Re-simulation-based self-supervised learning (RS3L)

6th IML Workshop
January 30, 2024

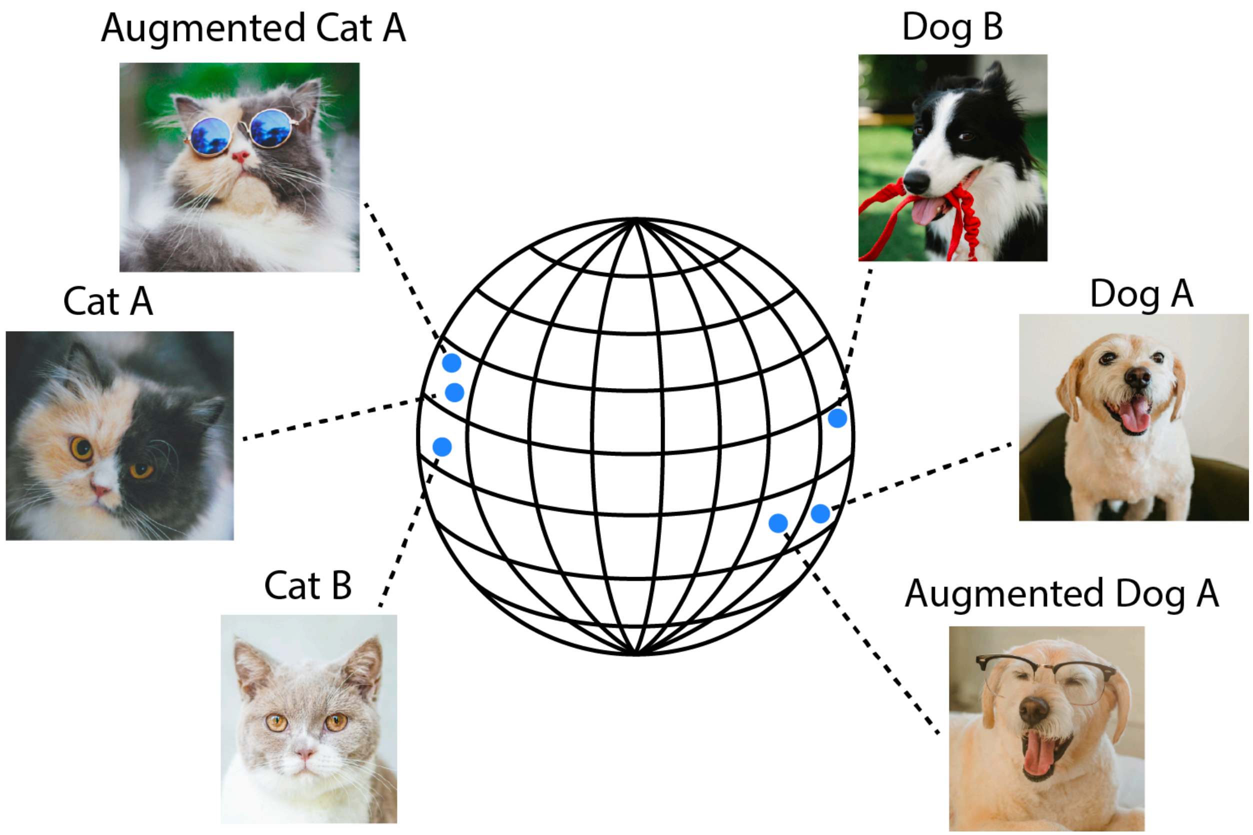
Jeffrey Krupa[1,2], Benedikt Maier[3], Michael Kagan[4], Nathaniel Woodward[1,2], Philip Harris[1,2], Maurizio Pierini[5]

Karlsruher Institut für Technologie

The NSF Institute for Artificial Intelligence and Fundamental Interactions

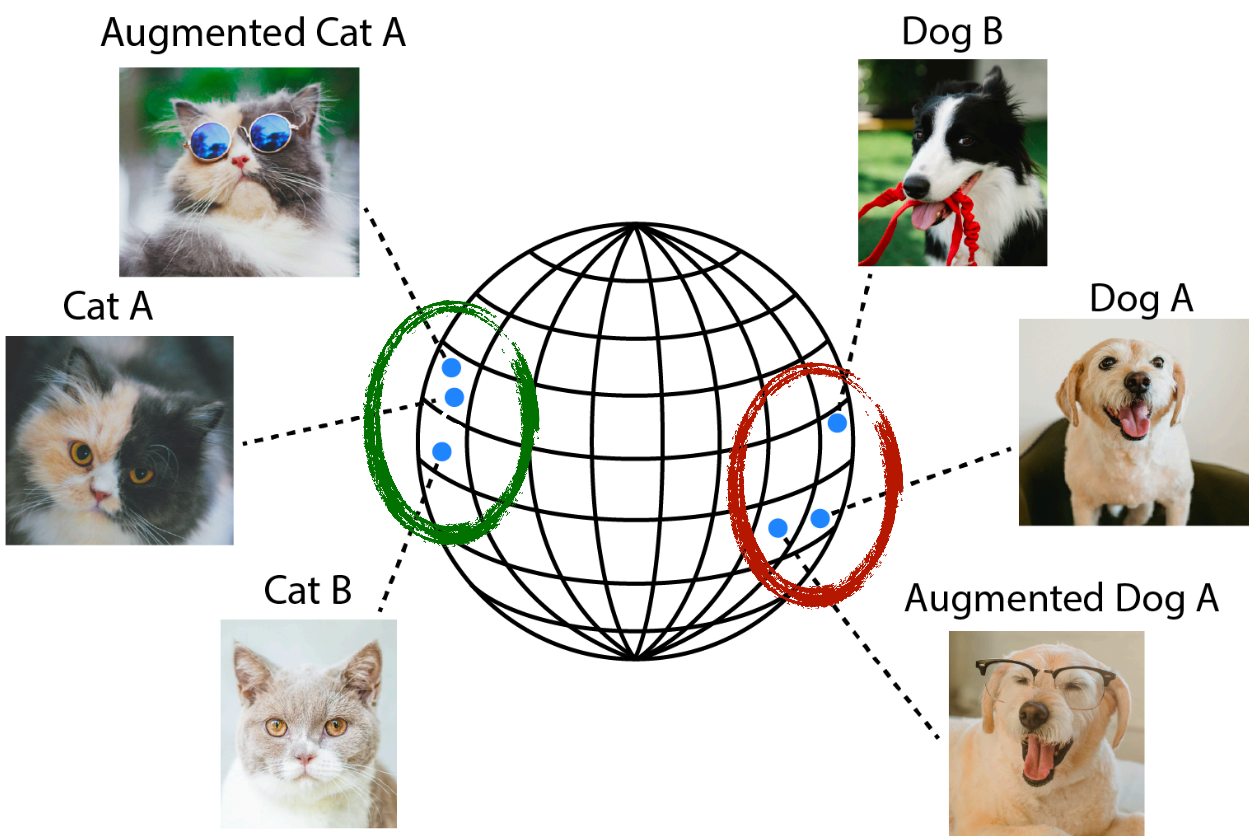
Representations

- Can we use AI to learn generic representations of jets?



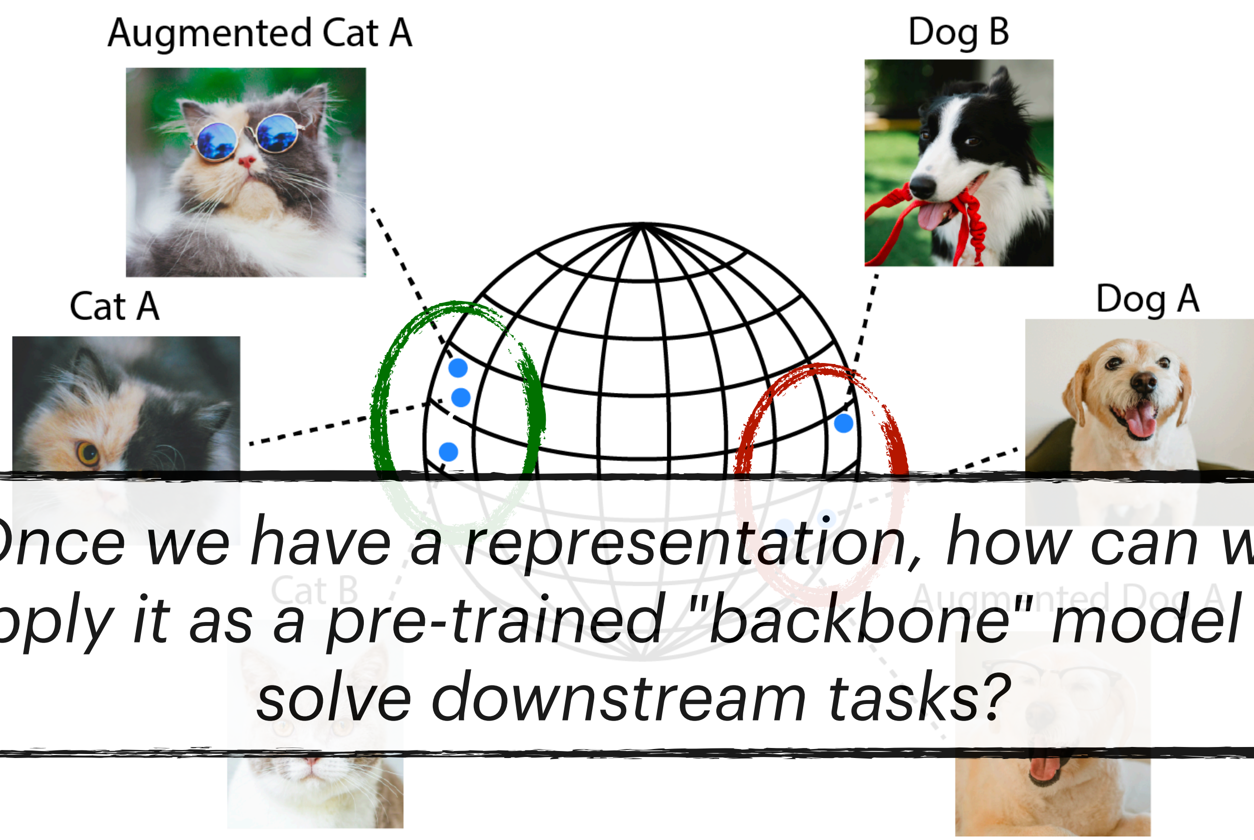
Representations

- Can we use AI to learn generic representations of jets?



