Self-Supervised Learning Strategies for Jet Physics

Garrett Merz

(with: Patrick Rieck, Kyle Cranmer, Etienne Dreyer, Eilam Gross, Nilotpal Kakati, Dmitrii Kobylanskii, Nathalie Soybelman)

Self-Supervised Learning Uses Unlabelled Data

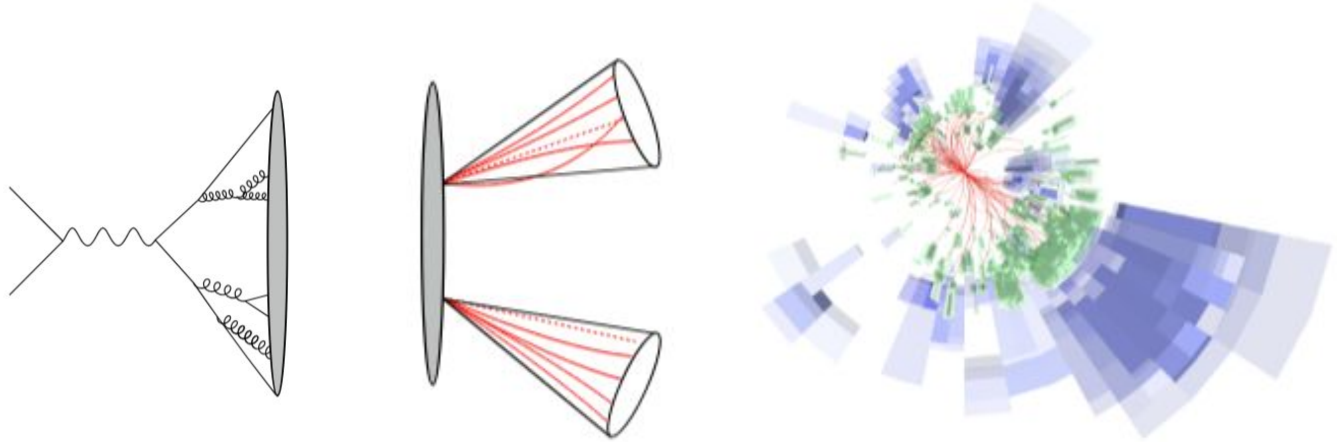
Core concept: when task-specific labelled data is expensive, but unlabelled data is plentiful, use self-supervised pretraining

Common type of SSL objective is contrastive learning: same or different?

Augment unlabelled data to create “positive pairs” and “negative pairs”

Augmentation through Re-Simulation

MC simulation chain is a factorizable Markov process!



$$p(x_{\text{Hard Scatter}}, x_{\text{Parton}}, x_{\text{Hadron}}, x_{\text{Detector}}) = p(x_{\text{Hard Scatter}})\; p(x_{\text{Parton}}|x_{\text{Hard Scatter}})\; p(x_{\text{Hadron}}|x_{\text{Parton}})\; p(x_{\text{Detector}}|x_{\text{Hadron}})$$

