Learning powerful jet representations via self-supervision

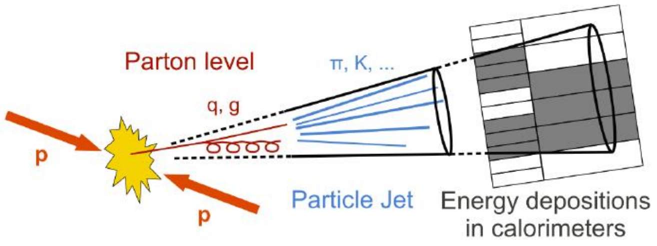
Qibin Liu, **Shudong Wang**, Congqiao Li, Huilin Qu

Quantum Computing & Machine Learning Workshop 2024, Changchun, China

7 August, 2024

Introduction

- Jets are collimated sprays of particles initiated by quarks or gluons
  - Ubiquitous at hadron colliders, carry rich information
- Jet tagging: identifying the hard scattering particle that initiates the jet
  - examples:
    - heavy flavor tagging (bottom/charm)
    - heavy resonance tagging (top/W/Z/Higgs)
    - quark/gluon discrimination
    - exotic jet tagging (displaced, 4-prong, …)
  - powerful tools for many new physics searches and standard model measurements



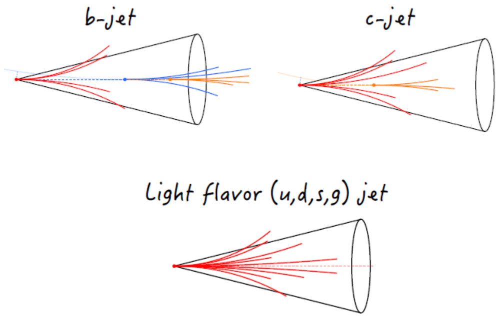


image credit

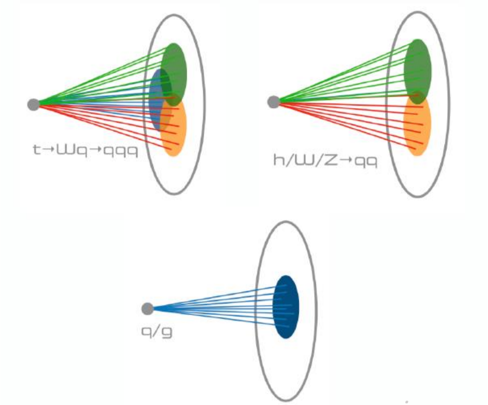
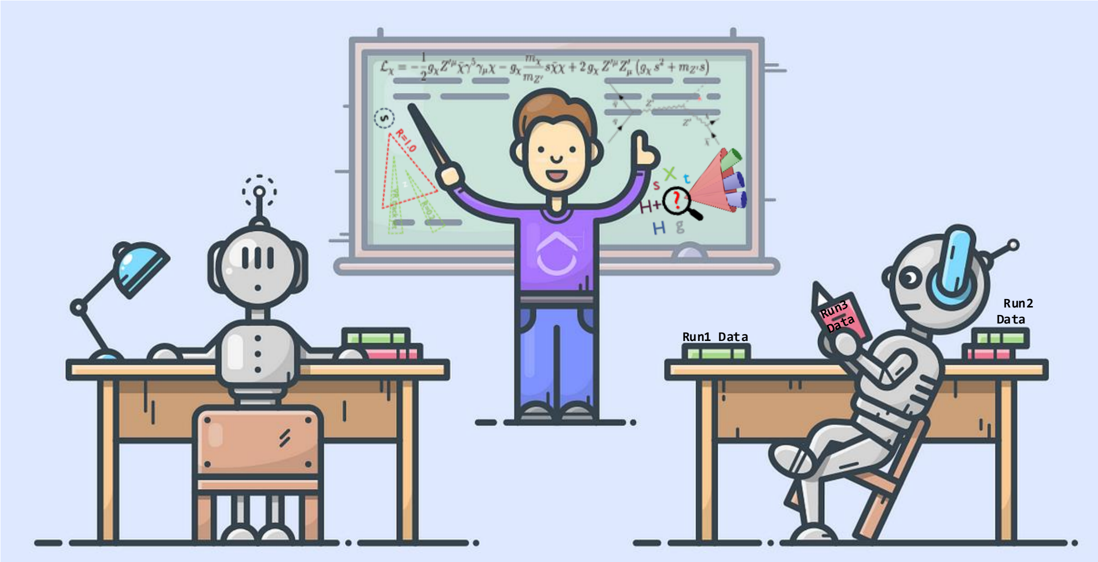


image credit

Introduction

- Various ML-based jet tagging algorithms have shown powerful performances
- Supervised Learning: strong performance while limited by labeled dataset
- We propose: learn jet representations through self-supervision using unlabeled dataset
- Applications on jet tagging and anomaly detection
- Outlook on future developments



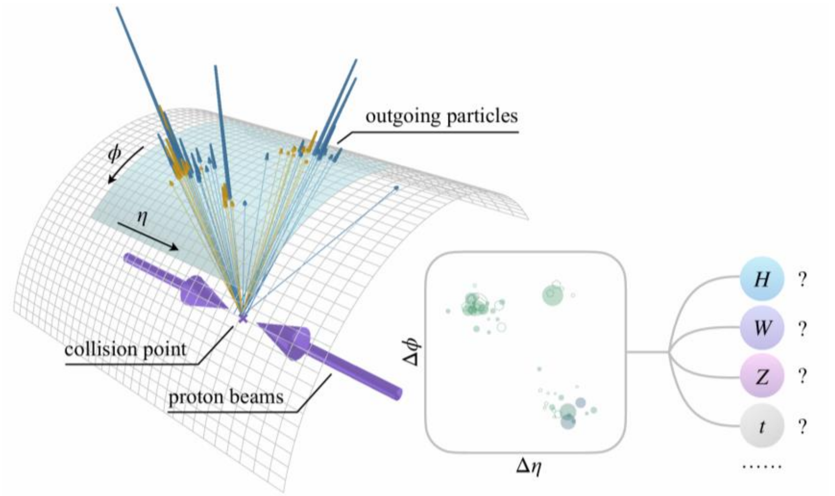
Plot modified from @Srinivas Rao

Prevailing intuition: train with truth label

Supervised learning on simulated dataset

Supervised ML for Jet Tagging

- Leverage information in a jet to identify it
- Incredible development over the past few years
    - EFN/PFN, ParticleNet, ParticleTransformer, PELICAN, OmniLearn, Sophon …
- All trained on simulated labeled dataset
    - Physics modelling, data-MC discrepancy

