GENOVA

BOOST 2024

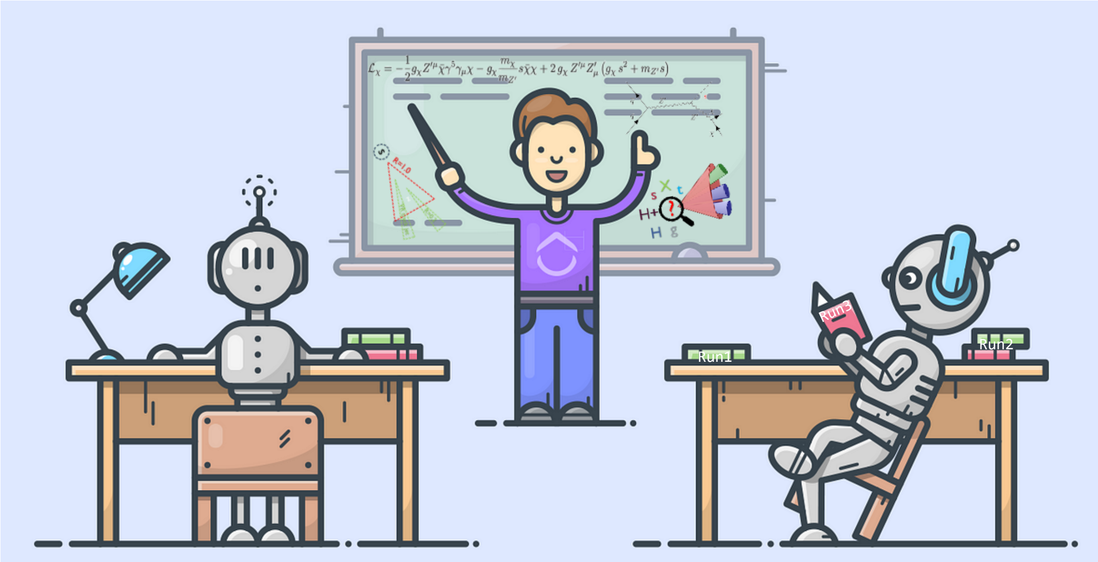
16th International Workshop on Boosted Objects Phenomenology

Learning powerful jet representations via self-supervision

Qibin LIU, Shudong Wang, Congqiao Li, Huilin Qu

Introduction

- Significant advances in jet tagging with wide application of ML
- Supervised learning model: strong performance while limited by labelled dataset
- We propose a new method to learn jet representations through self-supervision
- Applications to jet tagging and anomaly detection
- Outlook of future development



Plot modified from @Srinivas Rao

ML-based Jet tagging: the supervised way

- Exploit the information to assign correct jet label (Hbb/Hcc/tbqq/…)
- Focus on boosted jet reconstructed with PFlow algo

  Input: large-R jet composed of particles

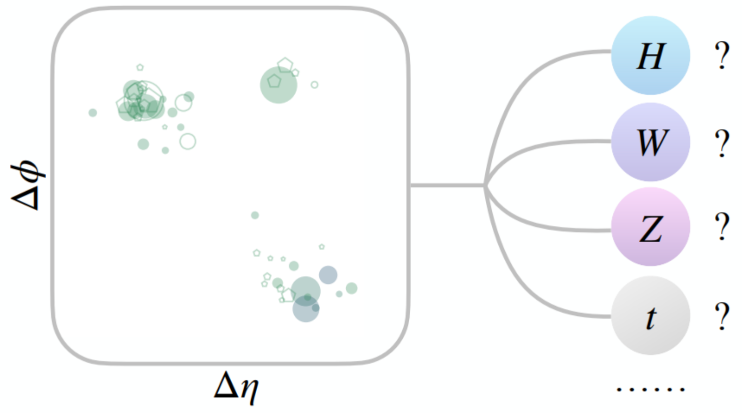
- Amazing development over years:

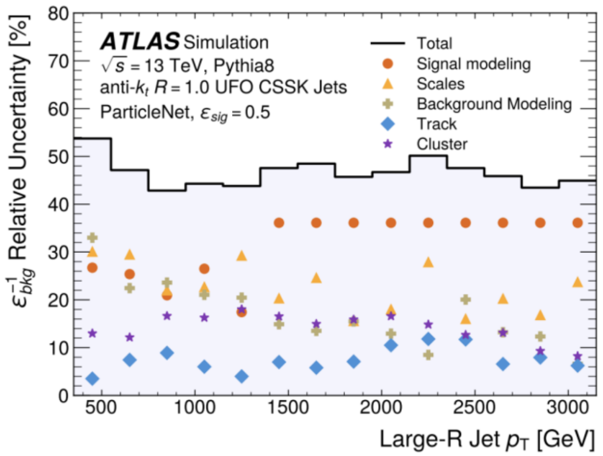
  ParticleNet, Particle Transformer, LundNet, PELICAN, OmniLearn, Sophon and many more!

- Common feature: trained from the **labelled dataset**

  Physics modelling, data-MC difference and statistics

Recall nice talks these days!





Plot taken from 2202.03772, CERN-EP-2024-159

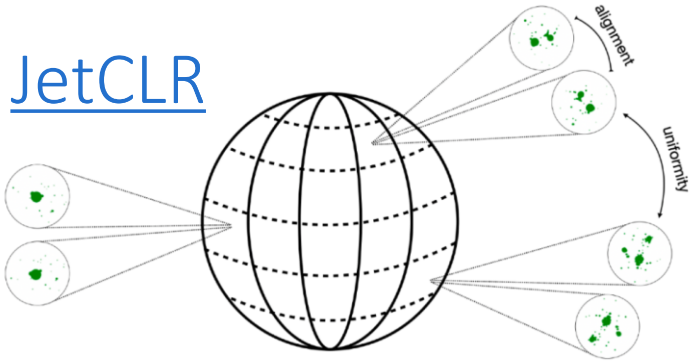
Can we learn from data?