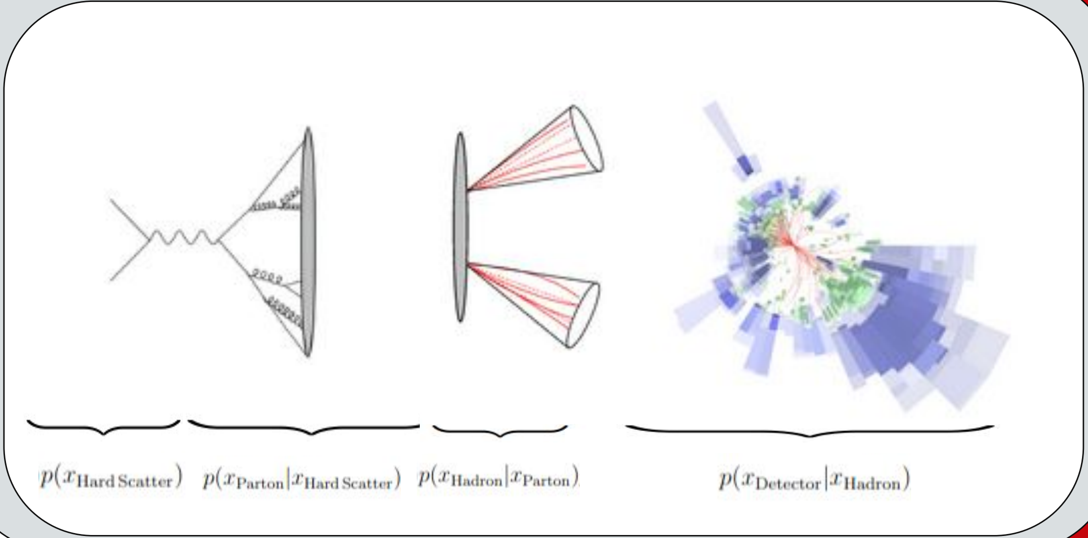
Augmentation through Re-Simulation

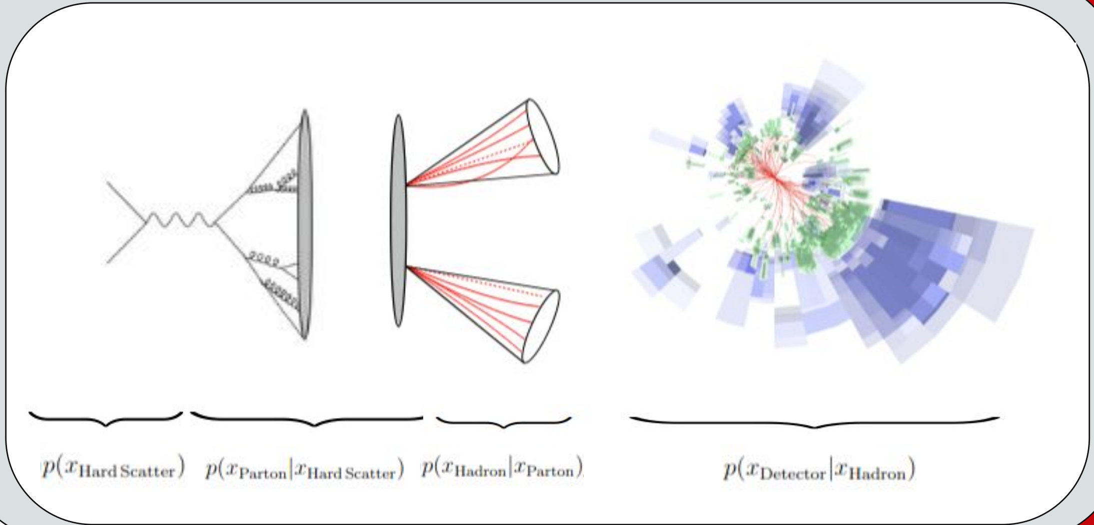
Generate augmentations by freezing generator up to a certain step and re-running.

We can produce events with same hard-scatter BUT:

- different hadronization (same parton shower)
- same hadronization & shower, but different detector response
- etc.

Introduced in (Harris et al, arXiv:2403.07066)