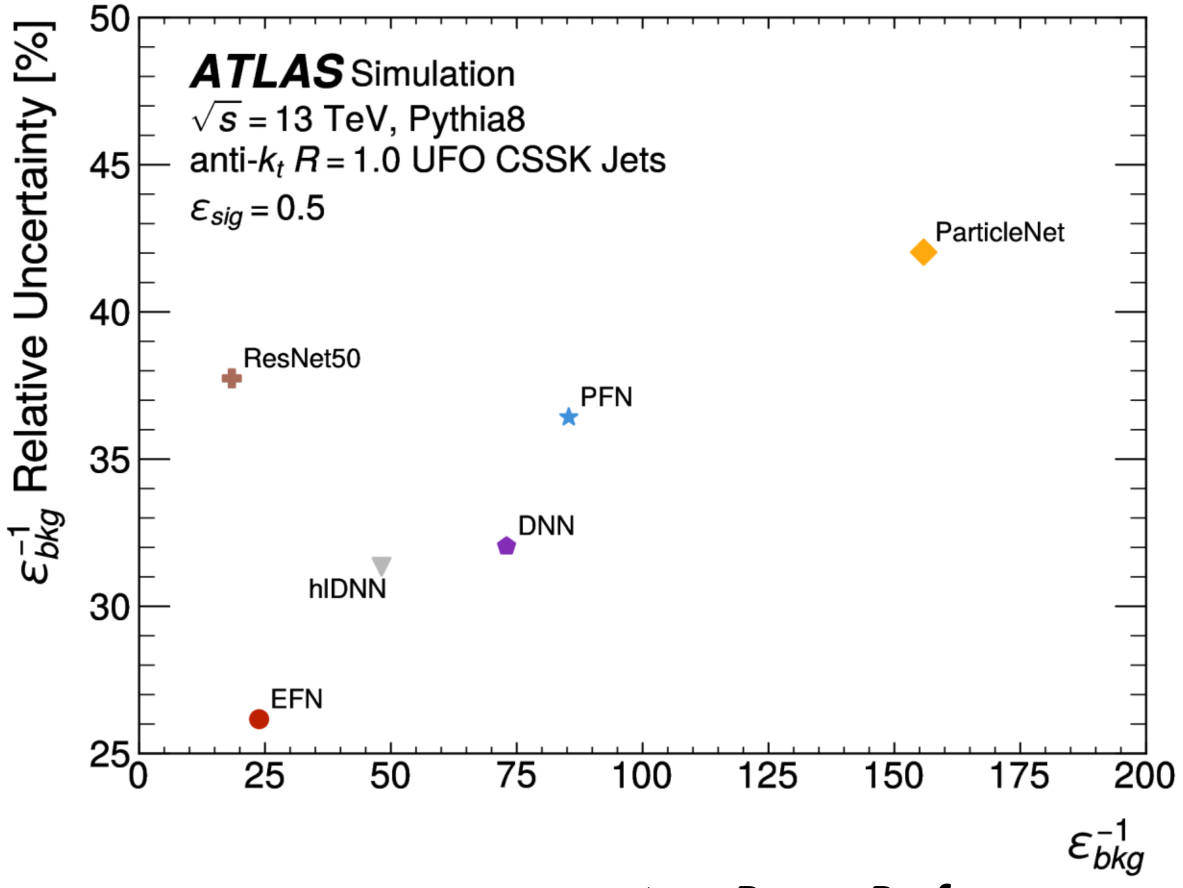


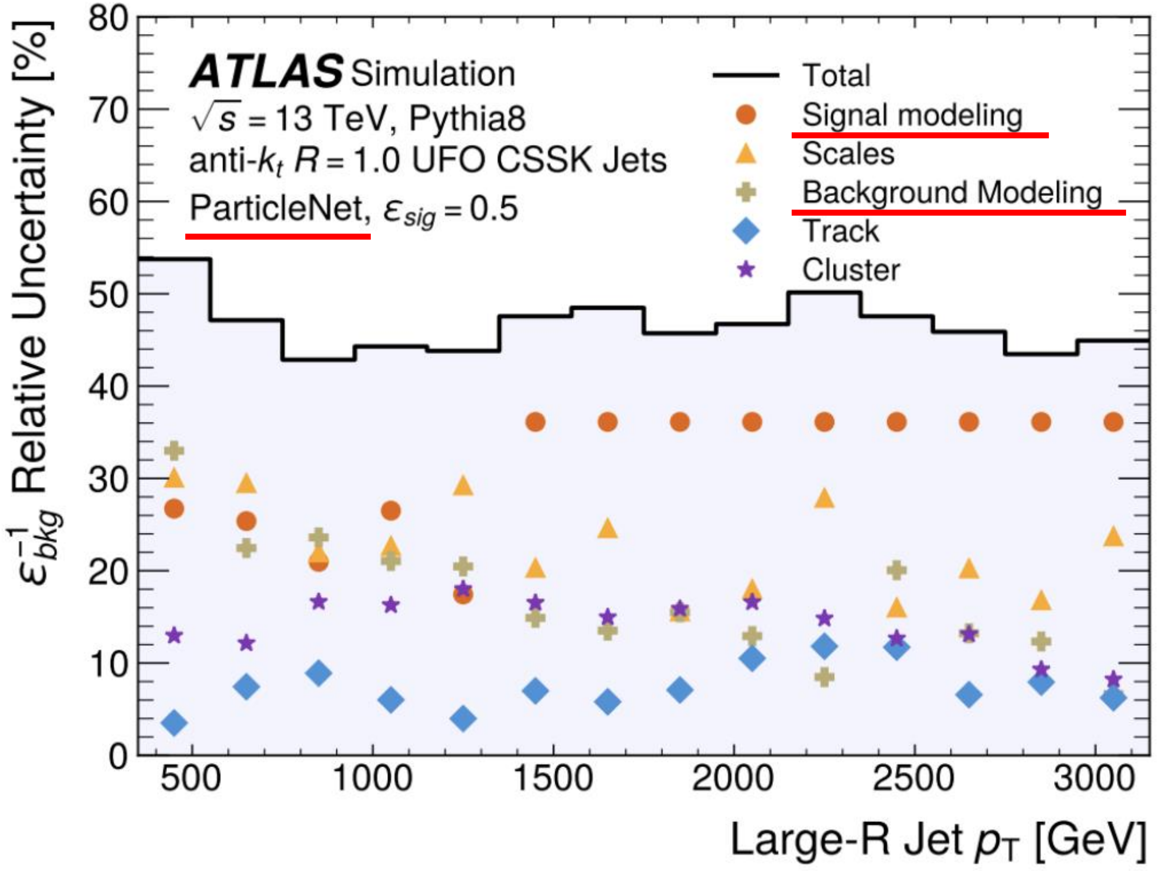
Plot credit: 2202.03772

Model Dependence of Jet Tagging

- **What ML algorithms are learning: Physics or Generators?**

Larger uncertainty





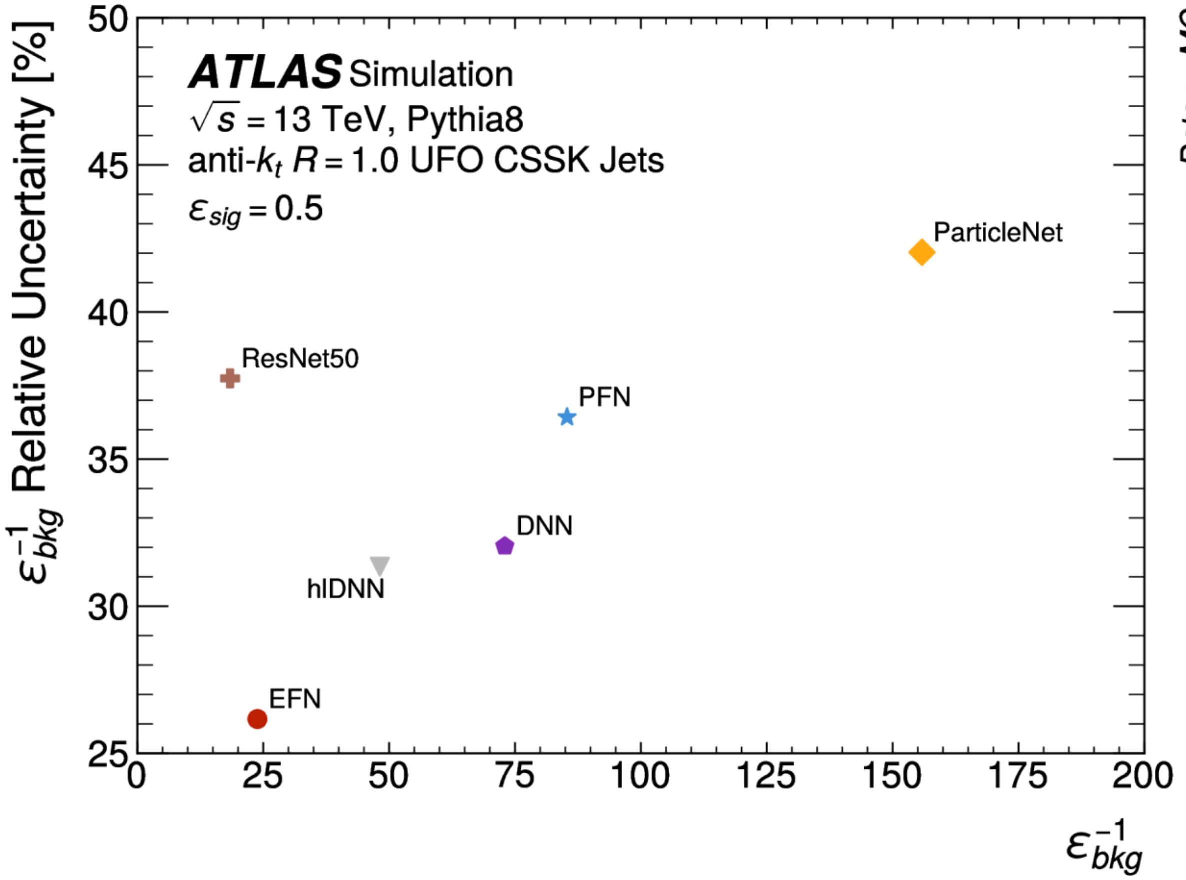
Better Performance