The self-supervised learning (SSL)

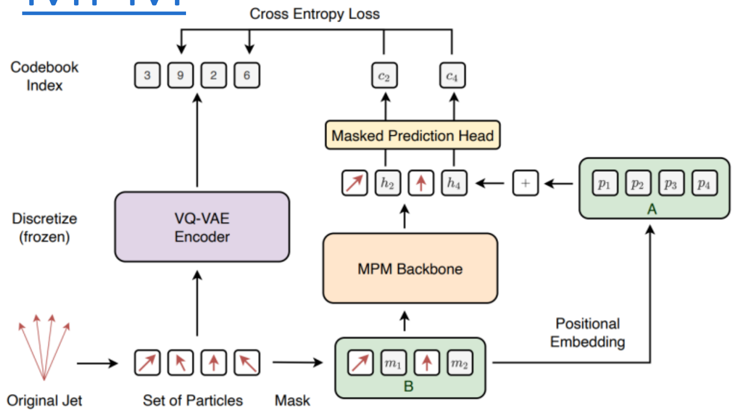
Self-supervised learning

- Physics knowledge embedded in jet even w/o label
  Color connection, hadronization, detector effect, …
- Self-supervised way to learn from unlabeled jet
  SimCLR, JetCLR (AD), AnomalyCLR, DarkCLR, RS3L, …
  Masked Particle Modelling
  OmniJet-α
- Jet representation shared between various applications
  Jet reconstruction, tagging, generation, anomaly detection, …
  → Bridge to the foundation jet model

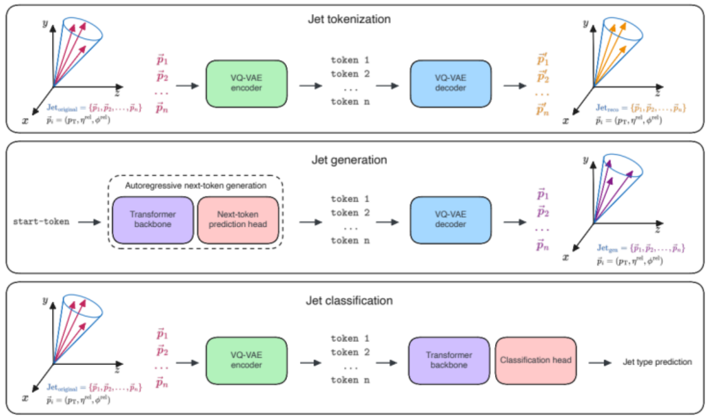
JetCLR



MPM



OmniJet-α

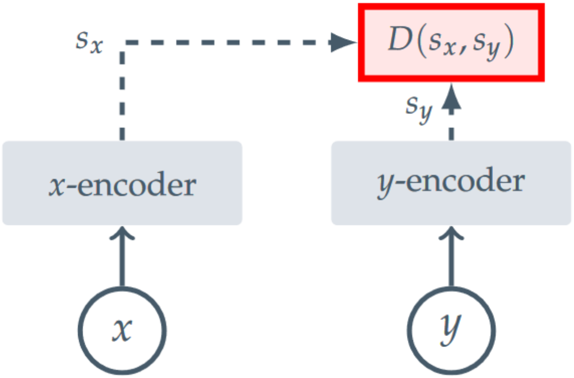


Designs of self-supervised learning

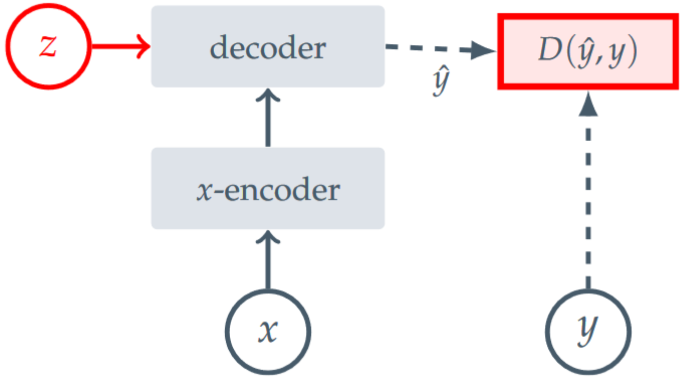
- a) Contrastive: min- or maximize the distance between representation of jet pair
- b) Generative: generate partial or the full jet
- c) Predictive: complete the jet representation

  Easy to train: no need to build pair or generate in physics space

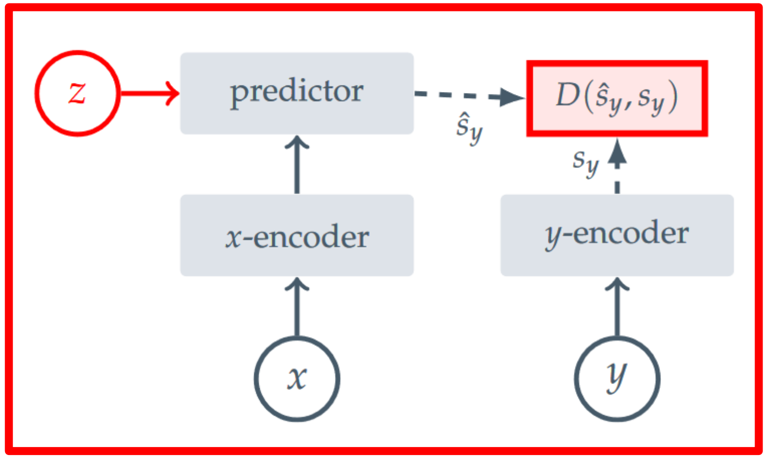
  Flexible to extend: handle any kind of jet input (more than kinematics)