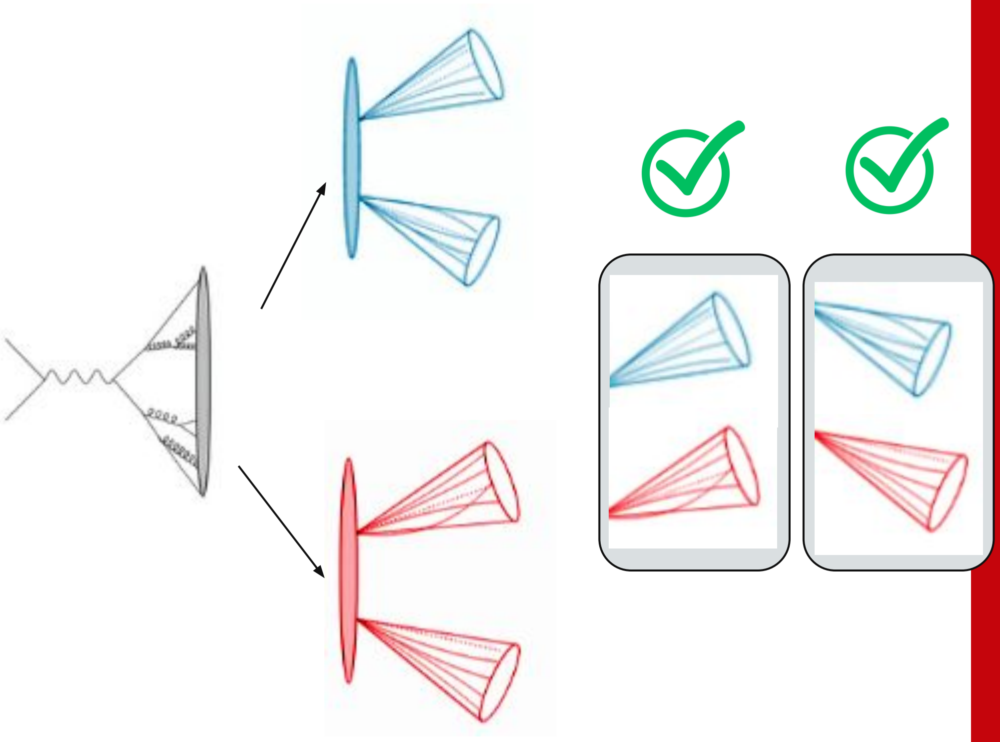




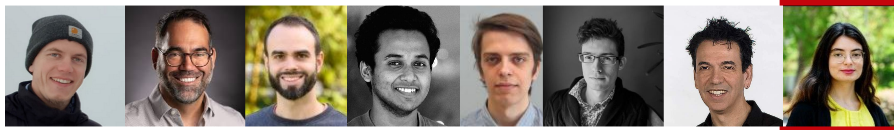
Augmentation through Re-Simulation

Generate 1M dijet events:

- Hard-scatter with MadGraph:
  - $|\eta| < 3.0$, all final-state pTs > 100 GeV
- Parton shower & hadronization with Pythia:
  - keep taus with 1 or 3 charged hadrons, require that resimulation doesn't change tau decay channel
- Form R=0.4 anti-kT jets:
  - associate tracks that are within R < 0.4 of jet axis
  - Require jet pT > 25 GeV (detector level), pT > 50 GeV (hadron level)
- Simulate detector response with COCOA (COnfigurable CalOrimeter for AI) differentiable calorimeter package