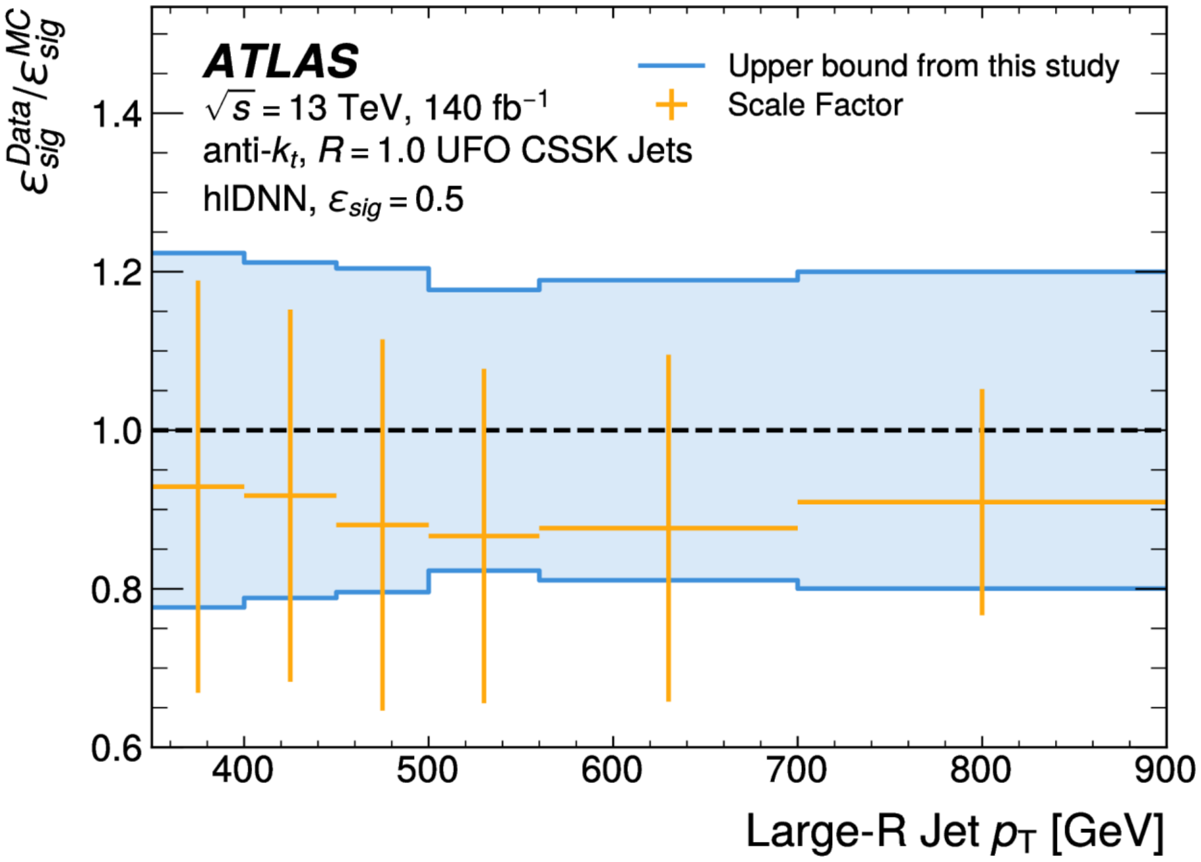
Plot credit: CERN-EP-2024-159

Model Dependence of Jet Tagging

- **Can we always calibrate it back to data?**

Larger uncertainty





Better Performance

Plot credit: CERN-EP-2024-159

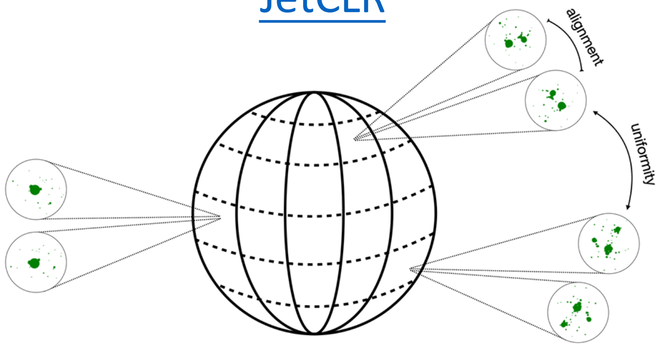
Learn from data directly?