Representations

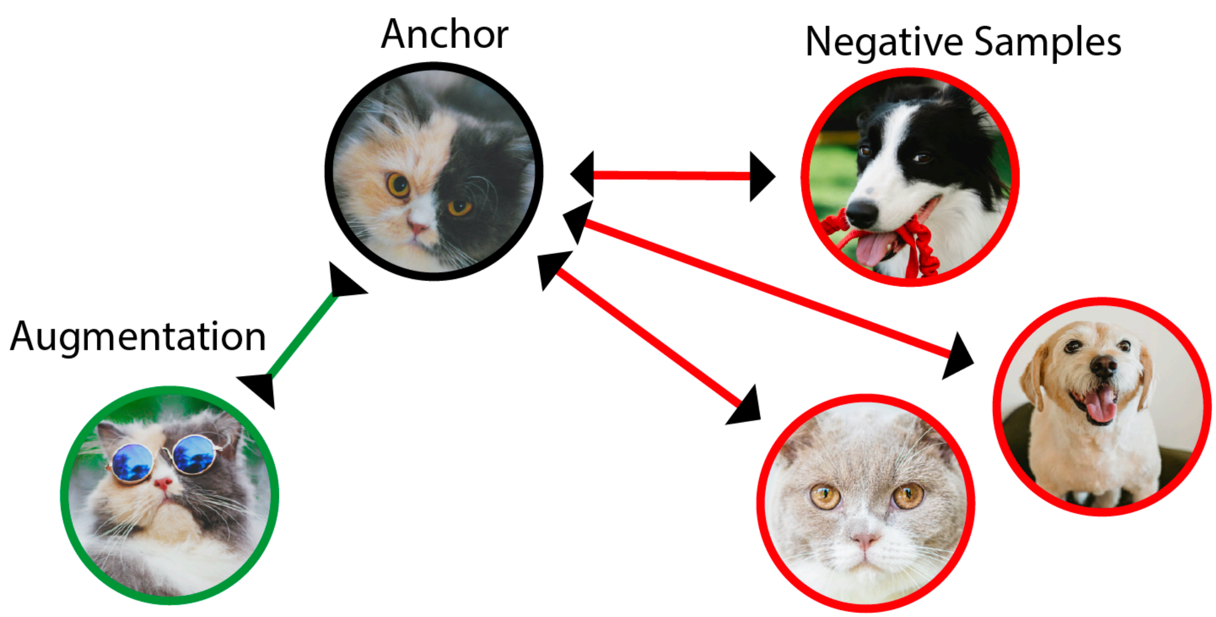
- Can we use AI to learn generic representations of jets?



Once we have a representation, how can we apply it as a pre-trained "backbone" model to solve downstream tasks?

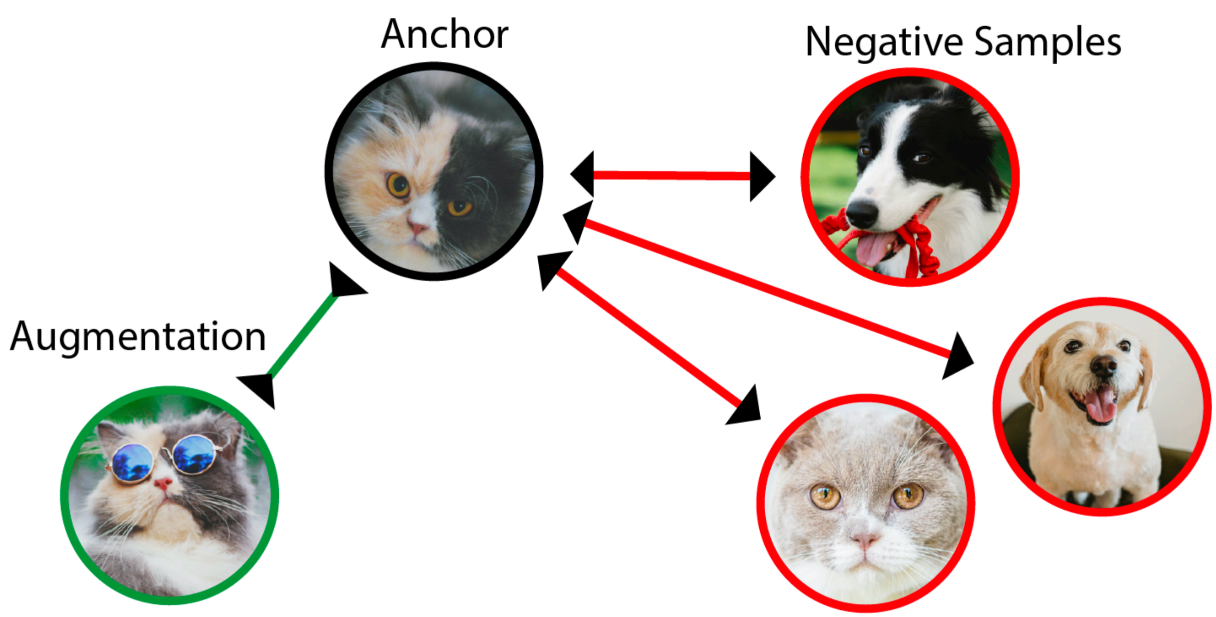
Self-supervised Learning (SSL)

- Learn using only relations between the **same objects from different perspectives**



Self-supervised Learning (SSL)

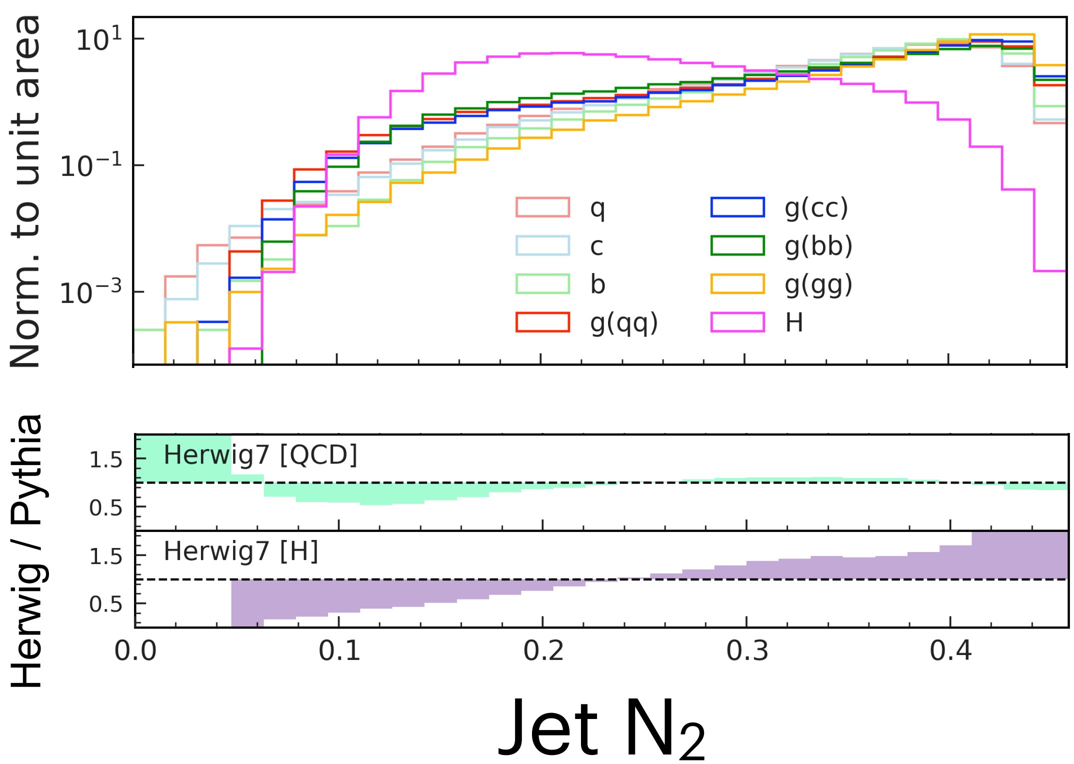
- Learn using only relations between the **same objects from different perspectives**



- *Classes = **Higgs and QCD***
- *Augmentations = **reshowering parton** with different simulators (pythia tunes + herwig)*

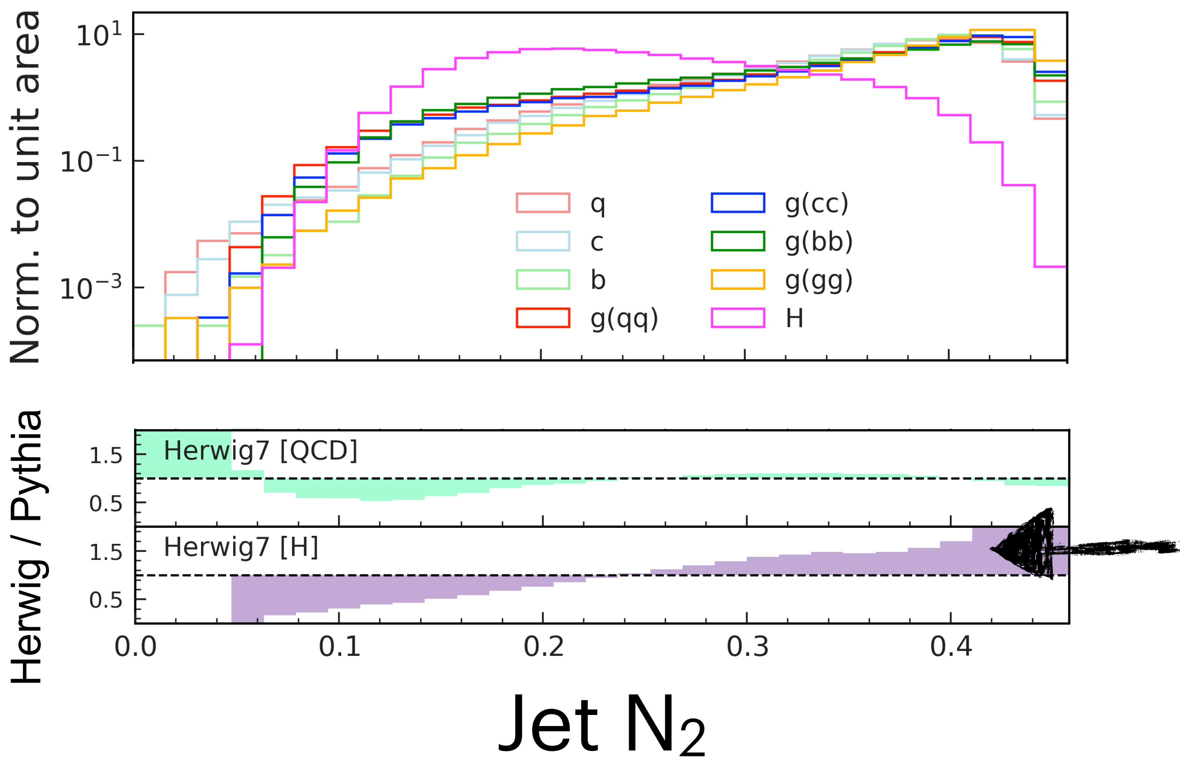
Where to start?

