Quantify how well } \mathrm{NN}_n \text{ matches the Autoencoder}}$

Find the “next best” observable

Decision Ordering

- For a pair of events x_1 and x_2, we say two networks have the same *Decision Ordering* if they agree on which event is more anomalous.

Decision Ordering

- For a pair of events x_1 and x_2, we say two networks have the same *Decision Ordering* if they agree on which event is more anomalous.
- We can then average over all possible pairs of events to give us a summary statistic, the *Average Decision Ordering* (ADO).
- An ADO of 1 corresponds to one network ordering all events in exactly the same way as another, an ADO of 0.5 means there is no consistency in how one network orders events relative to another.

Iterative Matching Procedure

Method inspired by 2010.11998

Train a Neural Network, NN_n, on a set of inputs, X_n

Quantify how well NN_n matches the Autoencoder

$\boxed{\text{Find the “next best” observable}}$

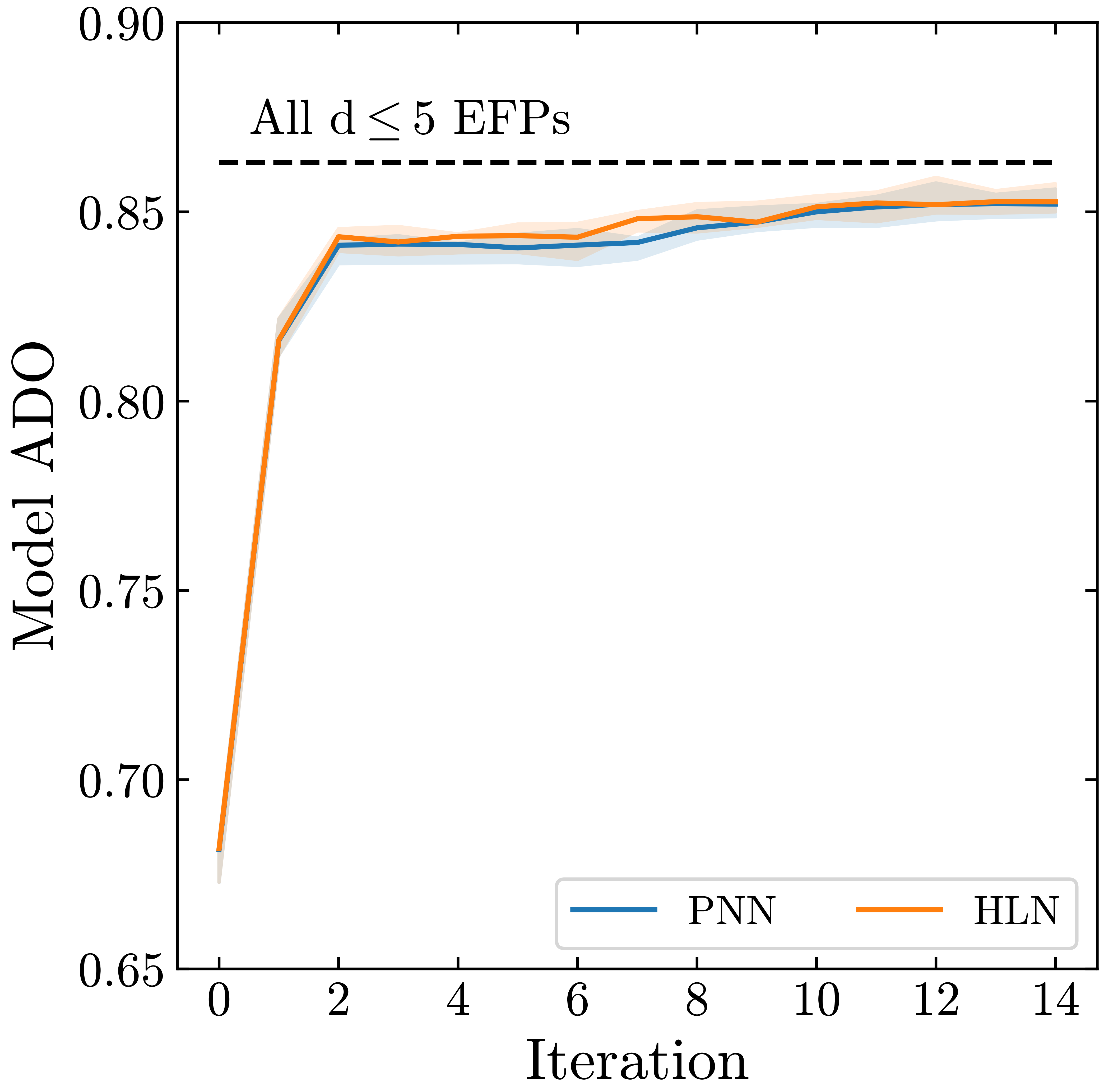
Add the “next best” observable to X_n

$n \rightarrow n+1$

Finding the “Next Best” Observable

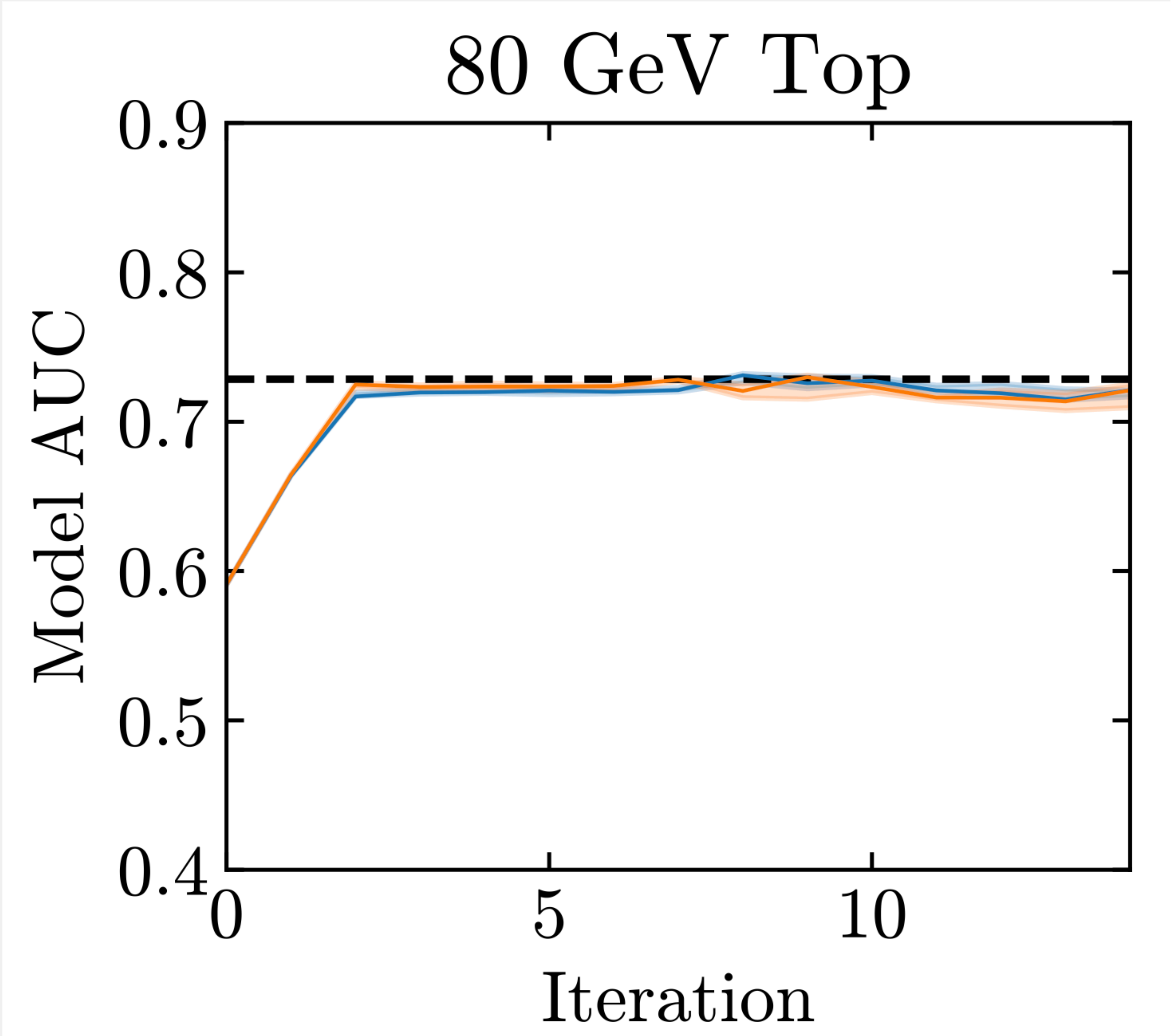
- Our set of observables are the *Energy Flow Polynomials*, a formally infinite set of jet substructure observables that form a discrete linear basis for all IRC safe observables. [arXiv: 1712.07124]
- Generalization of Energy Correlators, built on sums of momenta fractions and powers of angular distances.
- The EFP with the highest ADO on the pairs of events *misordered by* NN_n is the “next best” observable, and is added to our list of inputs.