arXiv: 2503.11632

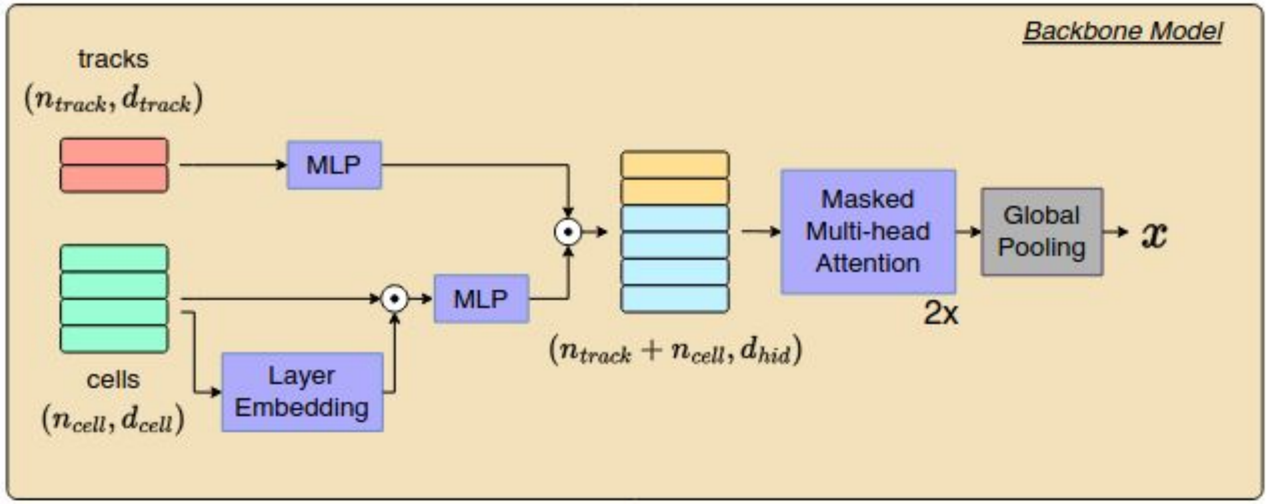


Contrastive Pre-Training

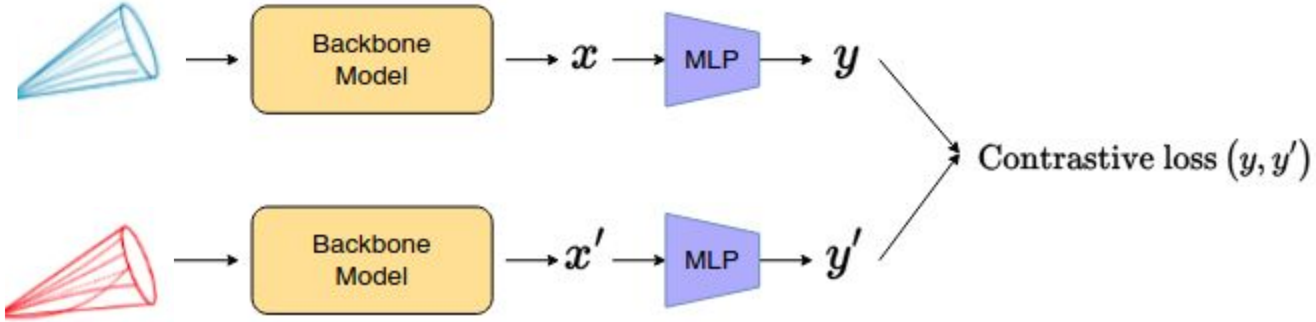
- Split dataset: 350k pairs for pretraining, 200k for downstream, 50k validation, 50k test
- Use Transformer backbone model comparable to that used in ATLAS
- Within a batch, construct <u>positive pairs</u> that originate from the same hard-scatter and <u>negative pairs</u> that do not
- Use SimCLR contrastive objective

$$\mathcal{L} = \frac{1}{2N}\sum_{k=1}^{N}\left[\ell_{(2k-1,2k)} + \ell_{(2k,2k-1)}\right],$$

$$\ell_{i,j} = -\log\frac{\exp(\mathrm{sim}(y_i, y_j)/\tau)}{\sum_{k=1}^{2N}\mathbb{1}_{[k\neq i]}\exp(\mathrm{sim}(y_i, y_k)/\tau)}$$