The self-supervised paradigm

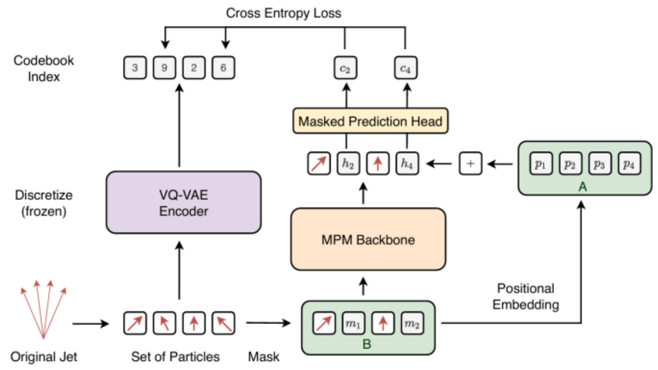
Self-supervised Learning

- Training with no label required
  - Could learn from data directly!
- Extract physics behind jets
  - Parton shower, hadronization, detector effects,…
  - Encourage algorithms to learn physics, rather than obsessed with minor details
- Learn comprehensive jet representations suitable for various applications
  - Jet tagging, generation, reconstruction, anomaly detection…
- Self-supervised learning on jets
  - JetCLR, MPM, OmniJet-α …

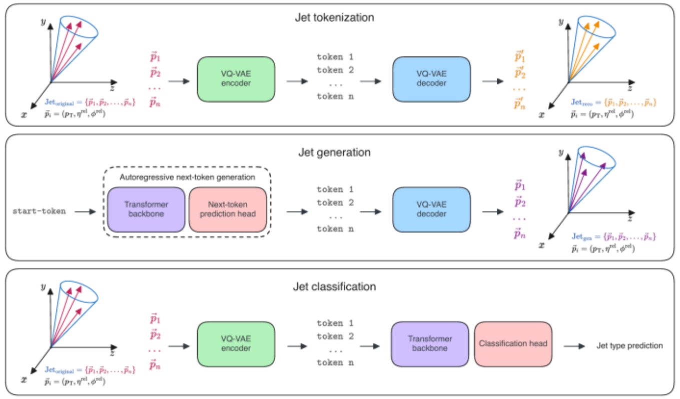
JetCLR



MPM

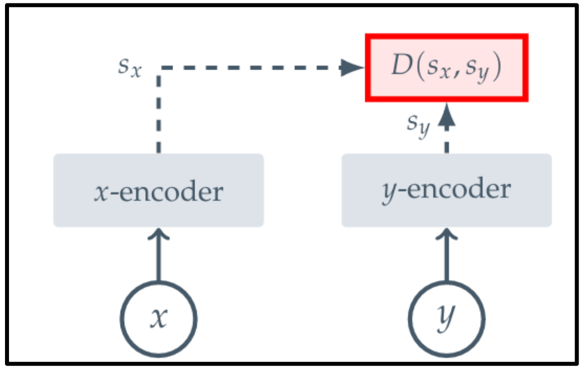


OmniJet-α

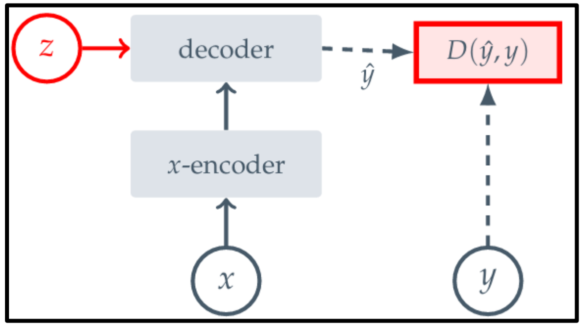


Architectures of Self-supervised Learning

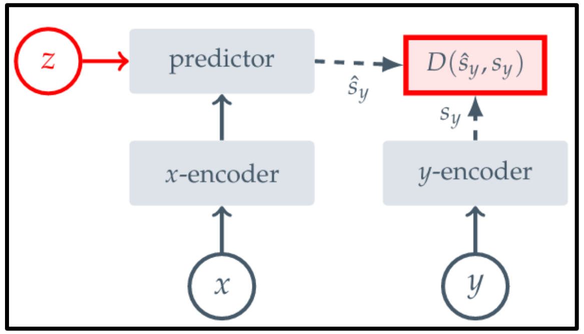
- Joint-Embedding Architecture (Contrastive)
  - Minimize/maximize distances between representations of similar/dissimilar jets
  - JetCLR



- Generative Architecture
  - Directly generate partial or full jets
  - MPM, OmniJet-α



- Joint-Embedding Predictive Architecture
  - Complete the representation of jets
  - Our approach, P-JEPA, inspired by I-JEPA