Model ADOs

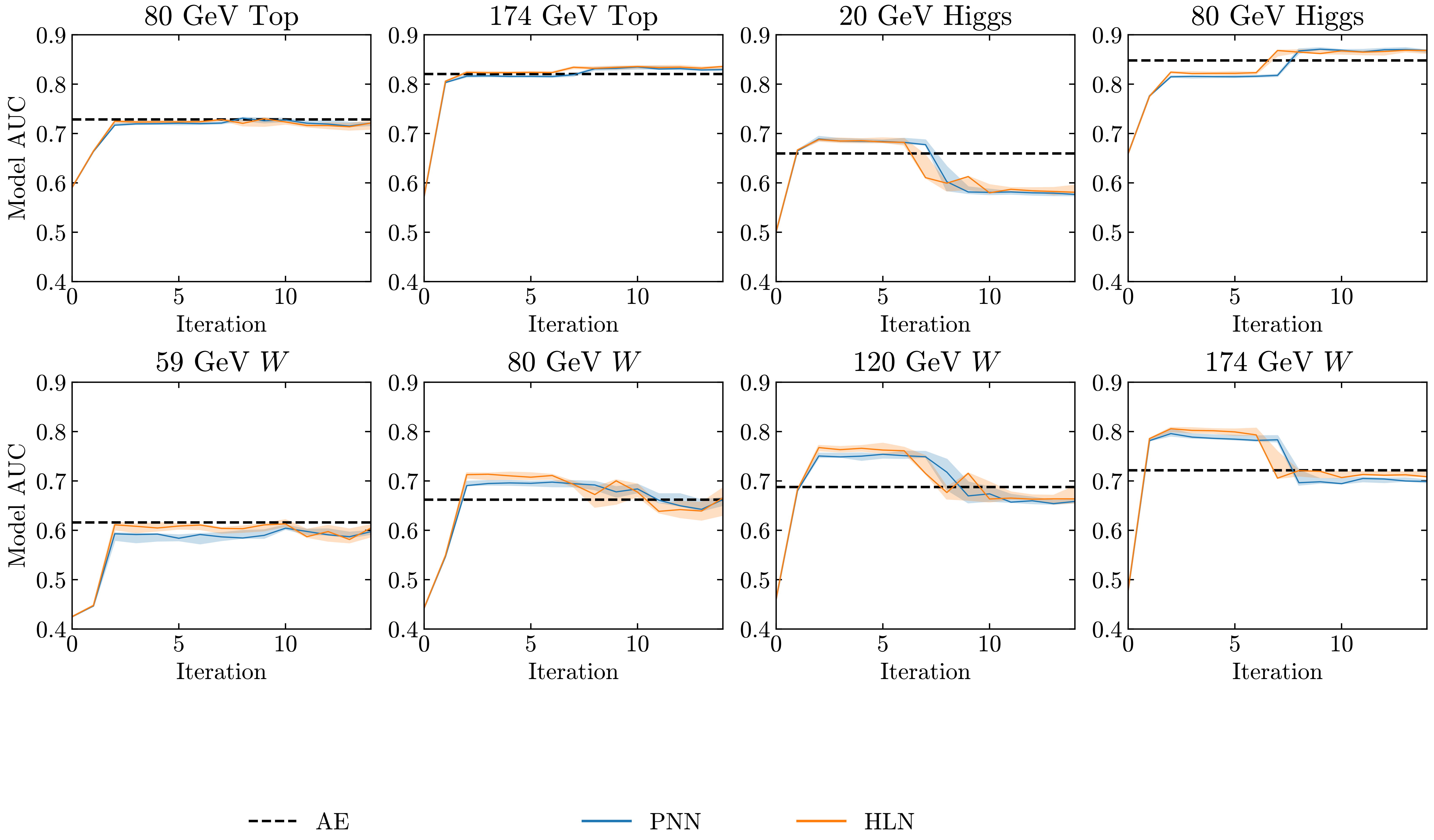


All d ≤ 5 EFPs
Model ADO
Iteration
0.90
0.85
0.80
0.75
0.70
0.65
0
2
4
6
8
10
12
14
PNN
HLN