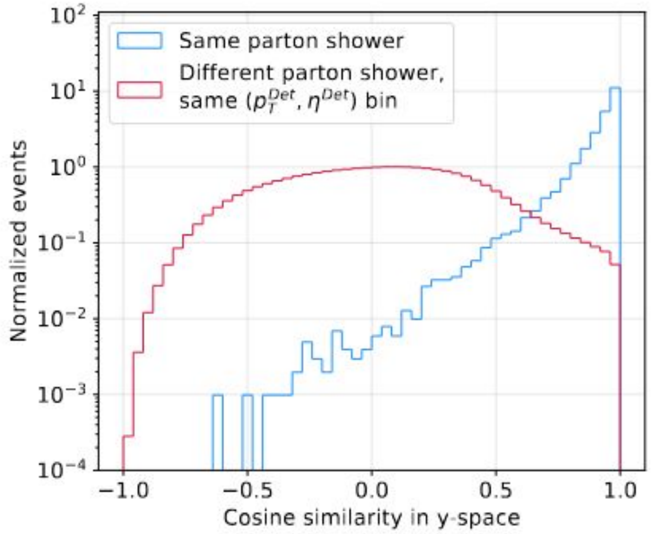


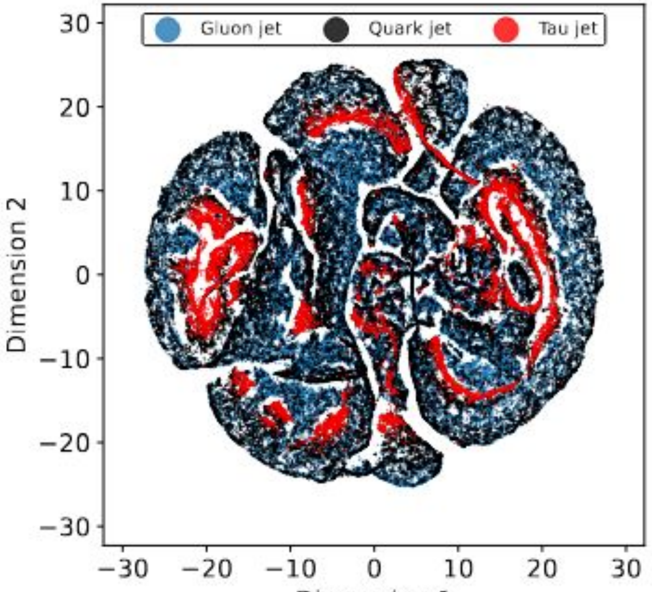
(a)



Evaluating Learned Representations

- Cosine similarity of representations for positive pairs is high and negative pairs is low
- q/g/tau jets form clusters in embedding space (per t-SNE embedding projection)
- kNN clustering (k=10) on pretrained reps correctly IDs 71% of gluon jets and 97% of taus