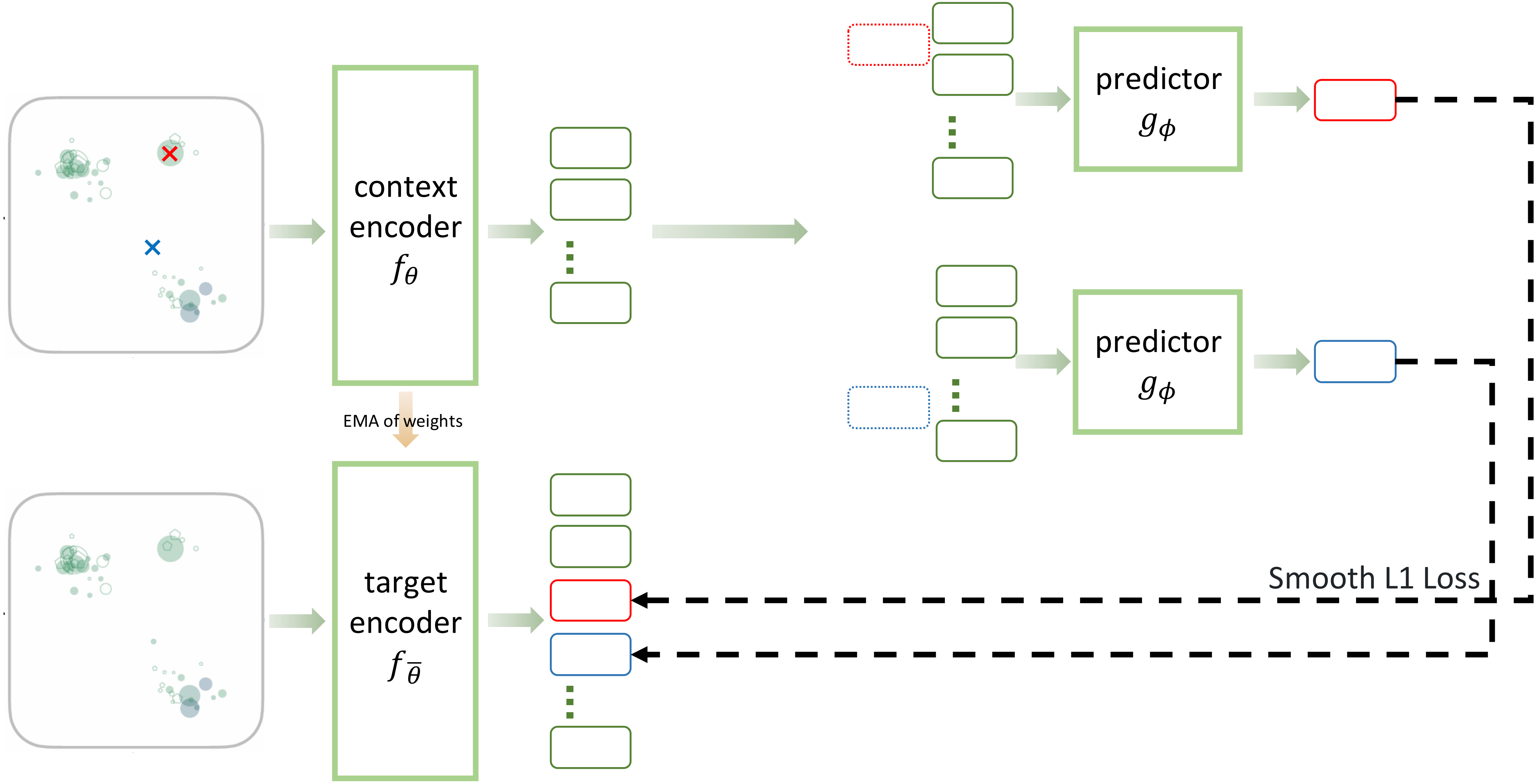


Plot credit: 2301.08243

Bringing the Concept to Life

Implementation of the P-JEPA network

Particle Joint-Embedding Predictive Architecture

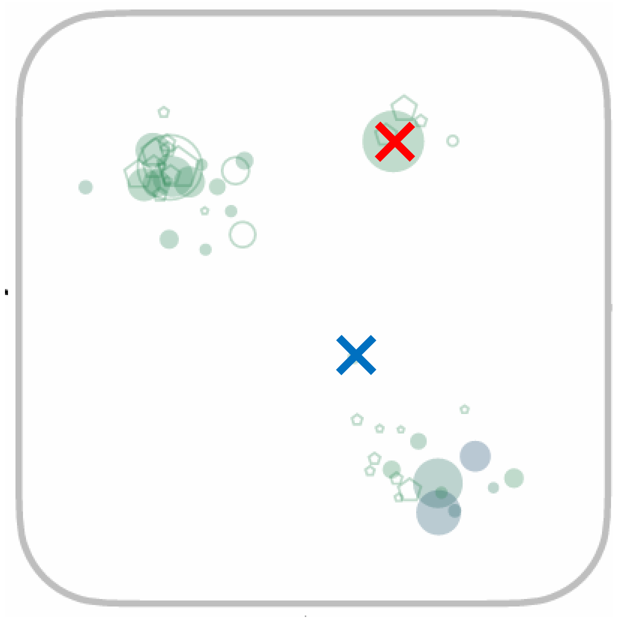


Particle Masking

- Jets are consist of particles
  - Own features: kinematics, charge, PID, track info
  - Correlation between each other: angular distance, invariant mass…

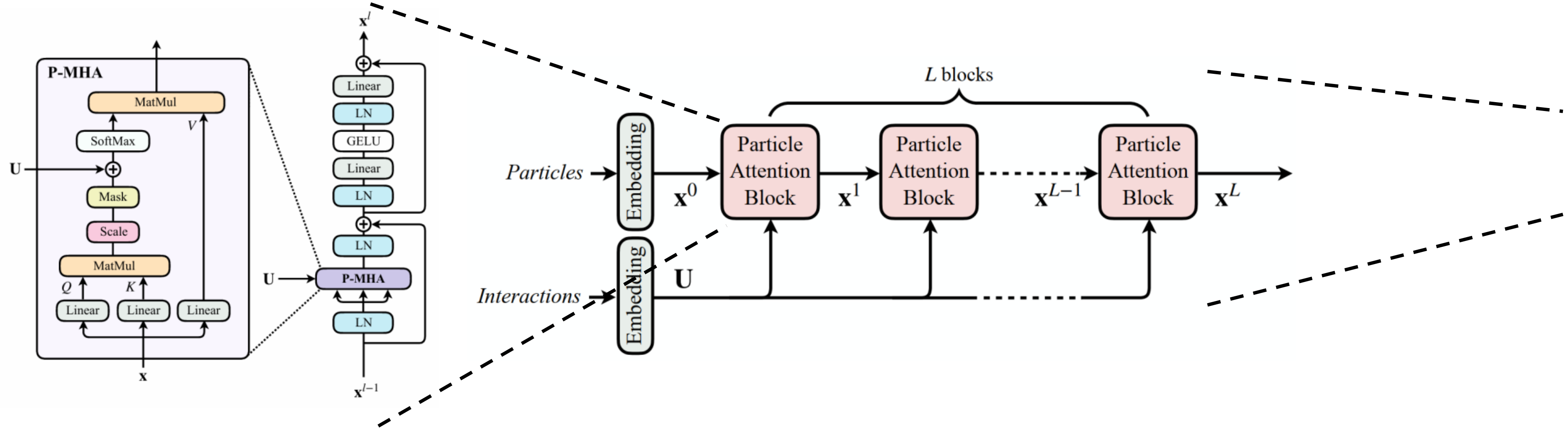
- Masking sets the task for training
  - Randomly masking ~30% of particles in a jet
  - Remaining particles provide “context” information for prediction

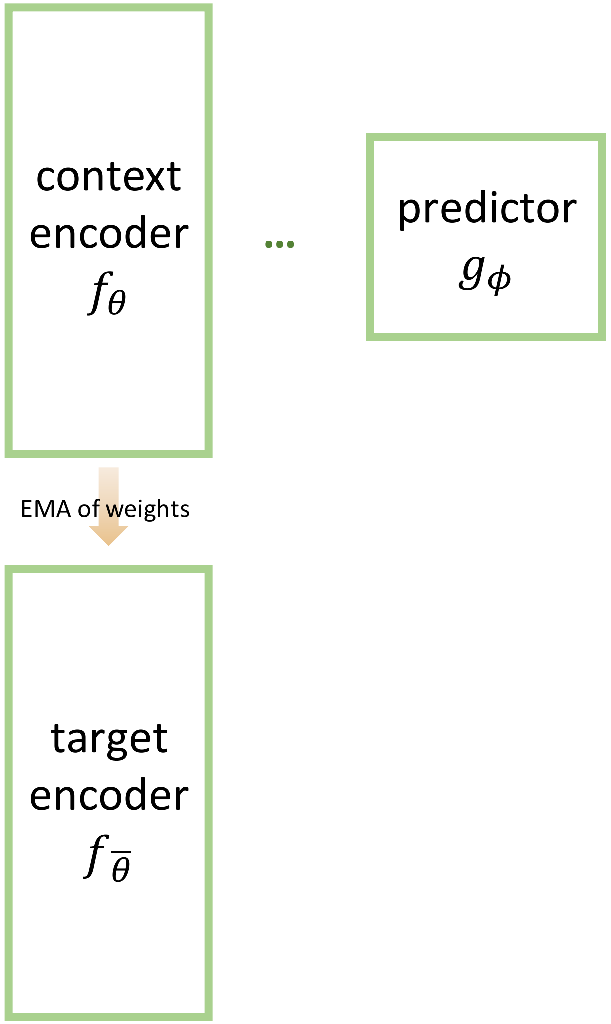
- Learning jet representations through predicting masked particles’ representations



Encoders and Predictor

- Particle Attention Block* as fundamental building block
  - Self-Attention & Pair-wise features
  - Accept all kinds of particle features
- Context encoder and target encoder share the same architecture
  - Weights of target encoder are updated via an exponential moving average (EMA) of the context encoder weights
- Predictor is narrower and shallower than encoders