- We begin by investigating Higgs and QCD jets with **various parton shower models**
  - Parton shower assumptions can be leading uncertainties in physics analyses (e.g. 2208.02751)



Where to start?

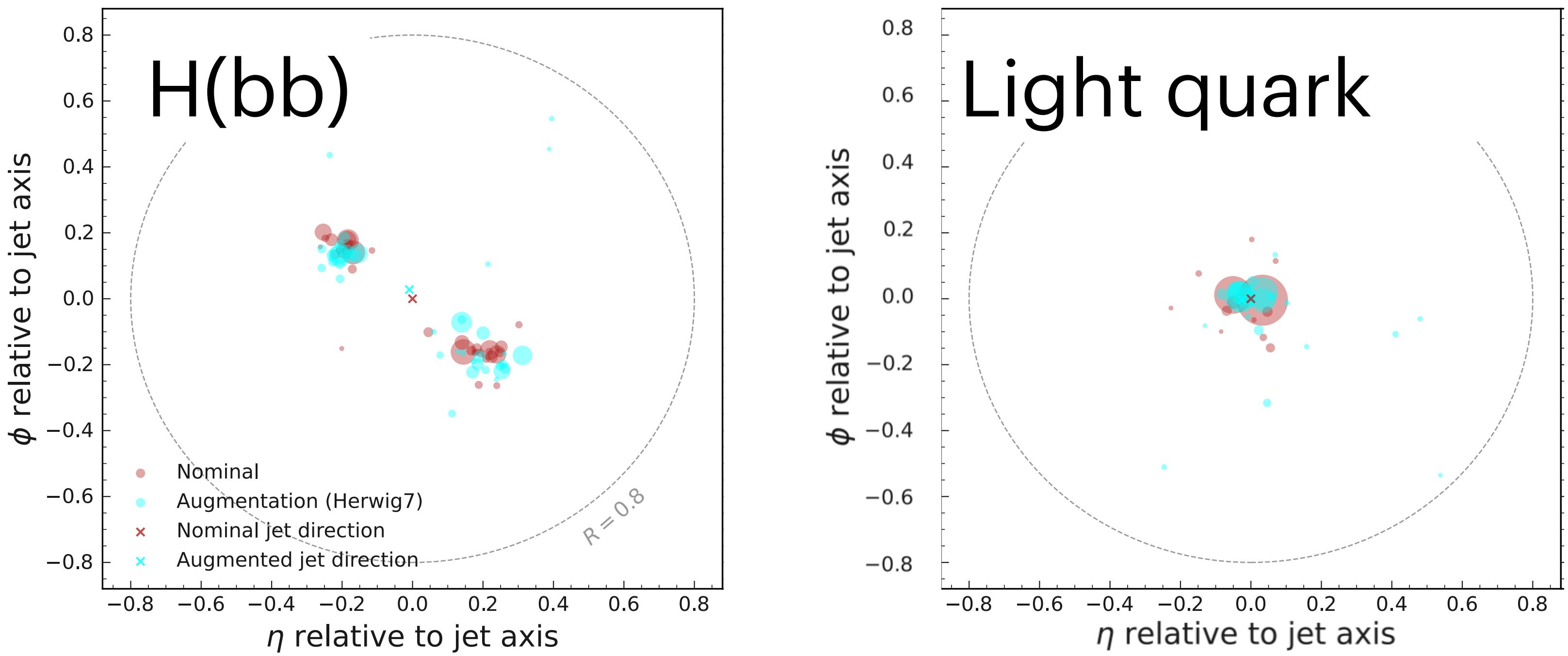
- We begin by investigating Higgs and QCD jets with **various parton shower models**
  - Parton shower assumptions can be leading uncertainties in physics analyses (e.g. 2208.02751)