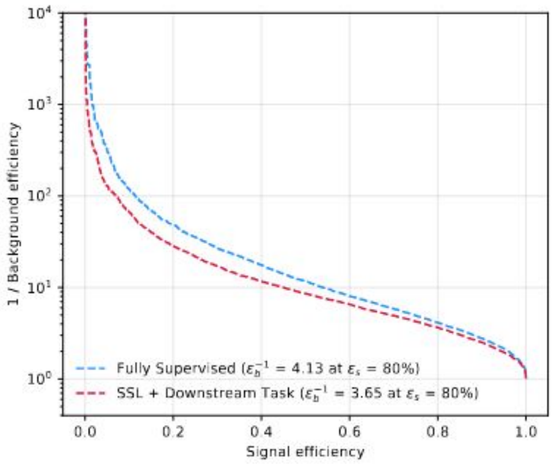


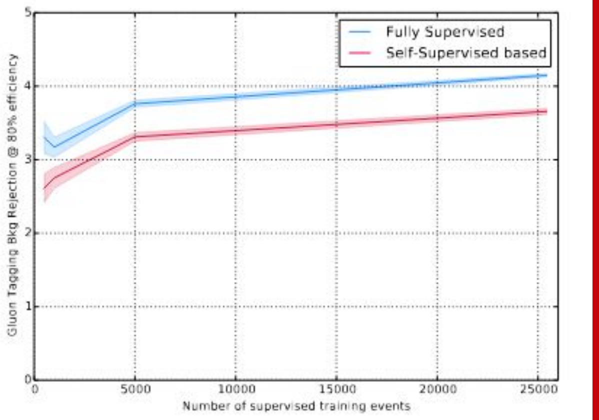


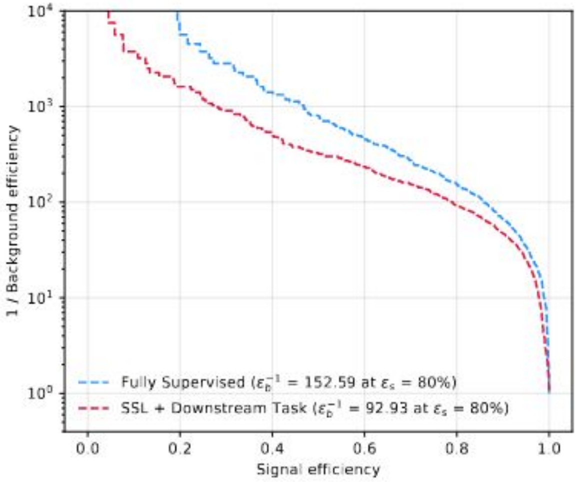
Downstream Tasks

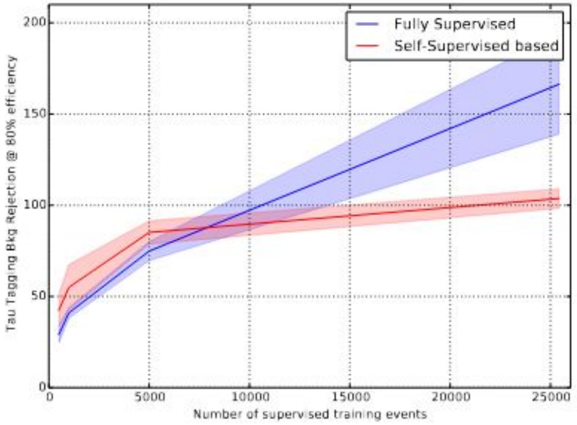
- Compare:
  - SSL-pretrained model w. LoRA fine-tuning
  - fully-supervised (no SSL)
- On:
  - q/g jet classification task
  - tau tagging task

- Fully-supervised is superior, but:
  - Comparable for q/g tagging
  - SSL outperforms fully-supervised when labelled data is scarce







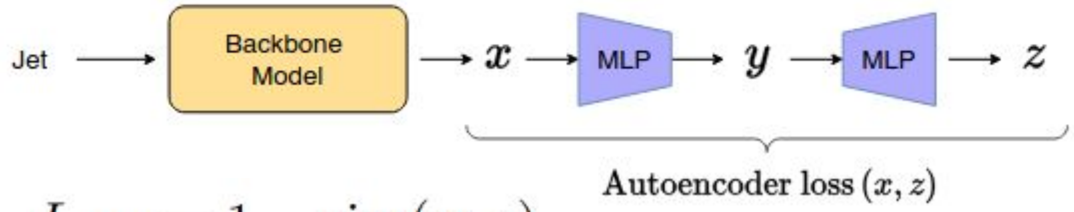


(d)

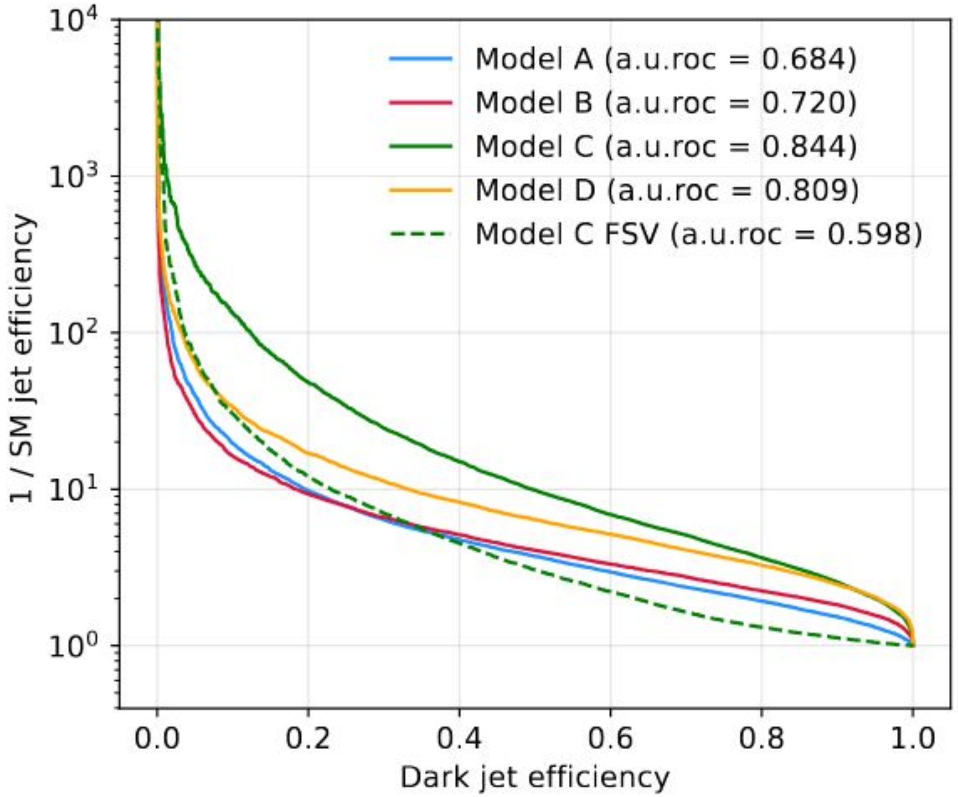
LoRA: arxiv.org/abs/2106.09685

Anomaly Detection (Dark Jets)