SimCLR, JetCLR (AD), AnomalyCLR, DarkCLR, RS3L, …



Masked Particle Modelling
OmniJet-α



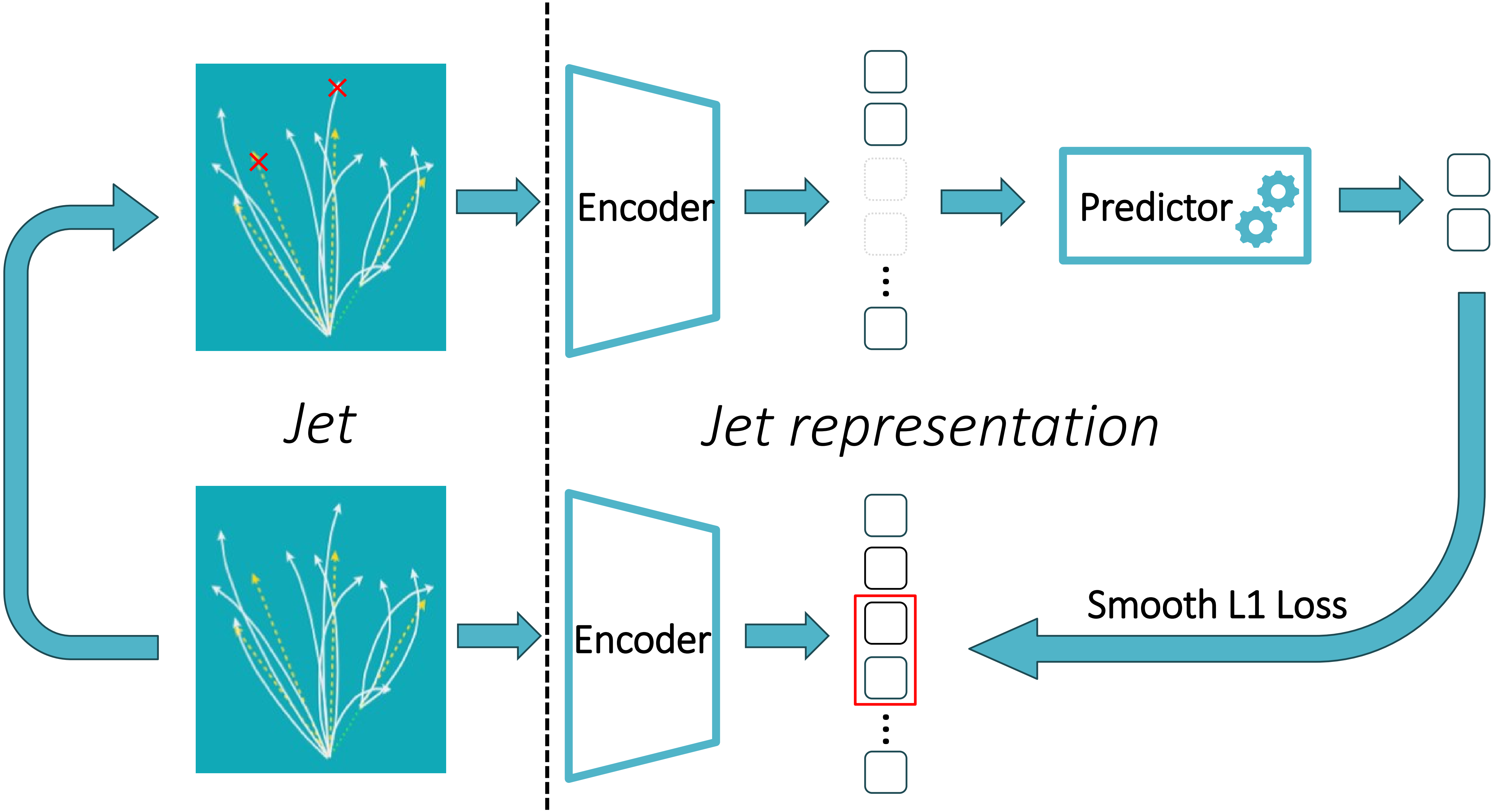
Our Approach

Plot modified from 2301.08243

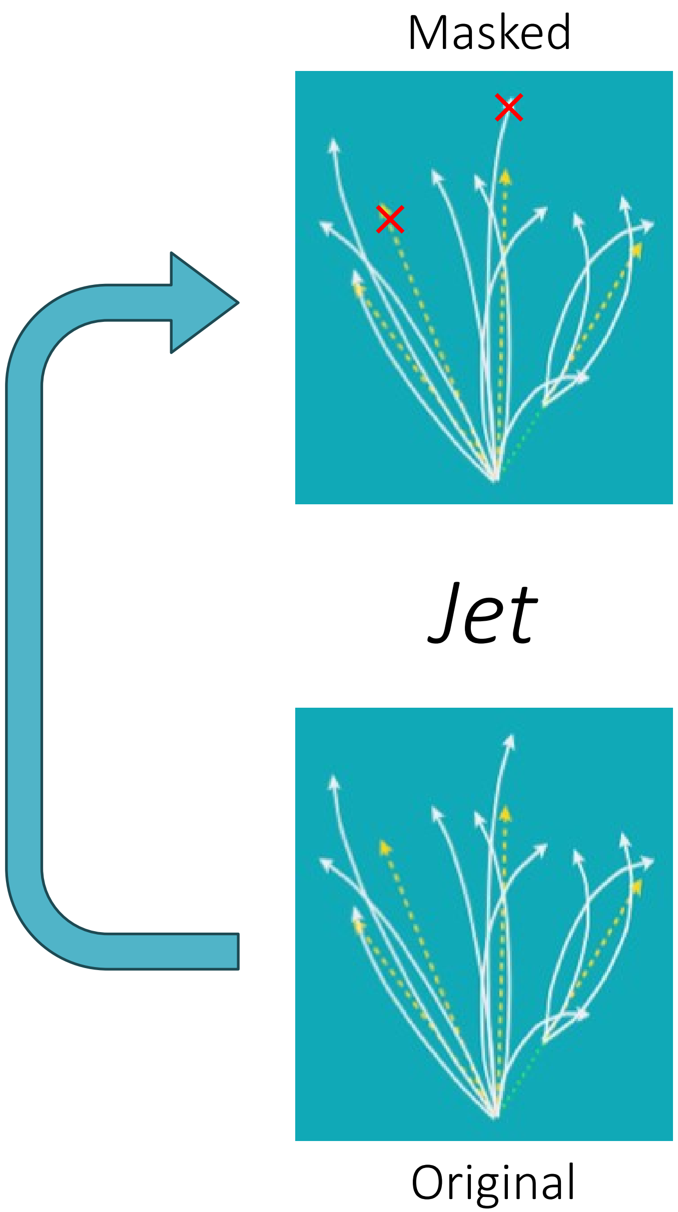
How to make it?