Large herwig/pythia ratio
=
Large modelling differences

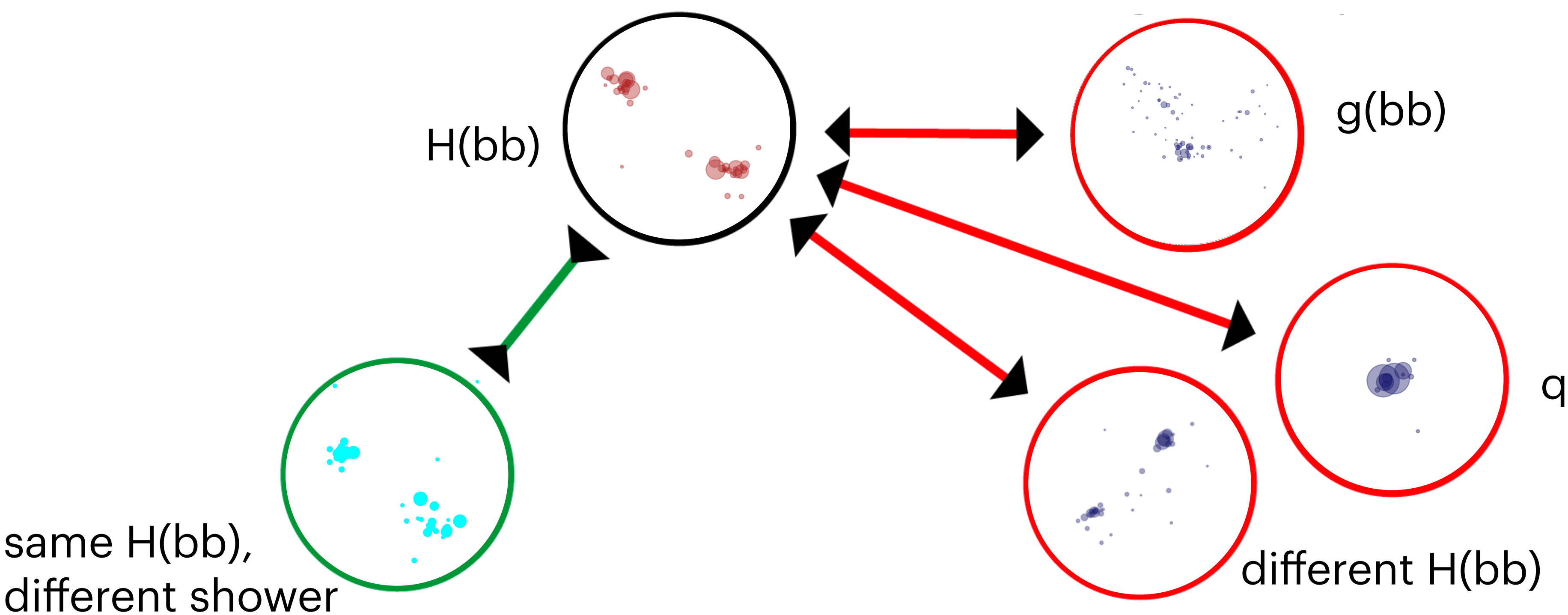
The Augmentations

Parton showered with Pythia8
Same parton showered with Herwig



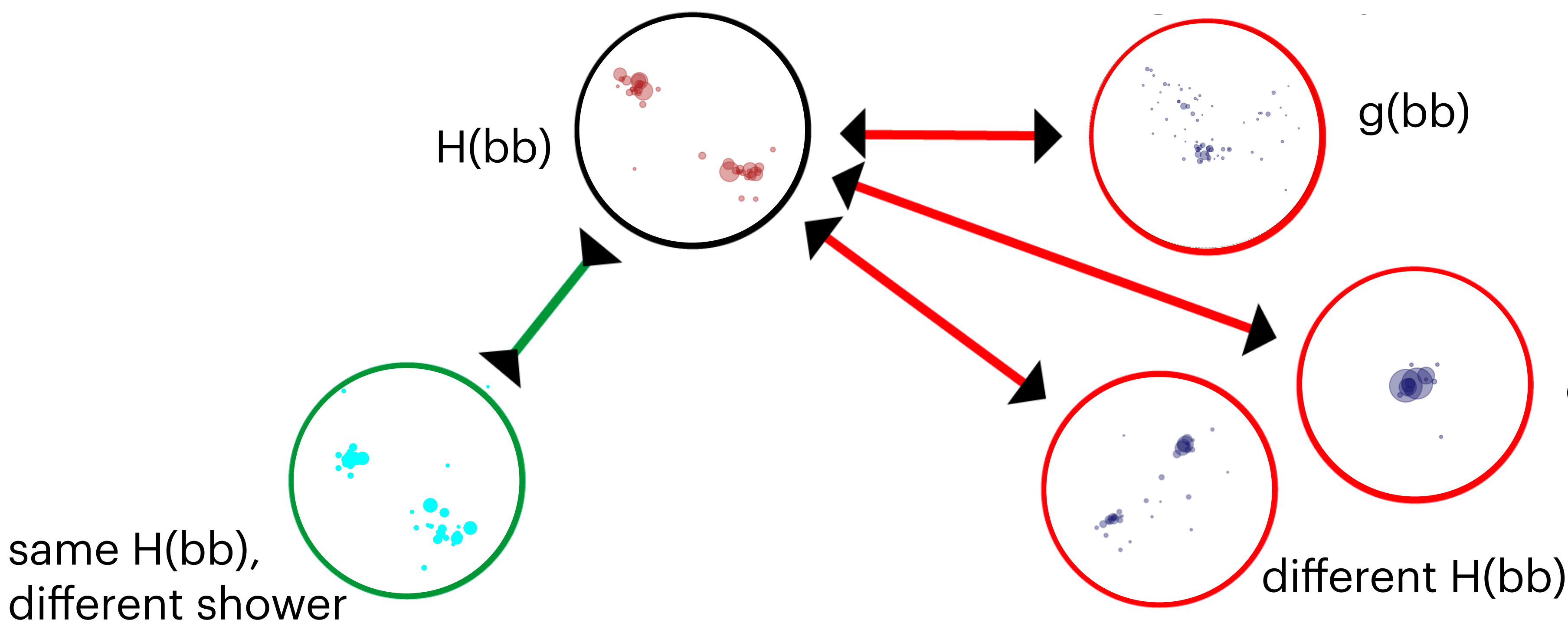
The Augmentations

Parton showered with Pythia8
Same parton showered with Herwig



The Augmentations

Parton showered with Pythia8
Same parton showered with Herwig

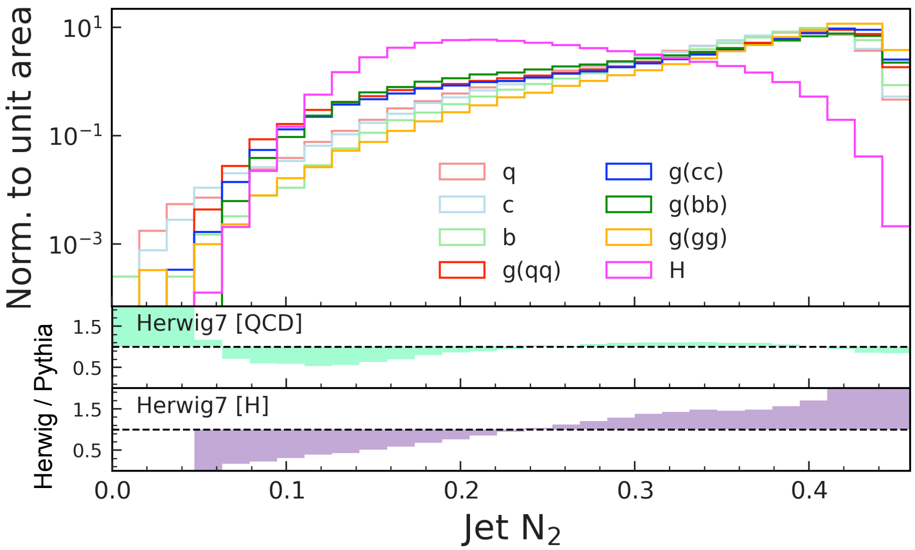


We use a self-supervised graph neural network to pull together the nominal and augmented (re-simulated) jet pair

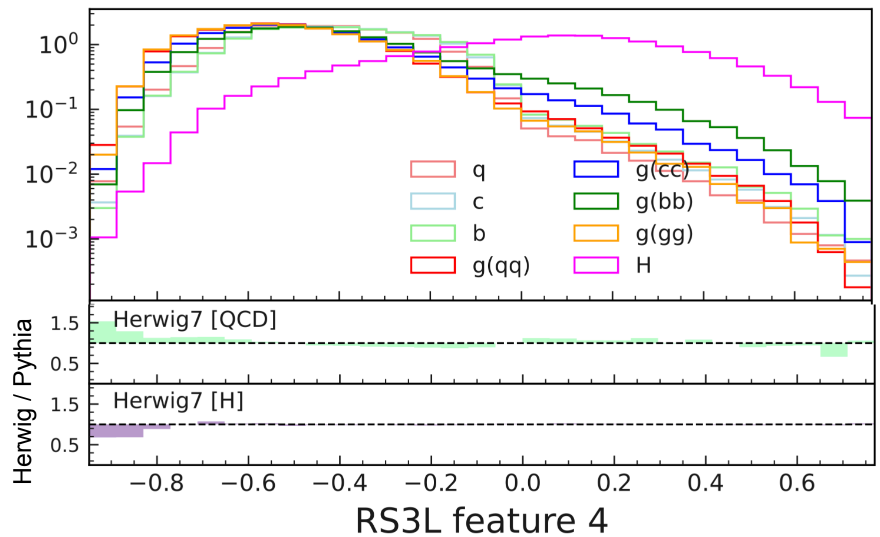
→ ***Re-simulation-based self-supervised learning (RS3L)***

The RS3L Space

Jet N_2

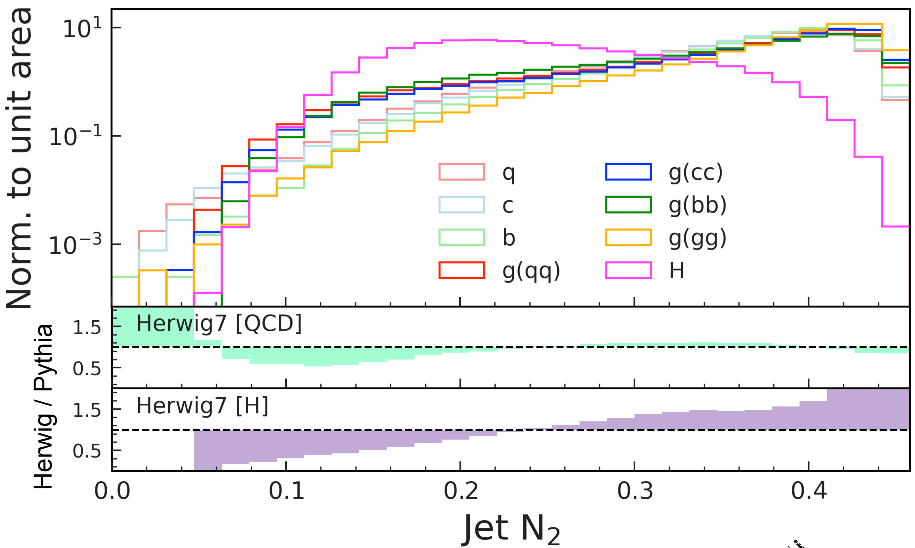


One RS3L feature

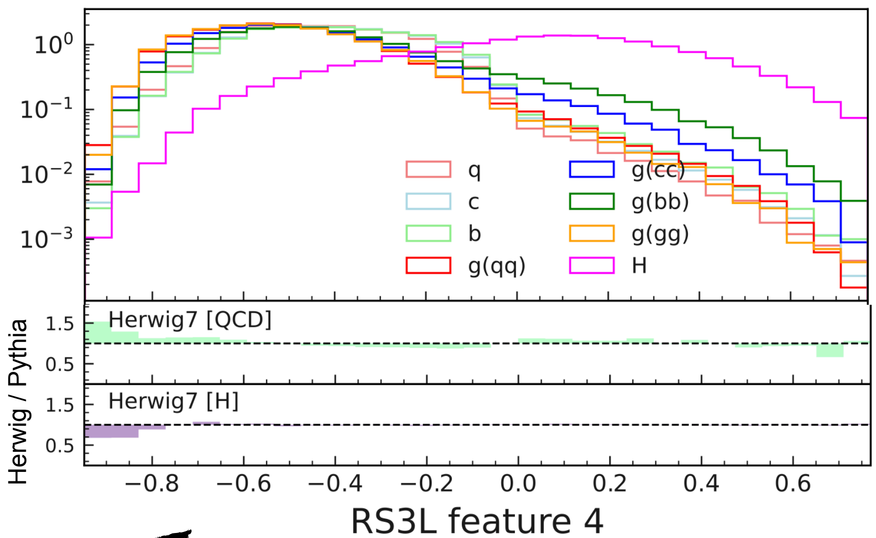


The RS3L Space

Jet N_2



One RS3L feature

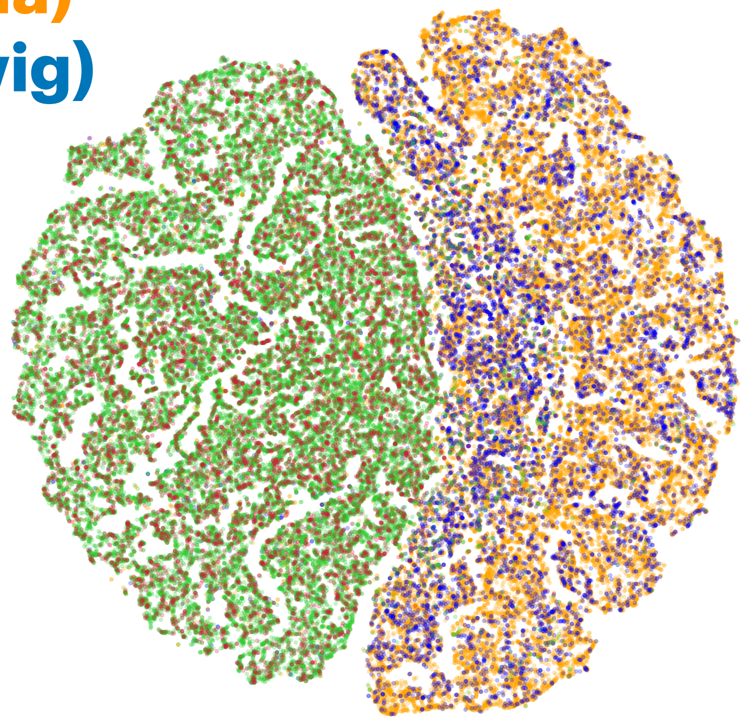


Pythia/Herwig ratio is reduced → our space learns ***features of H and QCD with less reliance on simulator***