- Generate four "dark jet" datasets (Models A-D)
  - Resample to match SM jet pT,η distribution; 20k jets each
- Train on SM jets ONLY; BSM jets marked by poor reconstruction
  - Anomaly score: $\log(L_{reco})$
- Model: SSL pretrained backbone + autoencoder
- Compare w. fully-supervised training
  - (Model C FSV, dashed line)



$$L_{\text{reco}} = 1 - \text{sim}(x, z)$$



ATLAS resonant dark jets: arxiv.org/abs/2311.03944

Model	n_f	Λ_d (GeV)	$\tilde{m}_{q'}$ (GeV)	m_{π_d} (GeV)	m_{ρ_d} (GeV)	π_d decay mode
A	2	15	20	10	50	$\pi_d \to c\bar{c}$
B	6	2	2	2	4.67	$\pi_d \to s\bar{s}$
C	2	15	20	10	50	$\pi_d \to \gamma'\gamma'$ with $m_{\gamma'} = 4.0$ GeV
D	6	2	2	2	4.67	$\pi_d \to \gamma'\gamma'$ with $m_{\gamma'} = 0.7$ GeV

Conclusion

- Extension of ReSim approach shows potential of SSL for developing robust representations for jet physics
  - Allows us to learn jet substructure in unsupervised way

- Many other potential applications of ReSim-based approaches!
  - Same event, different pileup profile
  - Same event, different detector
  - etc.

- ATLAS/CMS generate LOTS of data- clever pretraining strategies may enable us to extract as much info as possible