Implementation of the p-jepa network

Particle Joint-Embedding Predictive Architecture

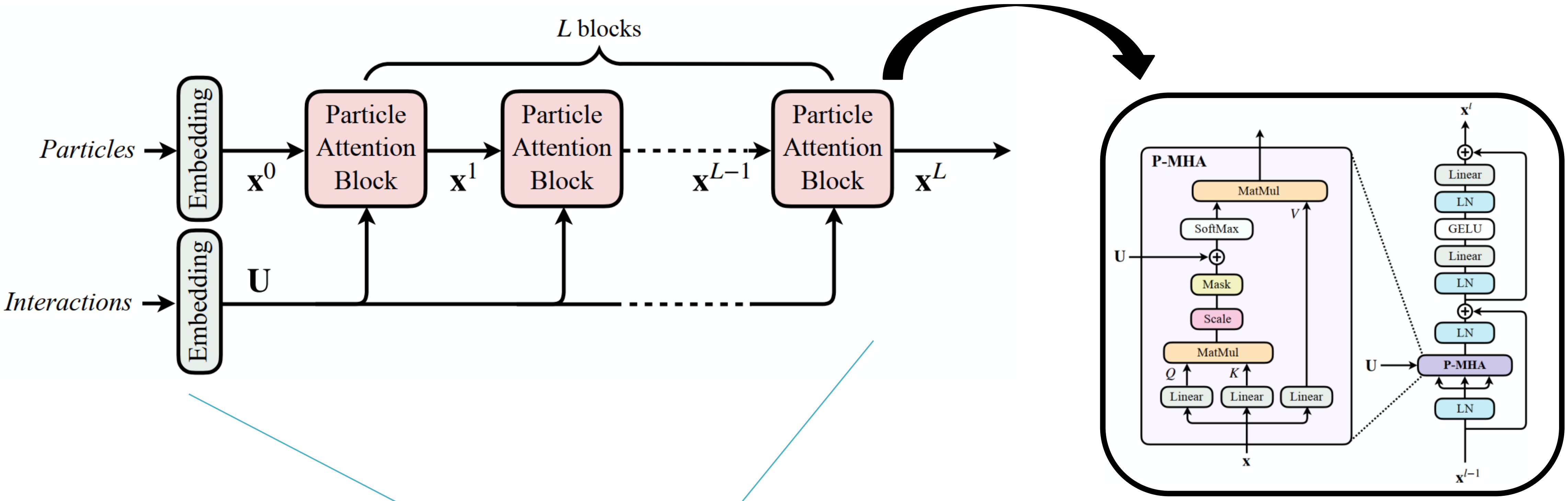


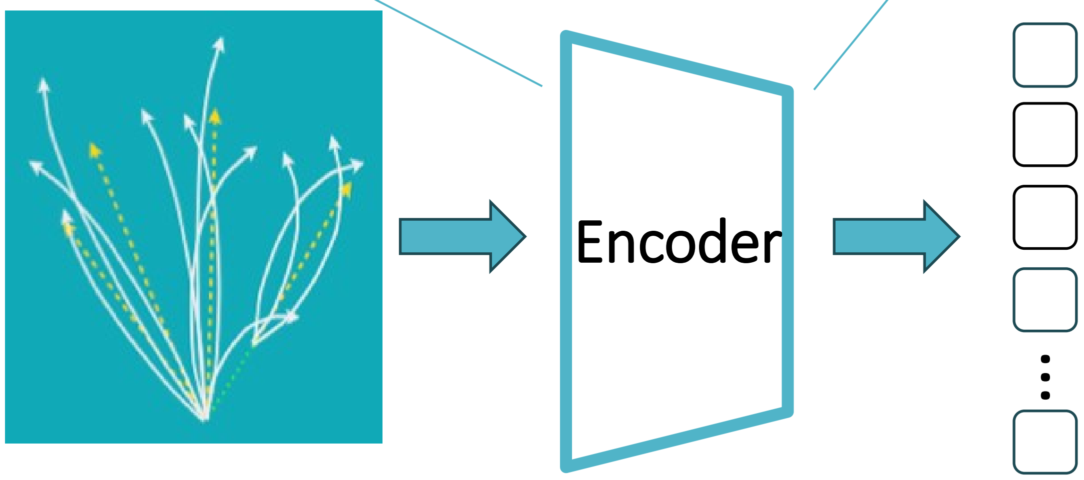
>>> Particle masking: the “question maker”



- Building blocks of a jet: particles
  - Kinematics (4-vec), PID, charge, track information
  - Correlation info, e.g. pairwise features and substructure
- Can ML learn to predict masked particles?
  - Randomly masking ~30% of particles in a jet
  - The remaining particles provide “context” information
  - Trying to recover the masked particles ("target") from the context
  - → Learn meaningful jet representations

>>> Encoder: representing the context





- Build on Particle attention block[1]
- Self-Attention
  - Extract information from particles
- Generalize to all particle info
  - Kinematics, charge, PID, track, etc

[1] *Particle Transformer* 2202.03772

>>> Predictor: the "question solver"