Model AUCs



Model AUCs



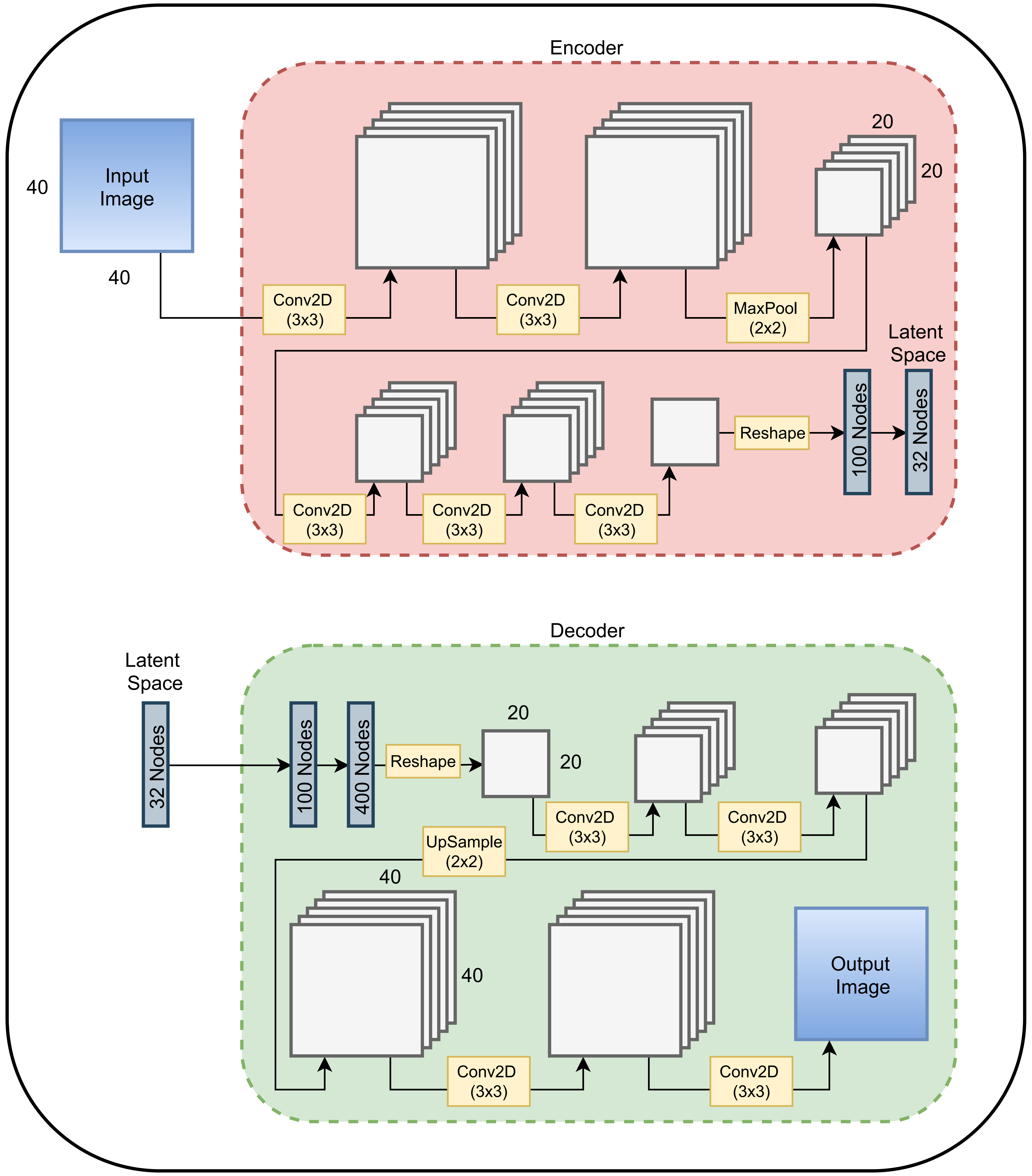
80 GeV Top
174 GeV Top
20 GeV Higgs
80 GeV Higgs
59 GeV W
80 GeV W
120 GeV W
174 GeV W
Model AUC
Iteration
0.4
0.5
0.6
0.7
0.8
0.9
0
5
10
AE
PNN
HLN

Conclusion and Future Work

- Simple architectures and inputs can be used to match the decision orderings of a much more complex anomaly detector.
- Learning to correctly order background events transfers to correctly ordering a variety of signal events.
- Future work: How can we get an ADO closer to 1? More EFPs or something more complicated?
- Future work: How well does this method work with other starting anomaly detection architectures?