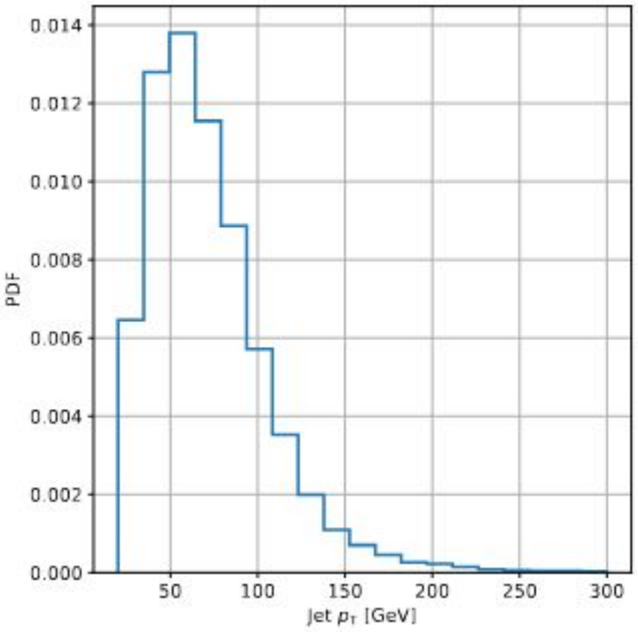
WISCONSIN
UNIVERSITY OF WISCONSIN–MADISON

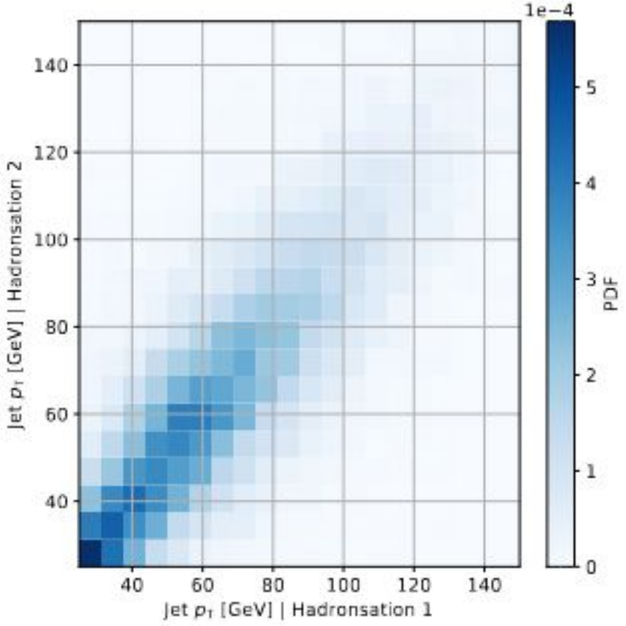
Hyperparameters

Hyperparameter	
Batch size	30
Optimizer	Adam [21]
Learning rate schedule	Single Cycle [22]
Maximum Learning rate	0.001
Number of epochs	150
Number of trainable parameters	
Cell layer Embedding	24
Cell MLP	17,536
Track MLP	17,280
Transformer	199,424
Output Network	8,776
Downstream MLP	1,281

Table 1: Hyperparameters used in the SSL pre-training and downstream supervised learning.

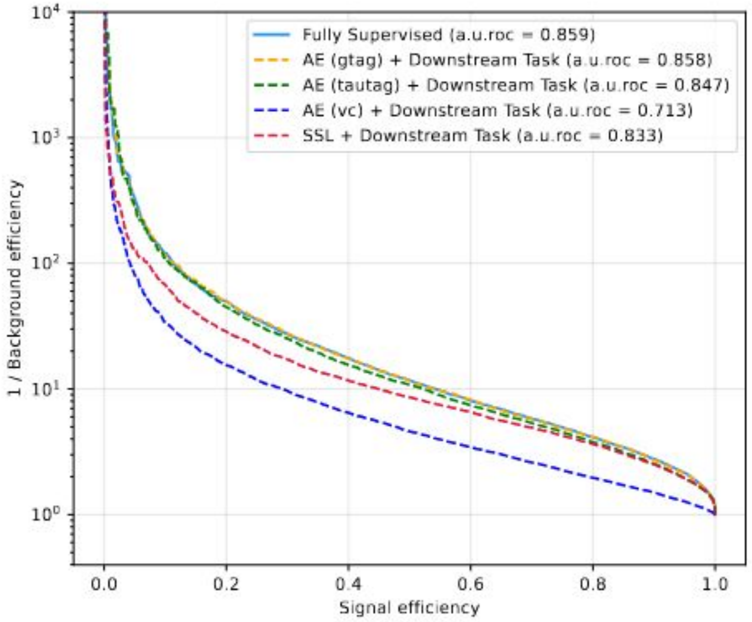
Jet Distributions



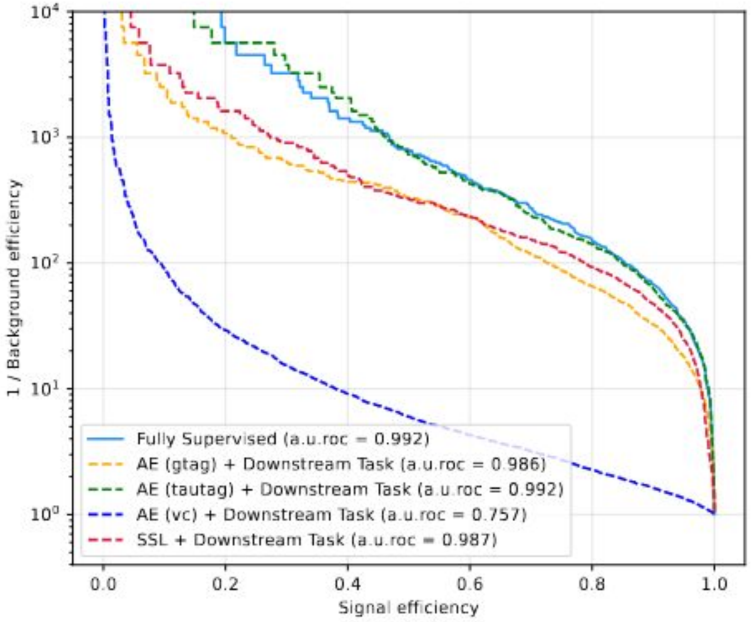


Comparison w. Autoencoder

- Encode each jet individually, rather than as pairs



(a) gluon tagging



(b) tau tagging