- Predict partial jet representation

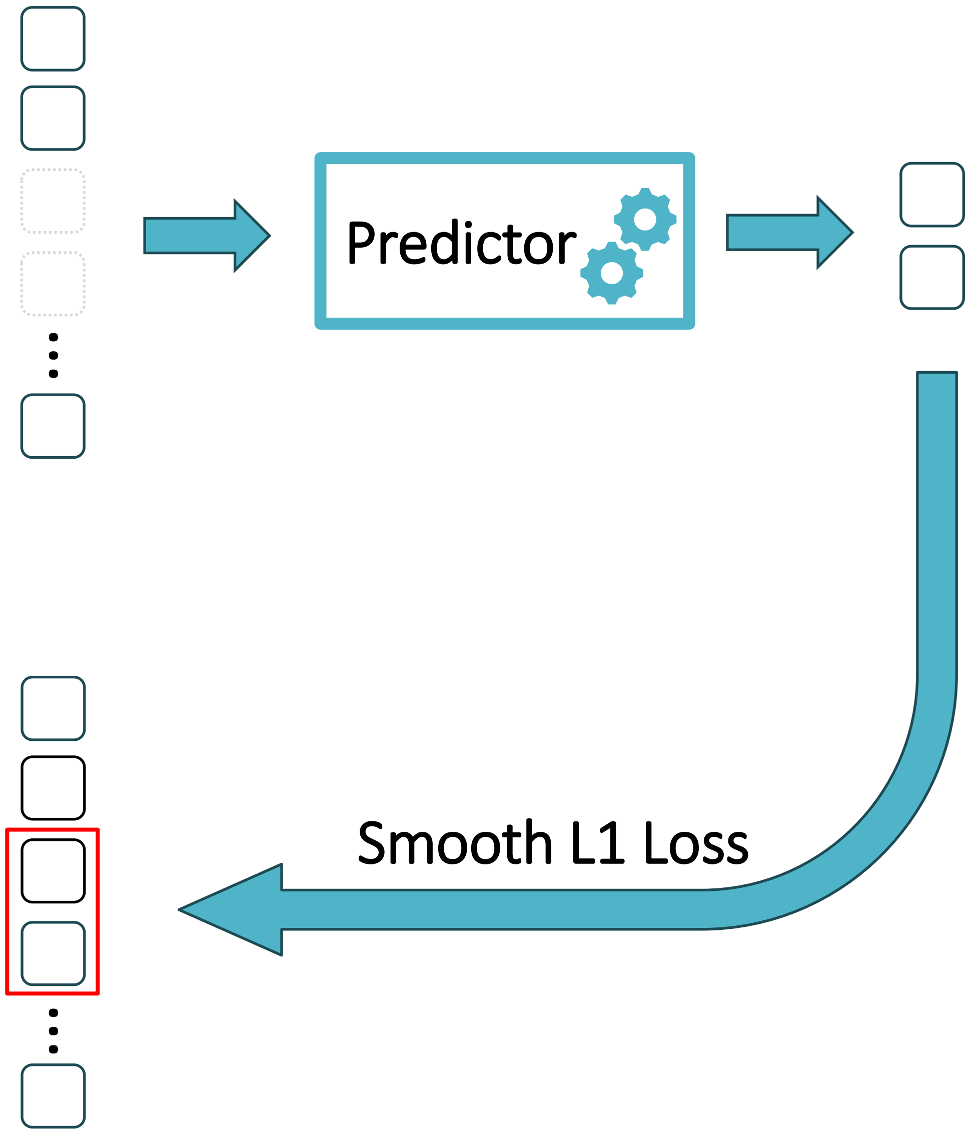
  Corresponding to the masked particles

- Smooth L1 loss

  Measure how close the predicted particles are to the truth in the representation space

- Encoder and predictor trained simultaneously

  → Aim to learn meaningful jet representation

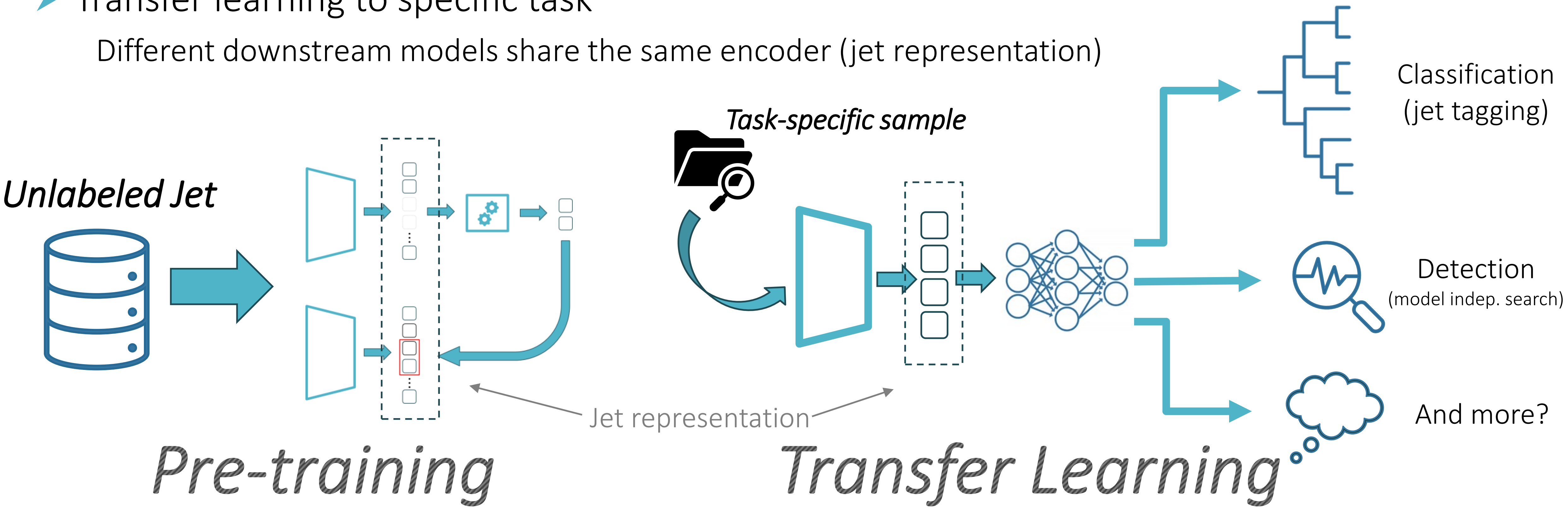


Does this work?