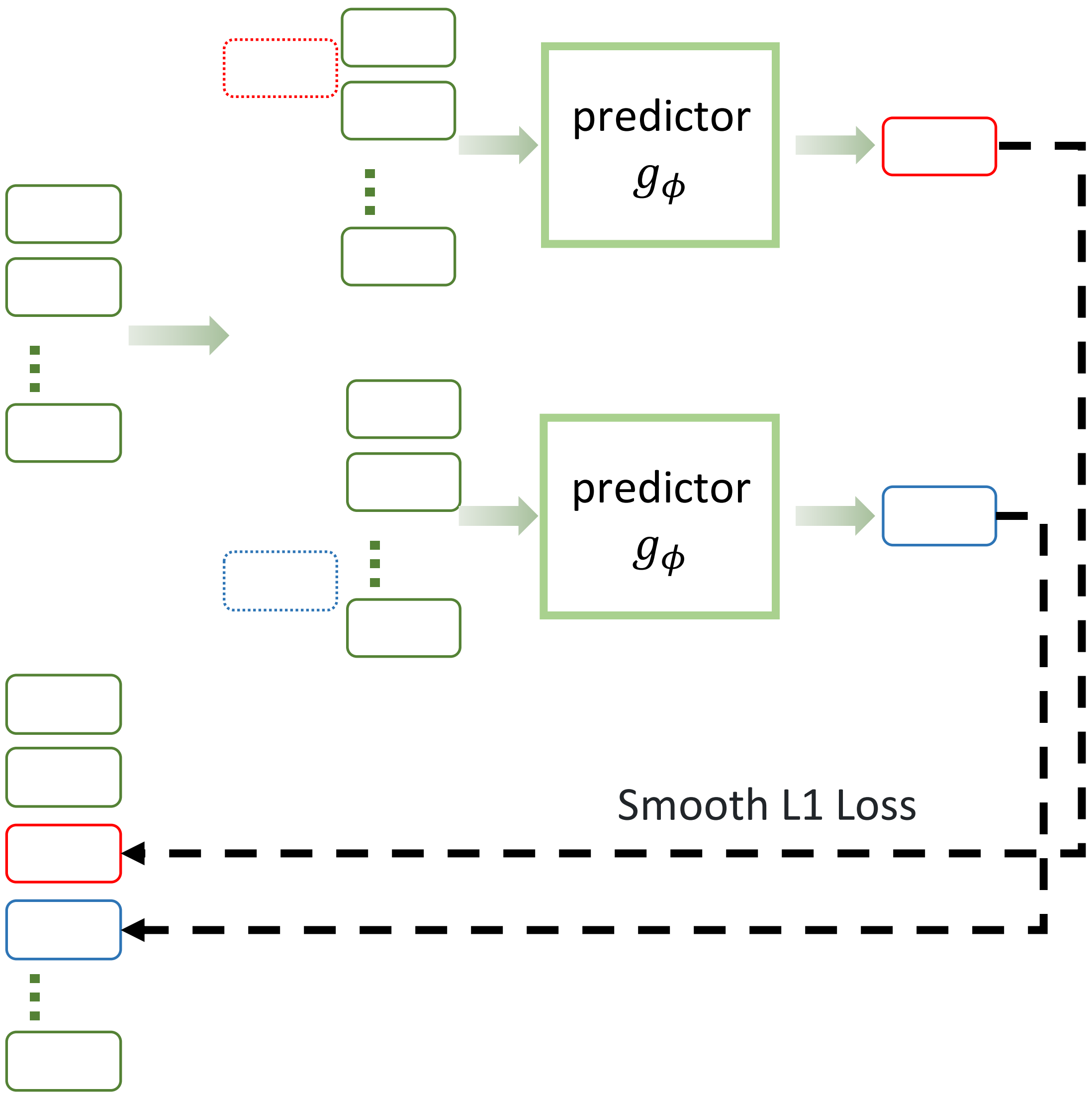




* From: Particle Transformer

Particle Representation Predicting

- Predictor: solve the task set by masking
  - Predict masked particles’ representations using context and auxiliary info (mask token)
- Smooth L1 loss
  - Measure how close the predicted particles are to the truth in the representation space
- Encoder and predictor are trained simultaneously
  - Aim to learn meaningful jet representation

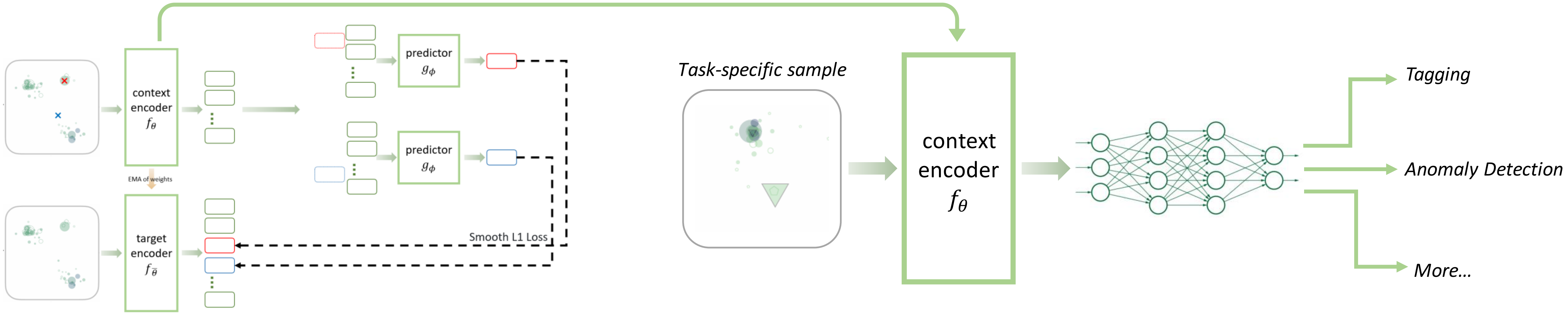


Does it work?

Experiments and Preliminary Results

Pre-training and Transfer Learning

- Performance evaluated with pre-training + transfer learning pipeline
- Foundation P-JEPA model pre-trained on "data"
  - From JetClass-II: anti-k_t(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)