Backup Slides

Autoencoder Architecture



Decision Ordering

- Given two decision functions f and g, the *Decision Ordering* given a pair of events, x_1 and x_2 is:

$$\text{DO}[f, g](x_1, x_2) = \Theta\left(\left[f(x_1) - f(x_2)\right]\left[g(x_1) - g(x_2)\right]\right)$$

Decision Ordering

- Given two decision functions f and g, the *Decision Ordering* given a pair of events, x_1 and x_2 is:

$$\text{DO}[f,g](x_1,x_2) = \Theta\left(\left[f(x_1)-f(x_2)\right]\left[g(x_1)-g(x_2)\right]\right)$$

- We can then average over all possible pairs of events to give us a summary statistic, the *Average Decision Ordering:*

$$\text{ADO}[f,g] = \int \mathrm{d}x_1 \mathrm{d}x_2\, p_1(x_1)p_2(x_2)\text{DO}[f,g](x_1,x_2)$$

Energy Flow Polynomials

- Formally infinite set of jet substructure observables that form a discrete linear basis for all IRC safe observables.

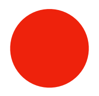
$$z_a^{(\kappa)} = \left(\frac{p_T}{\sum_{b=1}^{N} p_{T,b}} \right)^{\kappa}$$

$$\theta_{ab}^{(\beta)} = \left(\Delta\eta_{ab}^2 + \Delta\phi_{ab}^2 \right)^{\beta/2}$$

Energy Flow Polynomials

- Formally infinite set of jet substructure observables that form a discrete linear basis for all IRC safe observables.

$$z_a^{(\kappa)} = \left(\frac{p_T}{\sum_{b=1}^{N} p_{T,b}} \right)^{\kappa}$$



$$\Longleftrightarrow \sum_{a=1}^{N} z_a$$

$$\theta_{ab}^{(\beta)} = \left(\Delta\eta_{ab}^2 + \Delta\phi_{ab}^2 \right)^{\beta/2}$$

Energy Flow Polynomials

- Formally infinite set of jet substructure observables that form a discrete linear basis for all IRC safe observables.

$$z_a^{(\kappa)} = \left(\frac{p_T}{\sum_{b=1}^{N} p_{T,b}} \right)^{\kappa} \qquad \Longleftrightarrow \sum_{a=1}^{N} z_a$$

$$\theta_{ab}^{(\beta)} = \left(\Delta\eta_{ab}^2 + \Delta\phi_{ab}^2 \right)^{\beta/2} \qquad \Longleftrightarrow \left(\theta_{ab}\right)^k$$

Energy Flow Polynomials

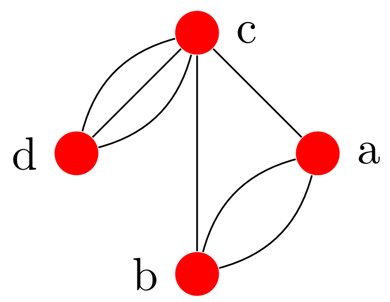
- Formally infinite set of jet substructure observables that form a discrete linear basis for all IRC safe observables.