Experiments and Preliminary Results

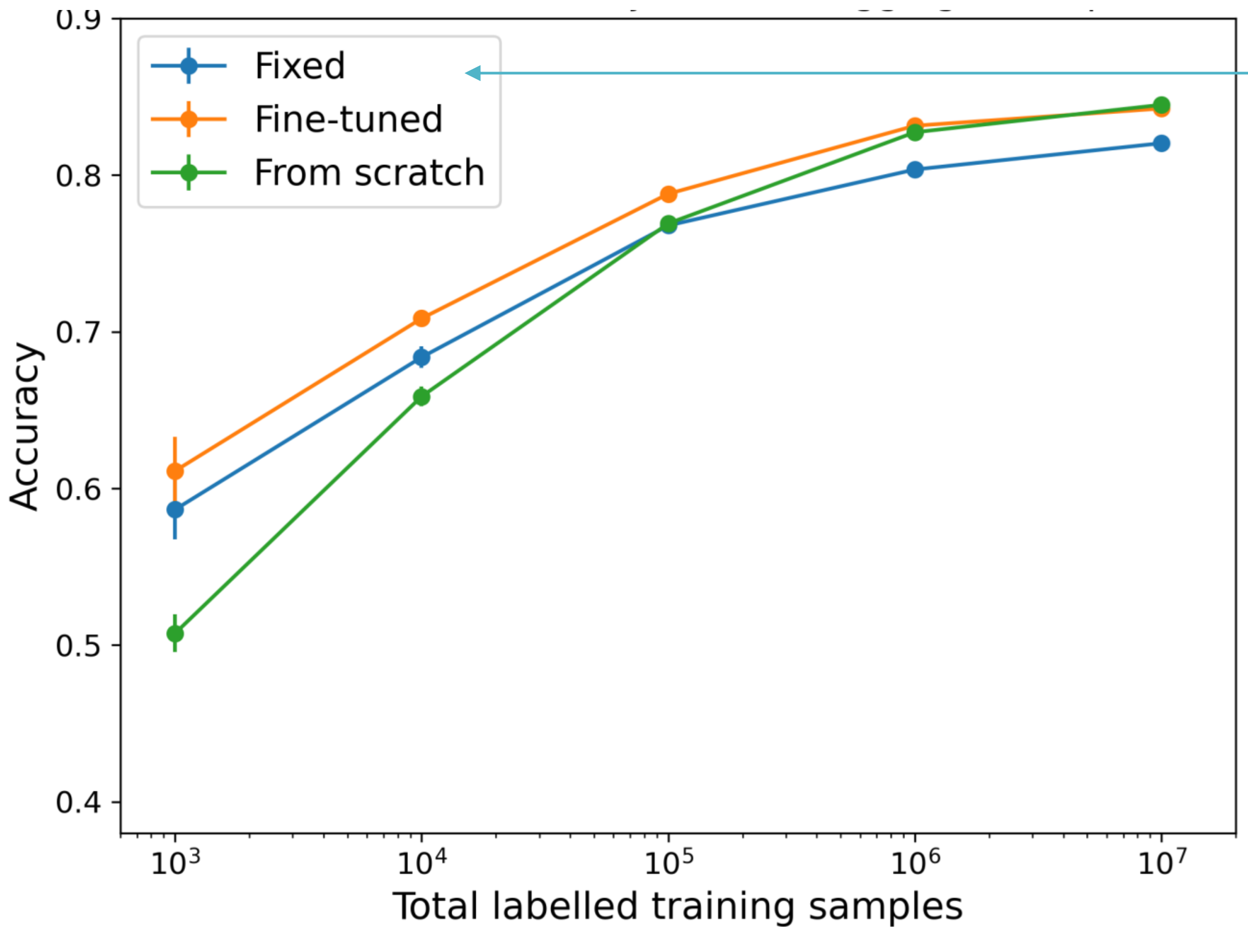
Pre-training and Transfer Learning

- Physics performance evaluated with pre-training + transfer learning pipeline:
- Foundation p-jepa model pre-trained on "data"
  - From JetClass-II: AntiKt(R=0.8), DELPHES simulation and realistic pileup effect (mu=50)
  - Composition emulated the real data (QCD >70% of training data, others follow cross-section)
- Transfer learning to specific task
  - Different downstream models share the same encoder (jet representation)



Application: Jet Tagging

- Few-shot transfer learning for jet tagging:
  10-class(*) classification on JetClass-I: different dataset with pre-train (no PU effect and balanced class)



Fixed: jet representation fixed when jet tagging task is trained

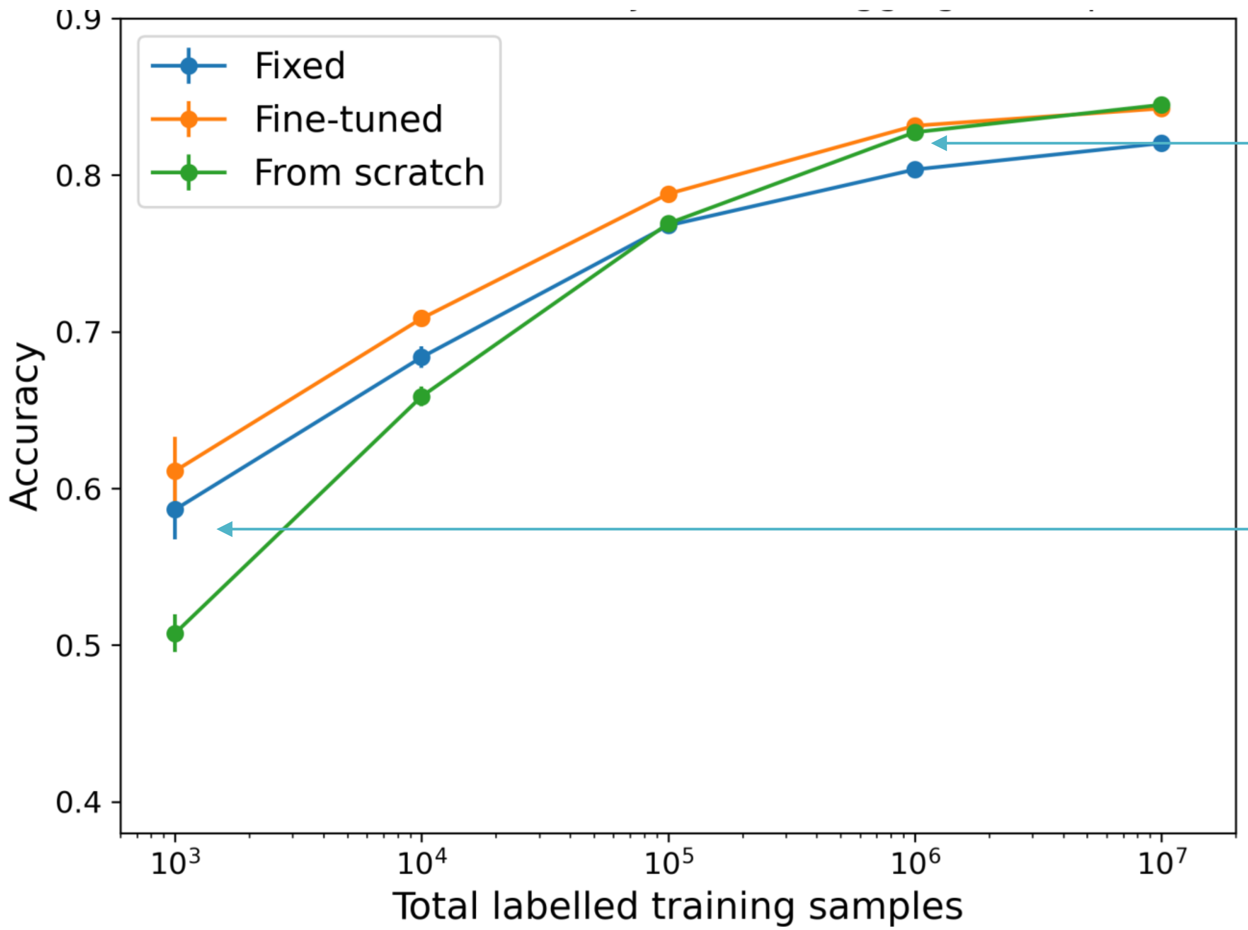
Fine-tuned: jet representation allowed slightly updating when tagging task is trained