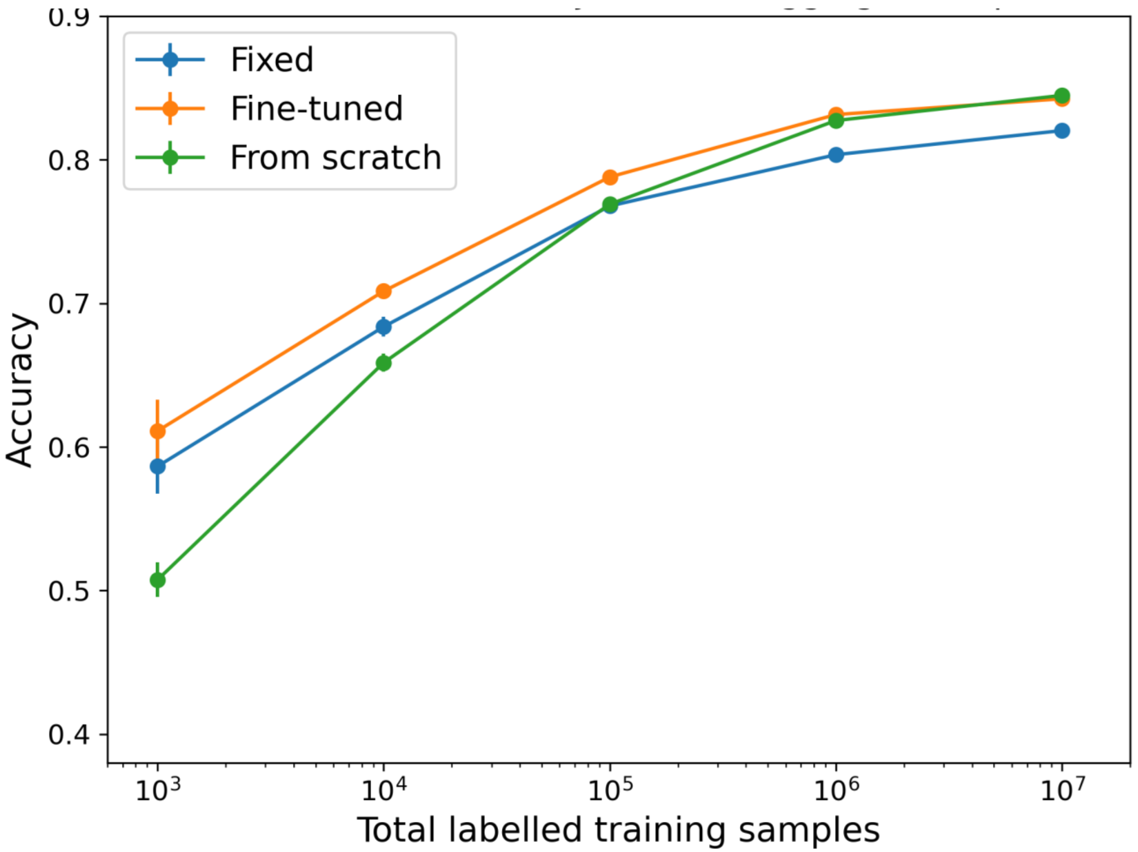


Pre-train on "data"

Transfer learning on MC

Application: Jet Tagging

- Few-shot transfer learning for jet tagging
  - 10-class* jet classification on [JetClass-I]



X-axis:
Training dataset size

Y-axis:
Fraction of correctly labeled jet in all 10 classes

Fixed:
Encoder fixed when jet tagging task is trained

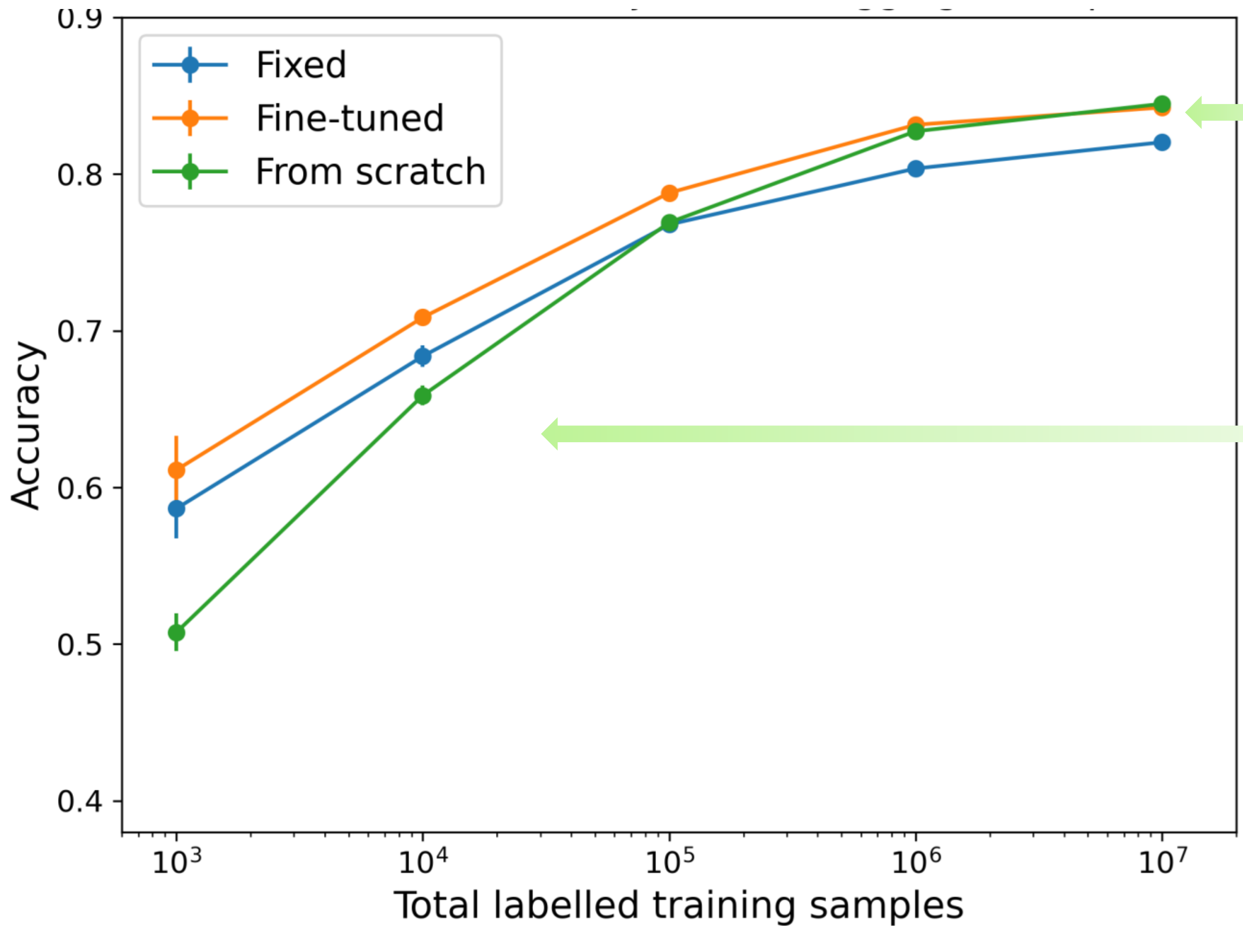
Fine-tuned:
Encoder allowed slightly updating when tagging task is trained

From scratch:
Identical network architecture but training started with randomly initialized weights.

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

Application: Jet Tagging

- Few-shot transfer learning for jet tagging
  - 10-class jet classification on JetClass-I