$$z_a^{(\kappa)} = \left(\frac{p_T}{\sum_{b=1}^{N} p_{T,b}} \right)^{\kappa}$$

$$\Longleftrightarrow \sum_{a=1}^{N} z_a$$

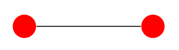
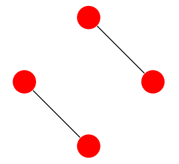
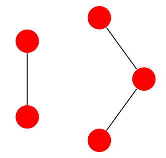
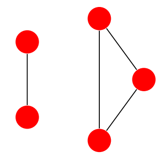
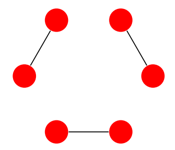
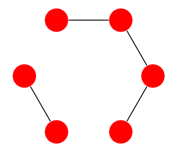
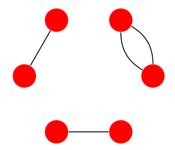
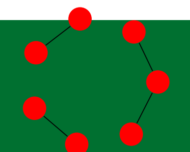
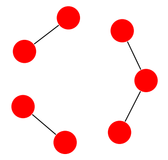
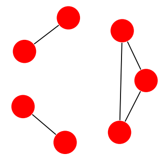
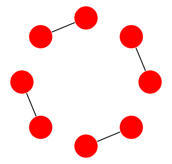
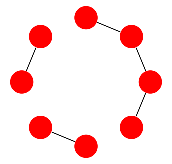
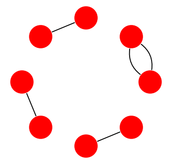
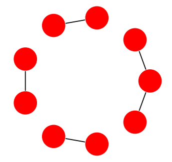
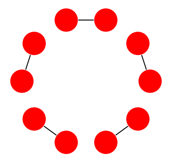
$$\theta_{ab}^{(\beta)} = \left(\Delta\eta_{ab}^2 + \Delta\phi_{ab}^2 \right)^{\beta/2}$$

$$\Longleftrightarrow \left(\theta_{ab} \right)^{k}$$



$$= \sum_{a=1}^{N} \sum_{b=1}^{N} \sum_{c=1}^{N} \sum_{d=1}^{N} z_a z_b z_c z_d \theta_{ab}^2 \theta_{ac} \theta_{bc} \theta_{cd}^3.$$

EFPs Selected

EFP No.	EFP Multigraph	EFP Expression
1		$\sum_{a,b=1}^{N} z_a z_b \theta_{ab}$
54		$\sum_{a,b,c,d=1}^{N} z_a z_b z_c z_d \theta_{ab} \theta_{cd}$
60		$\sum_{a,b,c,d,e=1}^{N} z_a z_b z_c z_d z_e \theta_{ab} \theta_{ac} \theta_{de}$
63		$\sum_{a,b,c,d,e=1}^{N} z_a z_b z_c z_d z_e \theta_{ab} \theta_{ac} \theta_{bc} \theta_{de}$
70		$\sum_{a,b,c,d,e,f=1}^{N} z_a z_b z_c z_d z_e z_f \theta_{ab} \theta_{cd} \theta_{ef}$
72		$\sum_{a,b,c,d,e,f=1}^{N} z_a z_b z_c z_d z_e z_f \theta_{ab} \theta_{bc} \theta_{cd} \theta_{ef}$
74		$\sum_{a,b,c,d,e,f=1}^{N} z_a z_b z_c z_d z_e z_f \theta_{ab}^2 \theta_{cd} \theta_{ef}$
86		$\sum^{N} z_a z_b z_c z_d z_e z_f z_g \theta_{ab} \theta \theta \theta_{fg}$
86		$\sum_{a,b,c,e,d,f,g=1}^{N} z_a z_b z_c z_d z_e z_f z_g \theta_{ab} \theta_{ac} \theta_{de} \theta_{fg}$
94		$\sum_{a,b,c,e,d,f,g=1}^{N} z_a z_b z_c z_d z_e z_f z_g \theta_{ab} \theta_{ac} \theta_{bc} \theta_{de} \theta_{fg}$
95		$\sum_{a,b,c,d,e,f,g,h=1}^{N} z_a z_b z_c z_d z_e z_f z_g z_h \theta_{ab} \theta_{cd} \theta_{ef} \theta_{gh}$
97		$\sum_{a,b,c,d,e,f,g,h=1}^{N} z_a z_b z_c z_d z_e z_f z_g z_h \theta_{ab} \theta_{bc} \theta_{cd} \theta_{ef} \theta_{gh}$
99		$\sum_{a,b,c,d,e,f,g,h=1}^{N} z_a z_b z_c z_d z_e z_f z_g z_h \theta_{ab}^2 \theta_{cd} \theta_{ef} \theta_{gh}$
100		$\sum_{a,b,c,d,e,f,g,h,i=1}^{N} z_a z_b z_c z_d z_e z_f z_g z_h z_i \theta_{ab} \theta_{ac} \theta_{de} \theta_{fg} \theta_{hi}$
101		$\sum_{a,b,c,d,e,f,g,h,i,j=1}^{N} z_a z_b z_c z_d z_e z_f z_g z_h z_i z_j \theta_{ab} \theta_{cd} \theta_{ef} \theta_{gh} \theta_{ij}$