From scratch: same network trained without pre-learned representation

$*: H(bb)/H(cc)/H(gg)/H(4q)/H(lvqq')/t(bqq')/t(blv)/W(qq')/Z(q\bar{q})/QCD$

Application: Jet Tagging

- Few-shot transfer learning for jet tagging:
  10-class(*) classification on JetClass-I: different dataset with pre-train (no PU effect and balanced class)



From scratch training takes over when the labelled dataset is large enough

→reduce to fully-supervised jet tagging

Pre-training + transfer learning gives a significant performance boost with very limited number of labelled samples (as lower as 100 jet/class)!

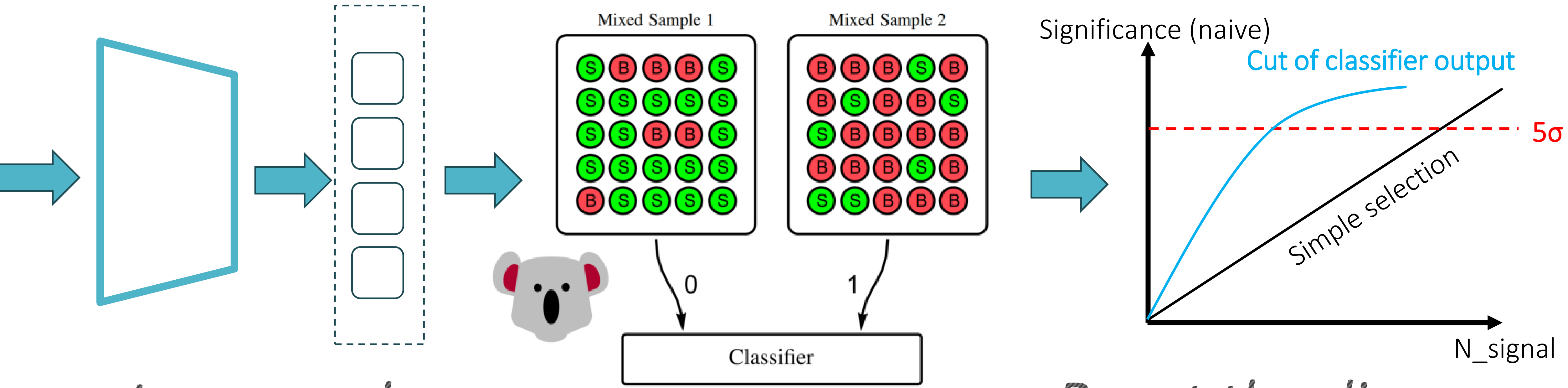
→ Benefit from jet rep. learned in SSL

*: $H(bb)/H(cc)/H(gg)/H(4q)/H(lvqq')/t(bqq')/t(blv)/W(qq')/Z(q\bar{q})/QCD$

Application: Anomaly Detection

- Test the pre-trained jet representations on anomaly detection
  Model independent search for new physics signals

Weakly supervised classification



Jet encode

Boost the discovery

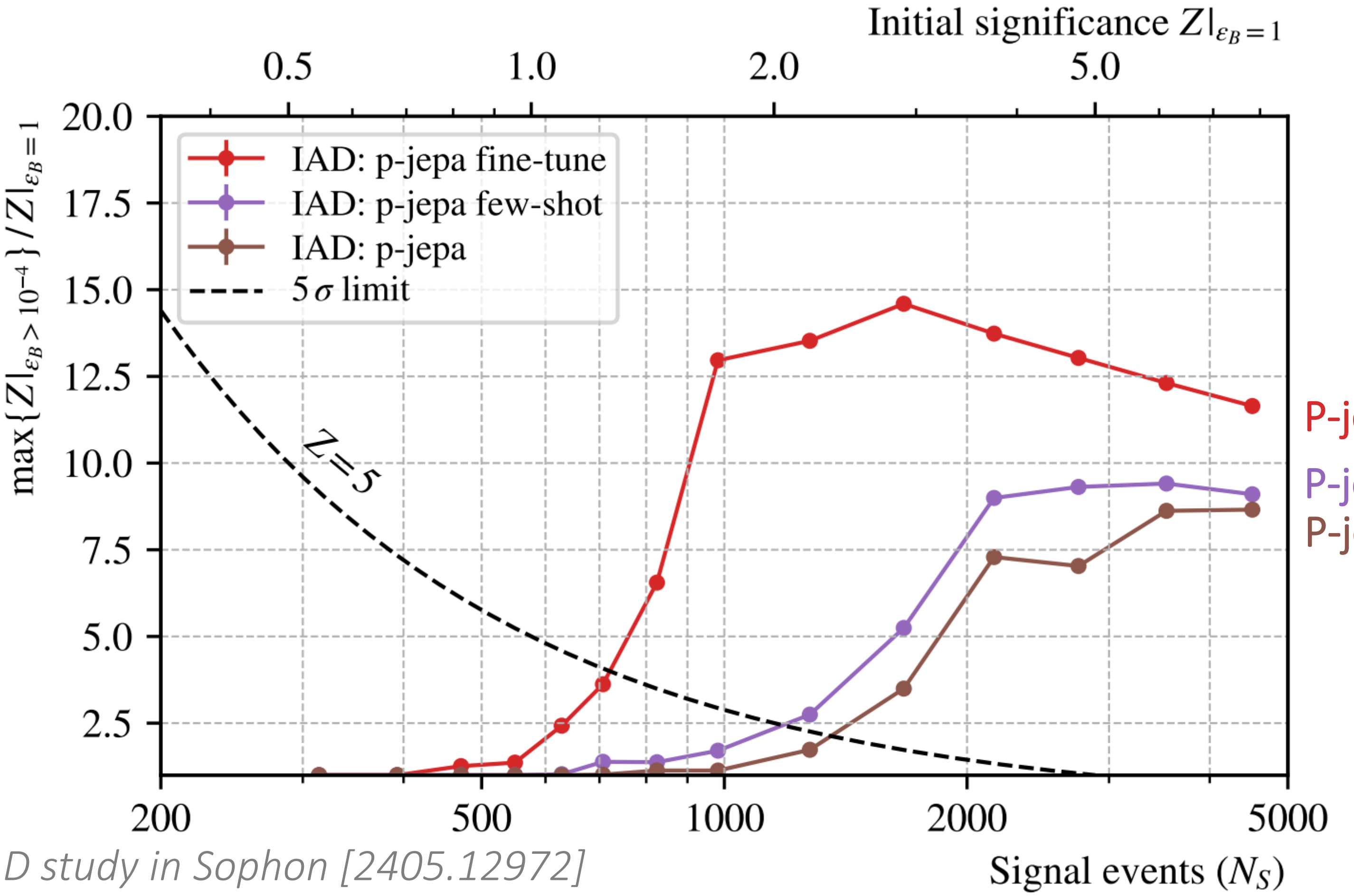
Share same framework of AD study in Sophon [2405.12972], originated from CWoLa [1708.02949]