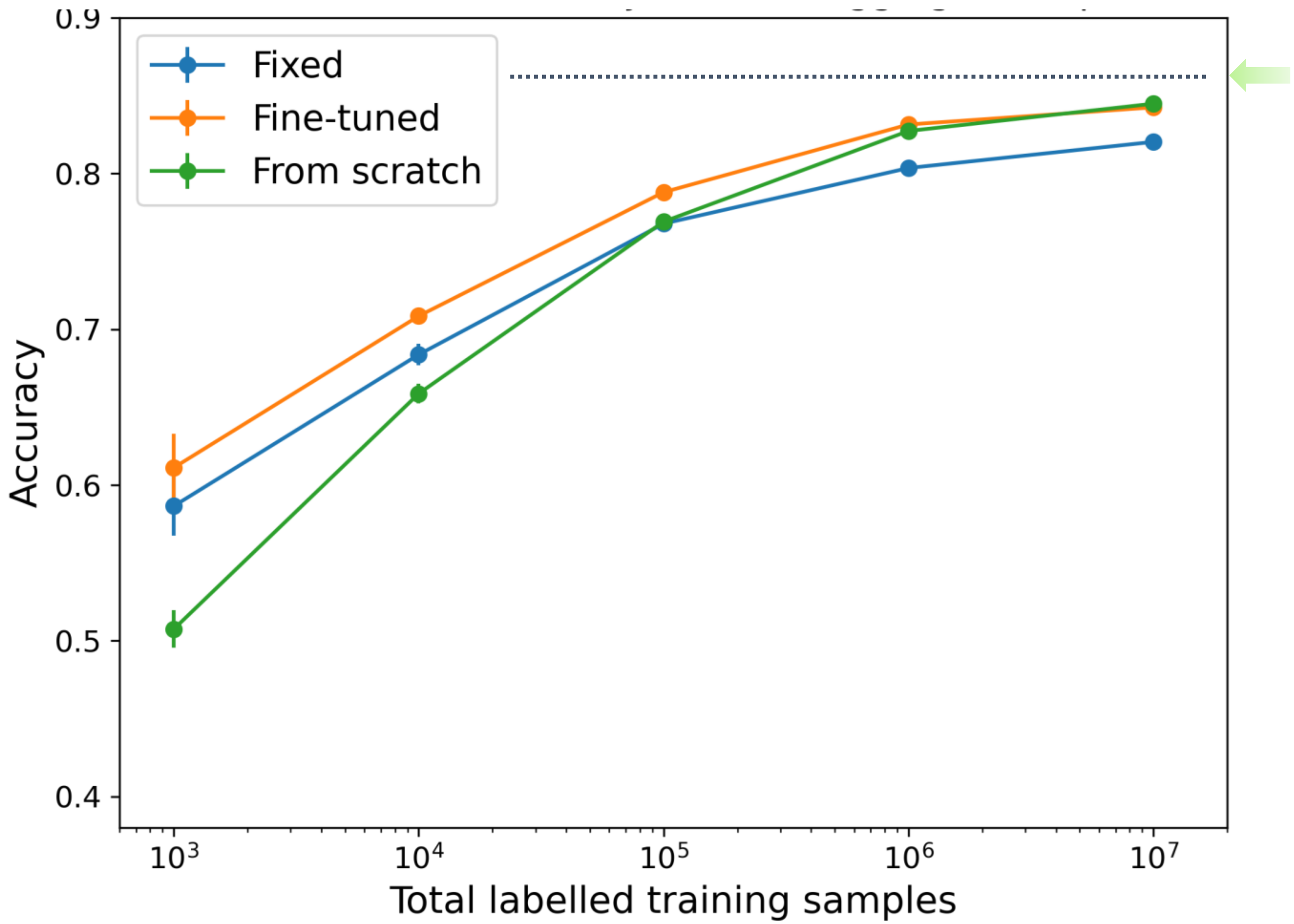


- From scratch training takes over when the labeled dataset is large enough
  - ➢ Converge to fully-supervised jet tagging

- Pre-training + transfer learning gives a significant performance boost with very limited number of labeled samples **(as low as 100 jet/class)!**
  - ➢ Benefit from jet representation learned via **self-supervised learning**

Application: Jet Tagging

- Few-shot transfer learning for jet tagging
  - 10-class jet classification on JetClass-I



How good the result is?

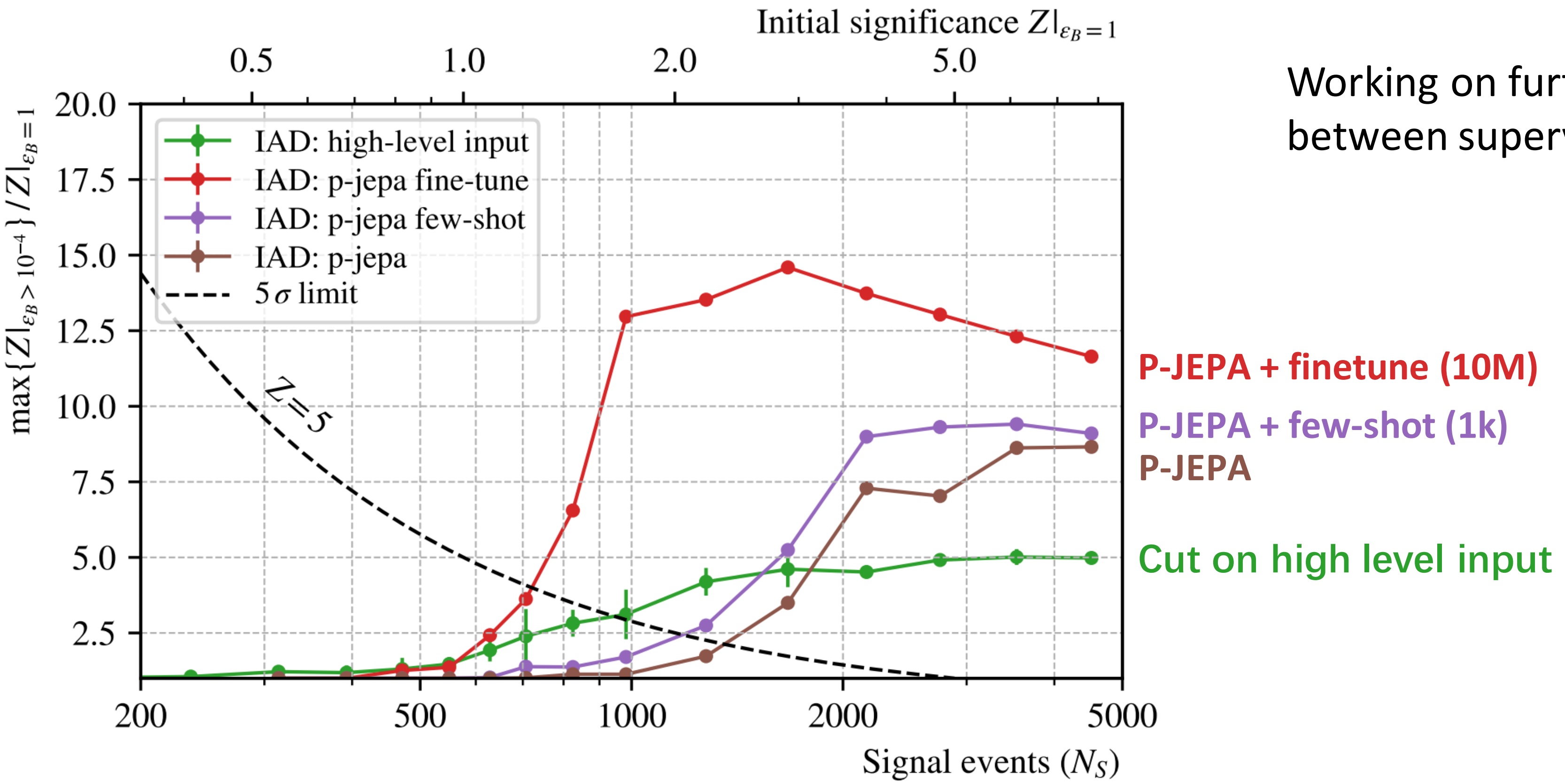
The SOTA of **supervised** model is 0.861 (trained on 100M jets)

This is current SOTA of **self-supervised** jet tagging

Working on further reducing the gap

Application: Anomaly Detection

- Test the effectiveness of pre-trained jet representations on anomaly detection
  - Share same framework of AD study in Sophon, originated from CWoLa
- Using the output of P-JEPA encoder as input to train the AD classifier
  - AD Significance enhanced using P-JEPA
  - More visible after transfer learning on labeled jets



Working on further reducing the gap between supervised way (e.g. Sophon)

Summary & Outlook

- Proposed the P-JEPA network for self-supervised learning on jets
- Performance tested on jet tagging and anomaly detection
- Effective jet representations can be learned from unlabeled dataset

- Outlook of further development and application: stay tuned!
  - Uncertainty-free or calibration-free jet tagging (ultimate goal though still long way to go)
  - Jet tagging at front-end: training on (unlimited) data stream; in-situ tagging on trigger level
  - MC-free anomaly detection (no MC used in full workflow, learn and inference all on data)