The RS3L Space

tSNE of 8 contrastive features

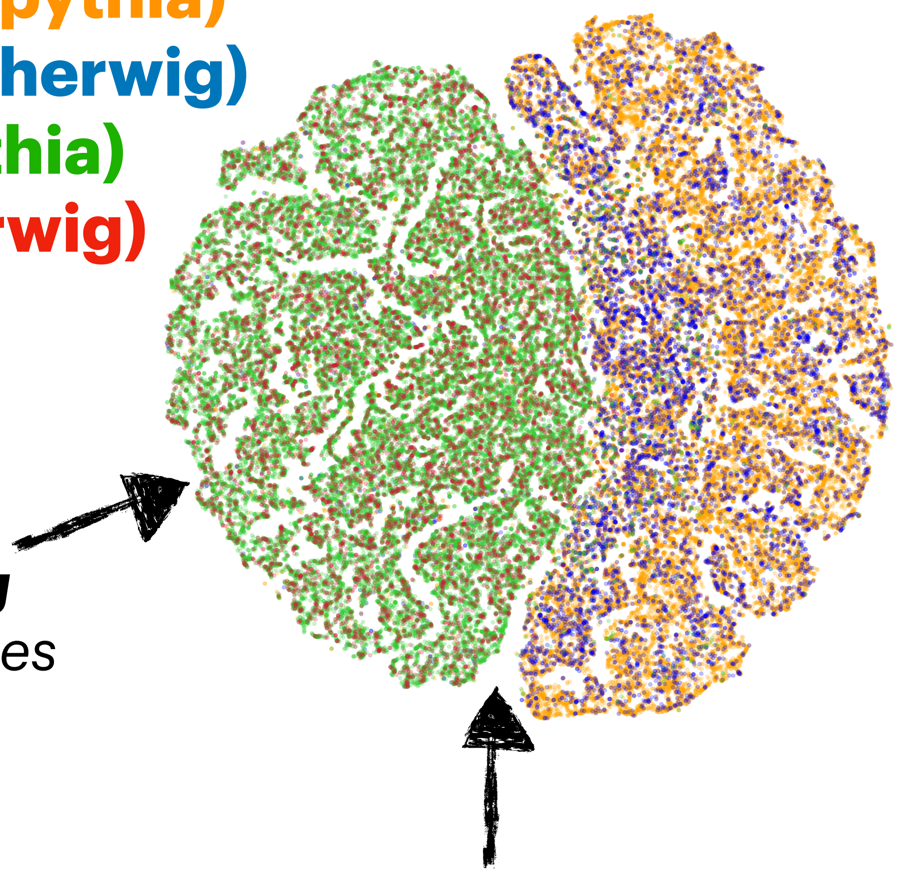
QCD (pythia)
QCD (herwig)
H (pythia)
H (herwig)



The RS3L Space

tSNE of 8 contrastive features

QCD (pythia)
QCD (herwig)
H (pythia)
H (herwig)

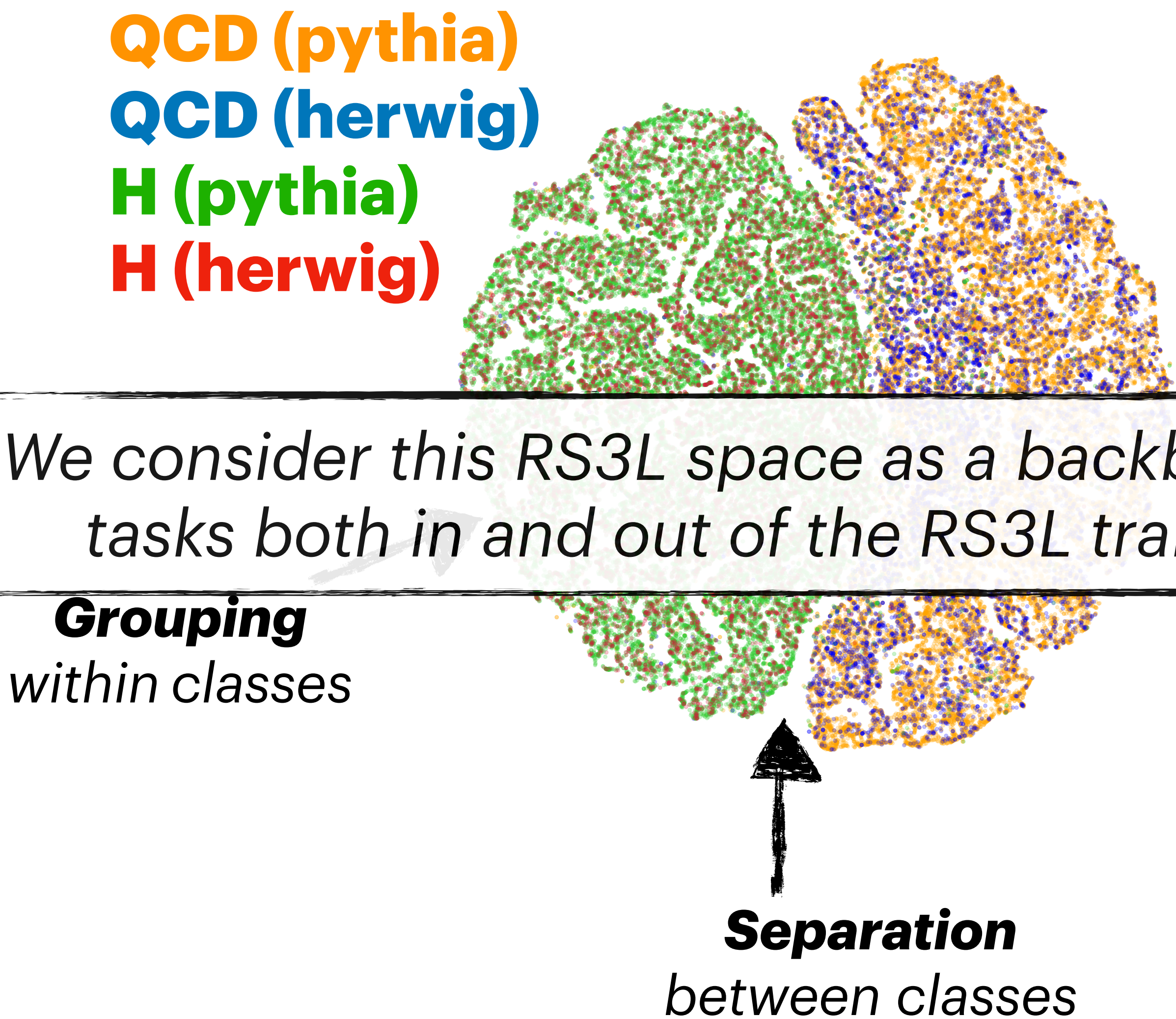


Grouping
within classes

Separation
between classes

The RS3L Space

tSNE of 8 contrastive features



We consider this RS3L space as a backbone in learning tasks both in and out of the RS3L training domain

#1: In-domain classification

QCD background rejection rates at various Higgs tagging efficiencies.

Higgs efficiency	0.3	0.5	0.7
RS3L fine-tuned (3M)	1340	379	135
Fully-supervised (8M)	1295	384	131

#1: In-domain classification

QCD background rejection rates at various Higgs tagging efficiencies.

Higgs efficiency	0.3	0.5	0.7
RS3L fine-tuned (3M)	1340	379	135
Fully-supervised (8M)	1295	384	131

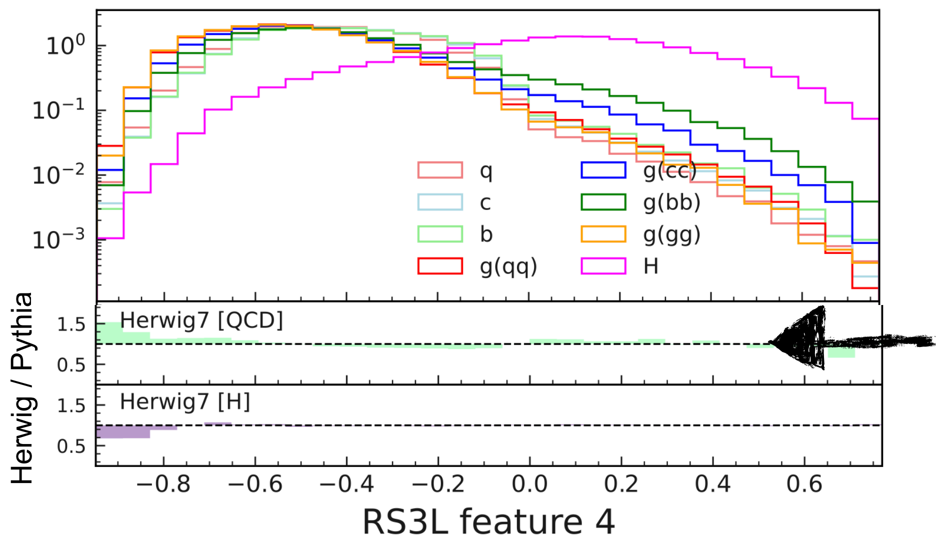
RS3L gives similar performance on in-domain classification

Robustness of RS3L space

- What effect does changing simulators (pythia→Herwig) have on the output of the final classifier?
  - We use the Wasserstein distance to quantify the **difference between tagger evaluated on nominal and augmented jets**

Robustness of RS3L space

- What effect does changing simulators (pythia→Herwig) have on the output of the final classifier?
  - We use the Wasserstein distance to quantify the **difference between tagger evaluated on nominal and augmented jets**



Compute Wasserstein distance between Herwig and Pythia

Robustness of RS3L space

Wasserstein distance between taggers distributions evaluated on pythia and herwig Higgs jets.

Training setup	Herwig
RS3L fine-tuned (3M)	$7.8\ \times 10^{-3}$
Fully-supervised (8M)	$9.4\ \times 10^{-3}$

Robustness of RS3L space

Wasserstein distance between taggers distributions evaluated on pythia and herwig Higgs jets.