Application: Anomaly Detection

- AD Significance enhanced using p-jepa:

  More visible after transfer learning on labeled jets

- Work in progress to reduce the gap with supervised way (e.g. Sophon)



P-jepa + finetune (10M)

P-jepa + few-shot (1k)

P-jepa

Share same framework of AD study in Sophon [2405.12972]

Summary and Outlook

- Proposed P-JEPA architecture for self-supervised learning on jets
- Jet representation learned from unlabeled data
- Performance tested on jet tagging and anomaly detection
- More applications in progress -- stay tuned!
- Take-away:

  Learning from jet without label is possible

  Joint-predictive architecture shows promising performance

  If data itself provides the knowledge, why not take it?

  BOOSTIAMO the new physics search in a self-supervised way!

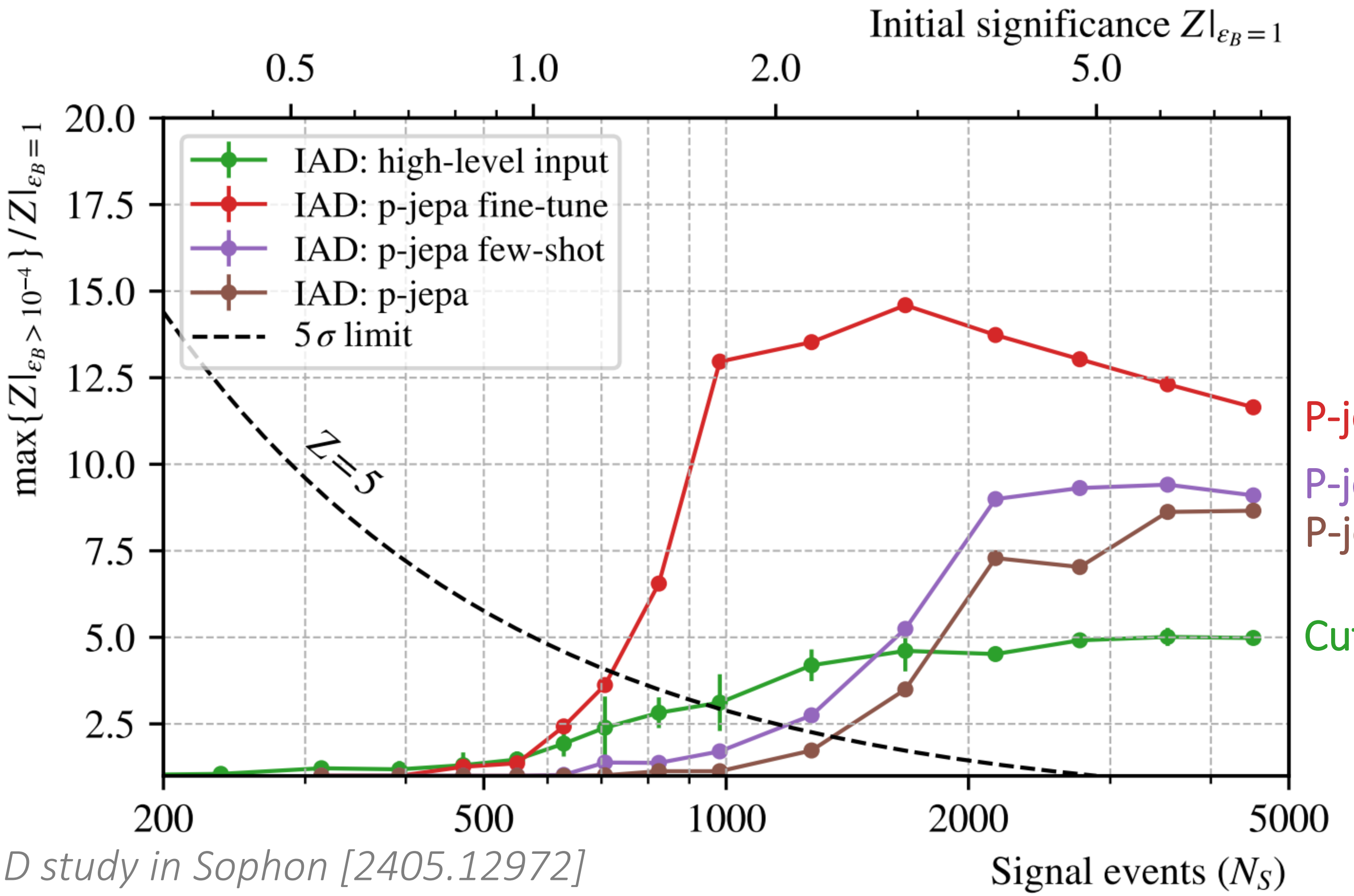
Backups

Application: Anomaly Detection

- AD Significance enhanced using p-jepa:

  More visible after transfer learning on labeled jets

- Work in progress to reduce the gap with supervised way (e.g. Sophon)



P-jepa + finetune (10M)

P-jepa + few-shot (1k)

P-jepa

Cut on high level input

Share same framework of AD study in Sophon [2405.12972]