Training setup	Herwig
RS3L fine-tuned (3M)	$7.8\ \times 10^{-3}$
Fully-supervised (8M)	$9.4\ \times 10^{-3}$

Fine-tuning RS3L reduces distance

→ ***SSL can provide more robust observables***

#2: Out-of-domain classification

- Does the RS3L space (trained on Higgs and QCD) contain useful information about **other processes**
  - out-of-domain learning: apply RS3L base to **W jets**

#2: Out-of-domain classification

- Does the RS3L space (trained on Higgs and QCD) contain useful information about **other processes**
  - out-of-domain learning: apply RS3L base to **W jets**

W efficiency	0.3	0.5	0.7
RS3L fine-tuned (3M)	1893	505	147
Fully-supervised (3M)	1781	457	134

#2: Out-of-domain classification

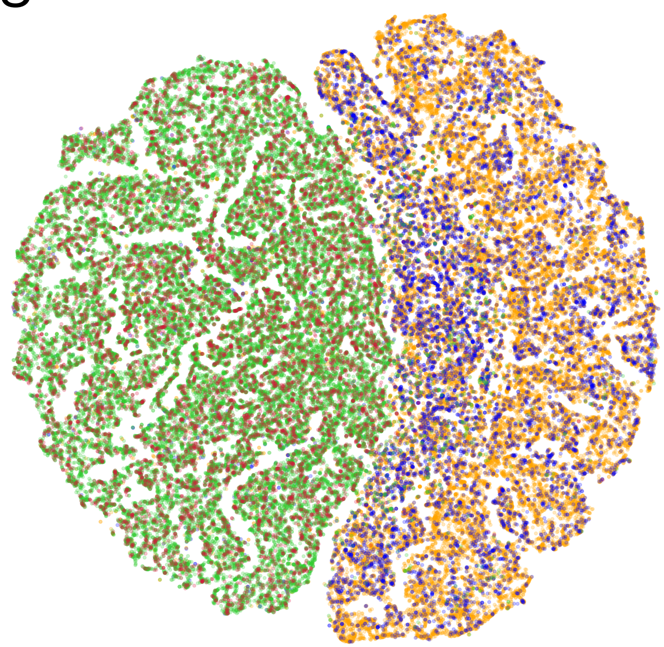
- Does the RS3L space (trained on Higgs and QCD) contain useful information about **other processes**
  - out-of-domain learning: apply RS3L base to **W jets**

W efficiency	0.3	0.5	0.7
RS3L fine-tuned (3M)	1893	505	147
Fully-supervised (3M)	1781	457	134

Starting from RS3L base → ***improvement in out-of-domain*** *classification*

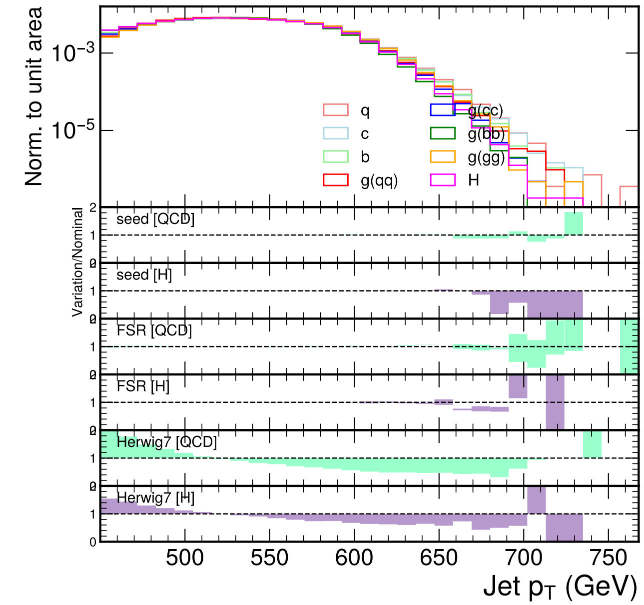
Conclusion

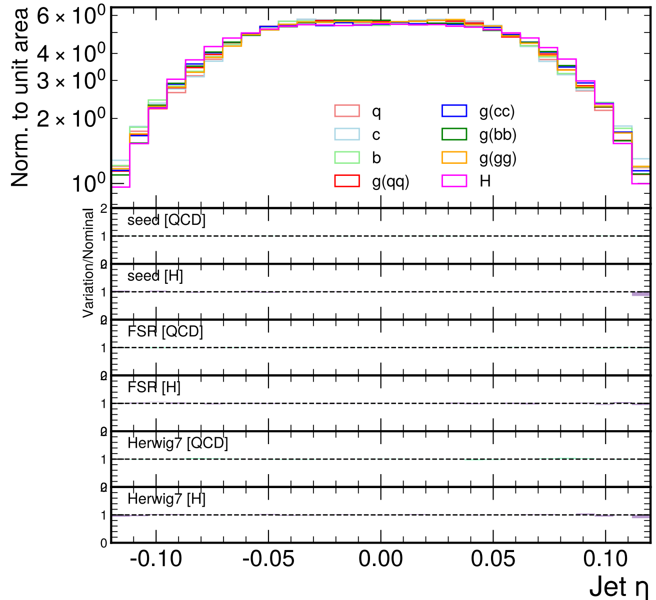
- Self-supervision helps us build a representation without labels
  - Could be a path toward a **foundation model** for HEP data
- RS3L has significant potential in downstream applications including:
  - Acting as a foundation for **discrimination tasks** (e.g. QCD/H)
  - Helping to **mitigate uncertainties** (e.g. parton showering)
  - Translating learning to out-of-domain classification tasks

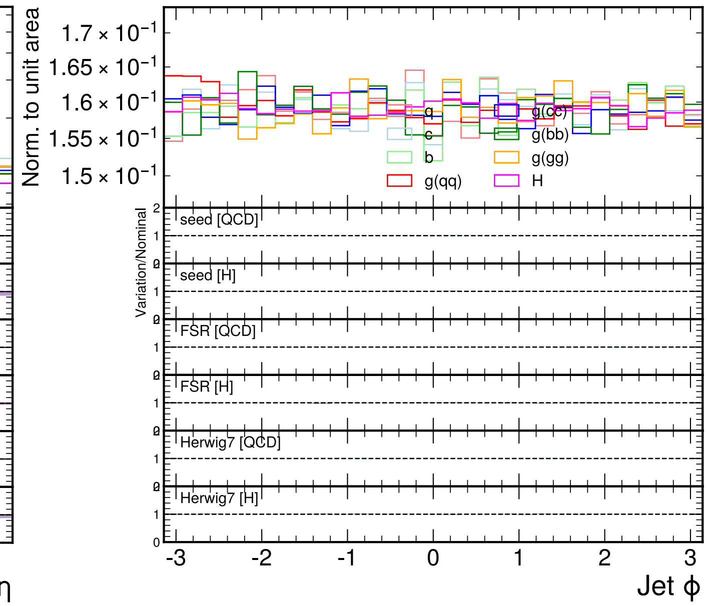


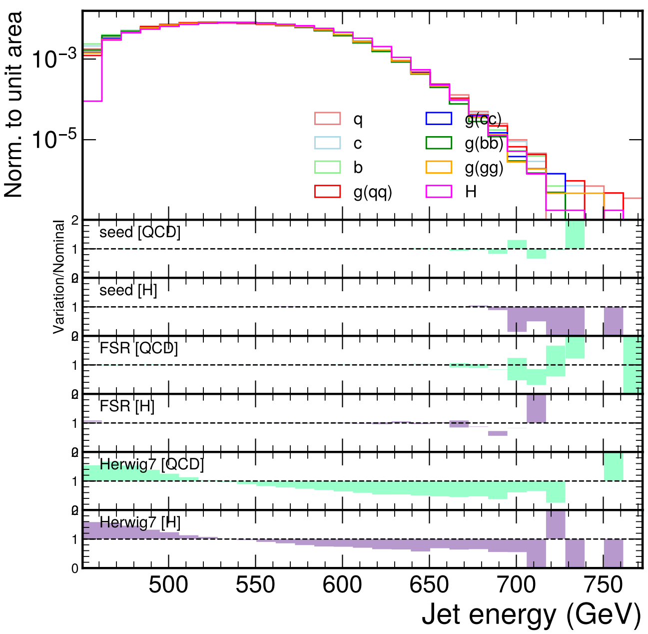
Backup

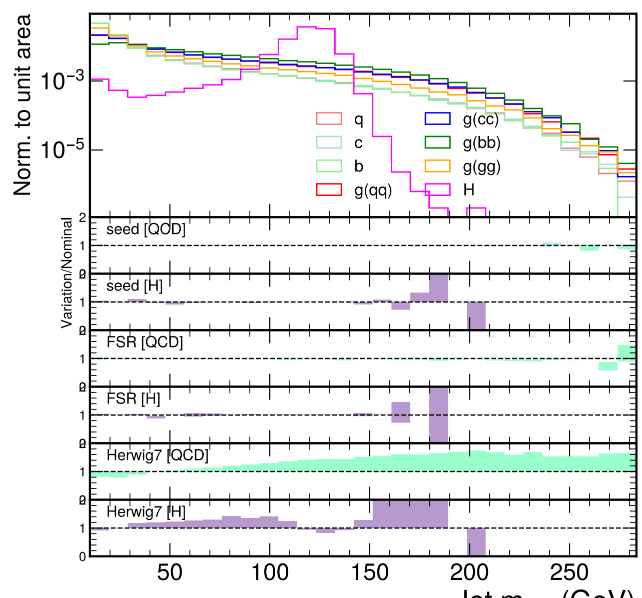
Jet kinematics

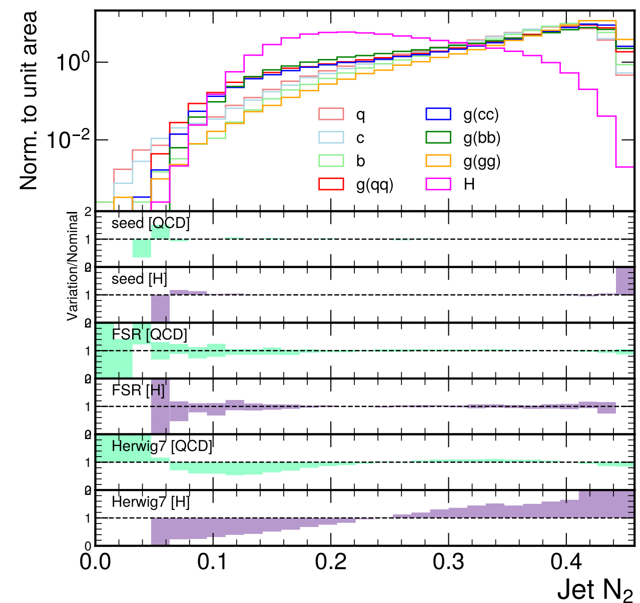








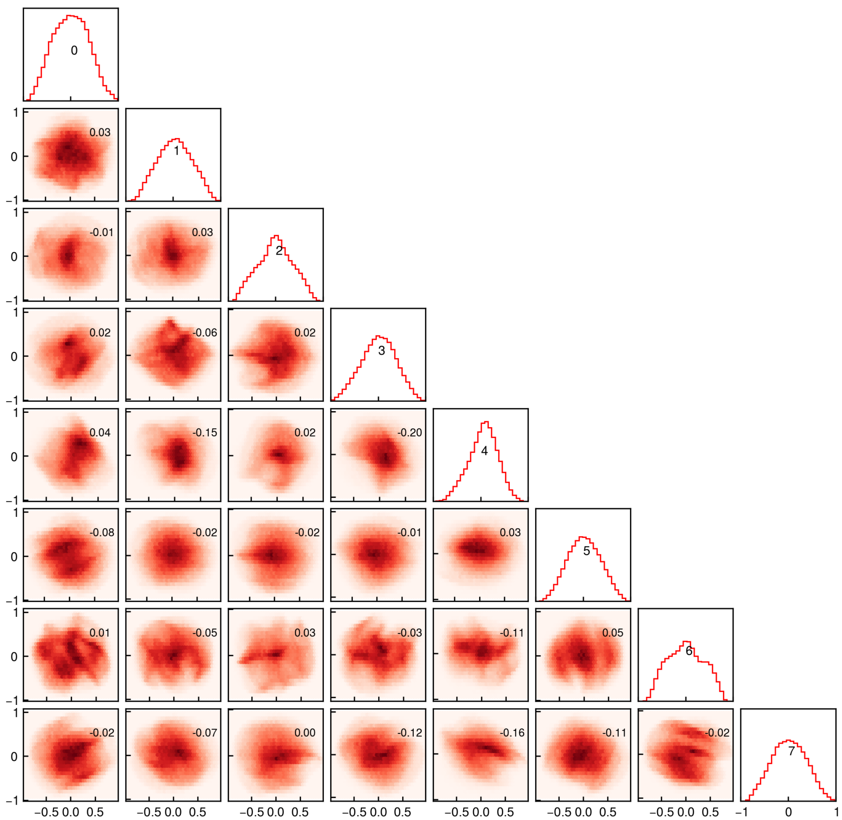




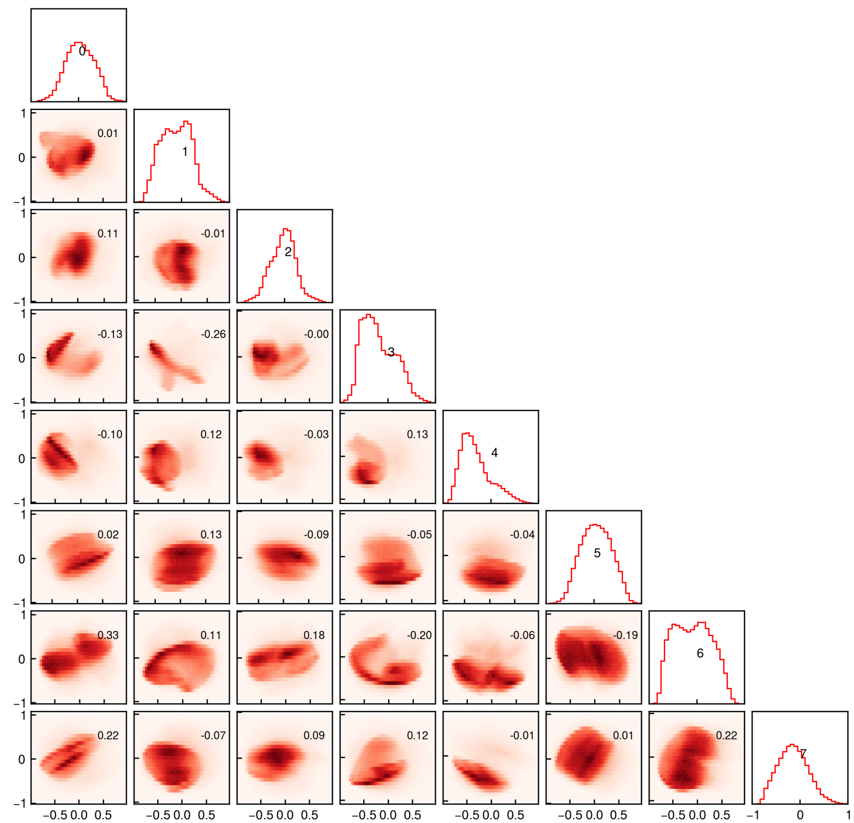
Corner Plots of contrastive features

- Higgs jets occupy a relatively larger volume of the space and tend to be distributed more uniformly

Higgs



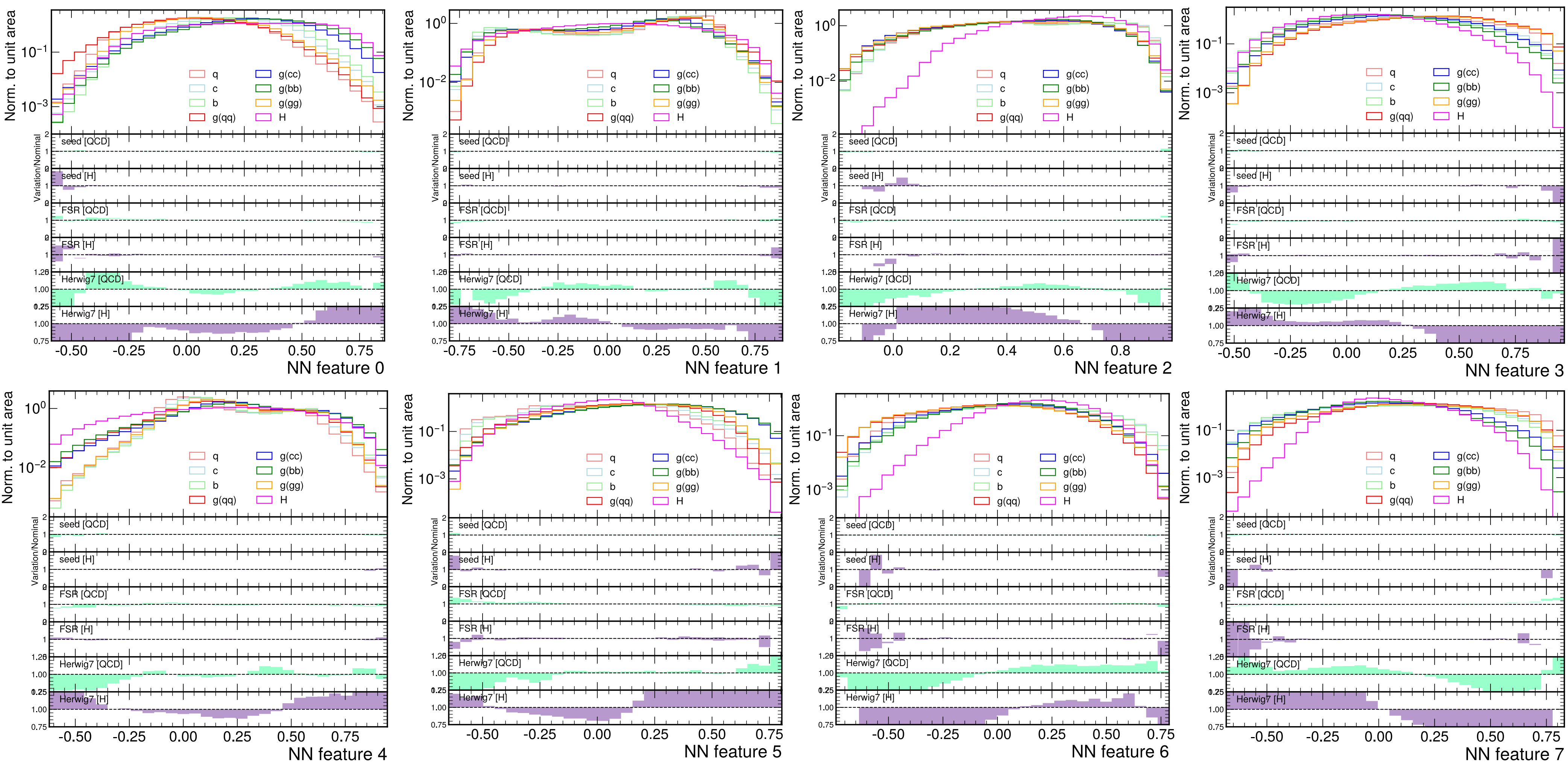
QCD



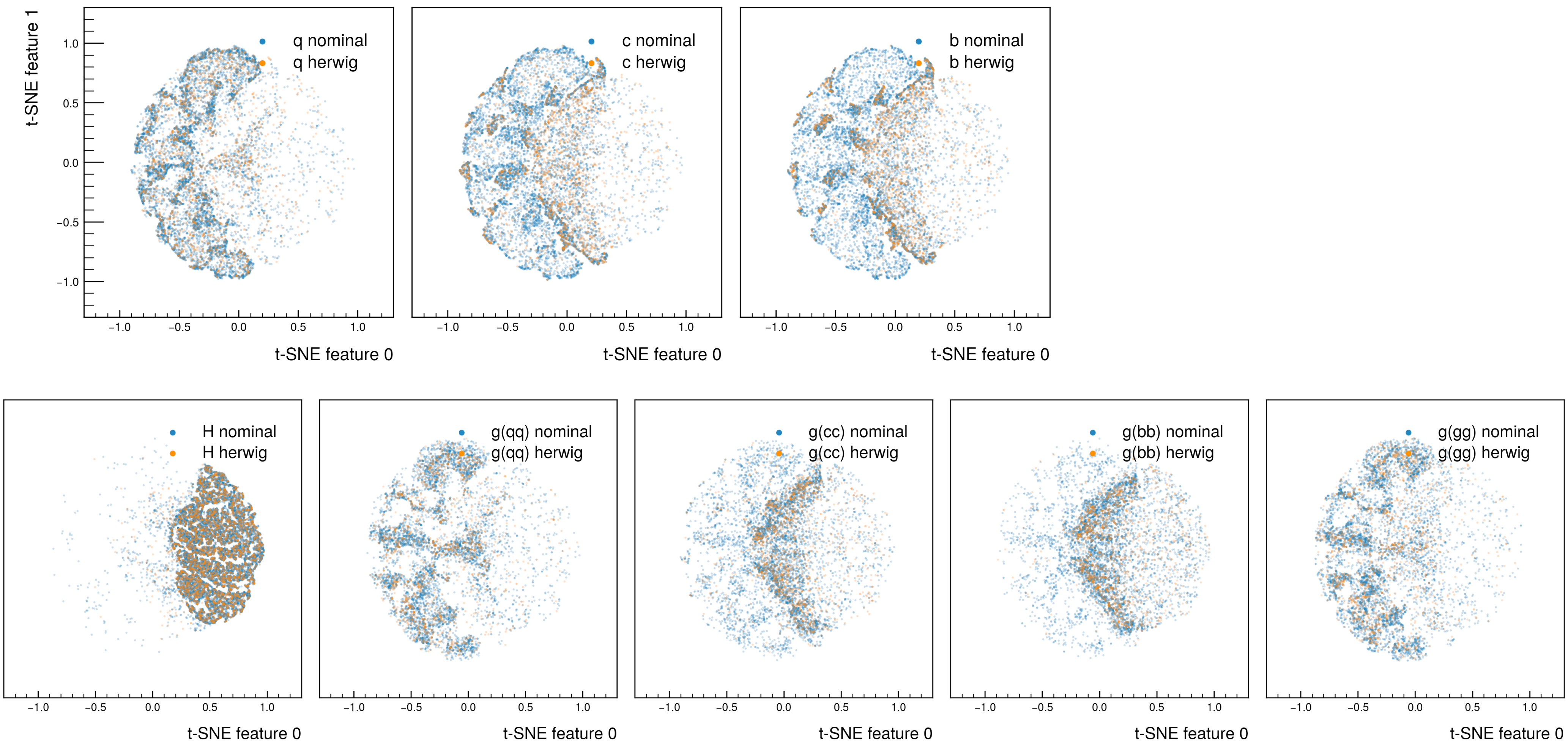
learned SSL features are largely independen
other

QCD-only training inferred on Higgs

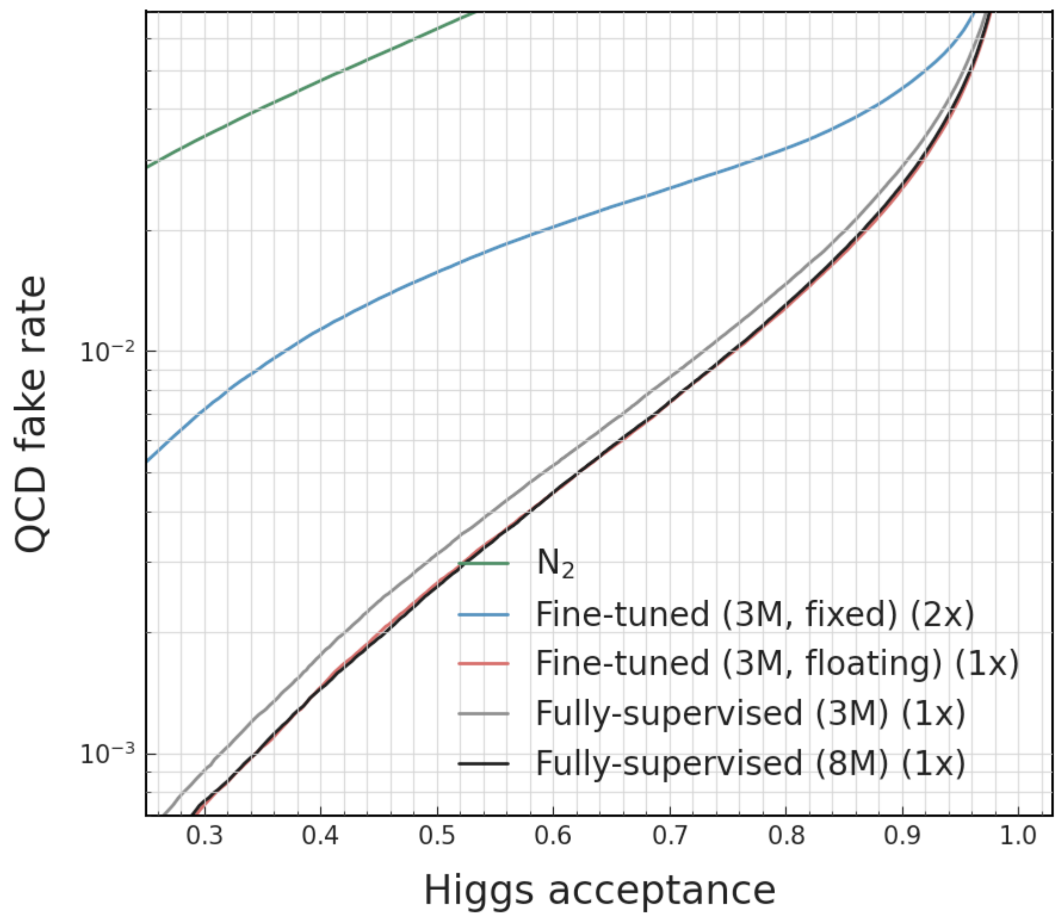
- We learn a reasonable Higgs representation despite training only on QCD



Full tSNE

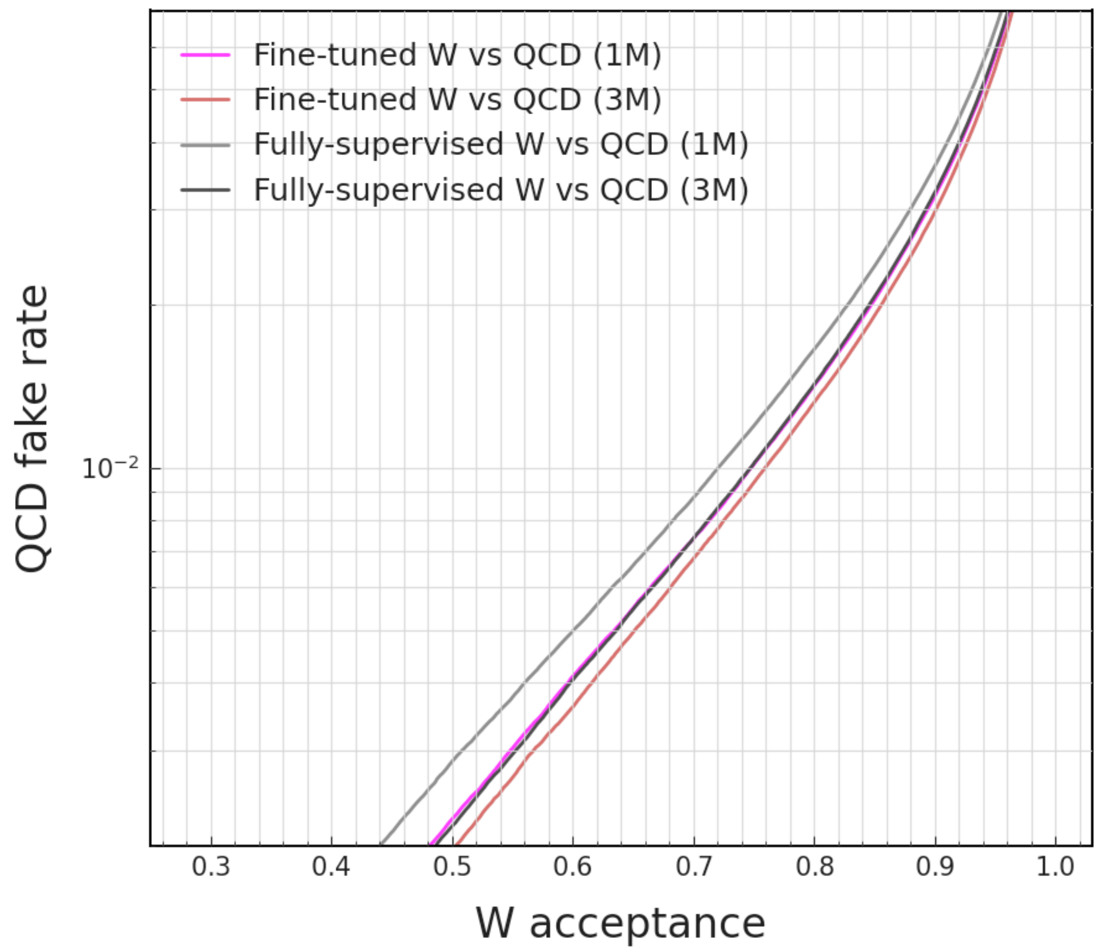

q nominal
q herwig
c nominal
c herwig
b nominal
b herwig
H nominal
H herwig
g(qq) nominal
g(qq) herwig
g(cc) nominal
g(cc) herwig
g(bb) nominal
g(bb) herwig
g(gg) nominal
g(gg) herwig
t-SNE feature 1
t-SNE feature 0

H vs QCD ROC



QCD fake rate
10^{-2}
10^{-3}
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
Higgs acceptance
N_2
Fine-tuned (3M, fixed) (2x)
Fine-tuned (3M, floating) (1x)
Fully-supervised (3M) (1x)
Fully-supervised (8M) (1x)

W vs QCD ROC



Fine-tuned W vs QCD (1M)
Fine-tuned W vs QCD (3M)
Fully-supervised W vs QCD (1M)
Fully-supervised W vs QCD (3M)
QCD fake rate
10^{-2}
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1.0
W acceptance