OmniJet-α: the first cross-task foundation model for particle physics

EuCAIFCon 2024, Amsterdam, May 1st 2024



Joschka Birk Anna Hallin Gregor Kasieczka

Universität Hamburg
DER FORSCHUNG | DER LEHRE | DER BILDUNG

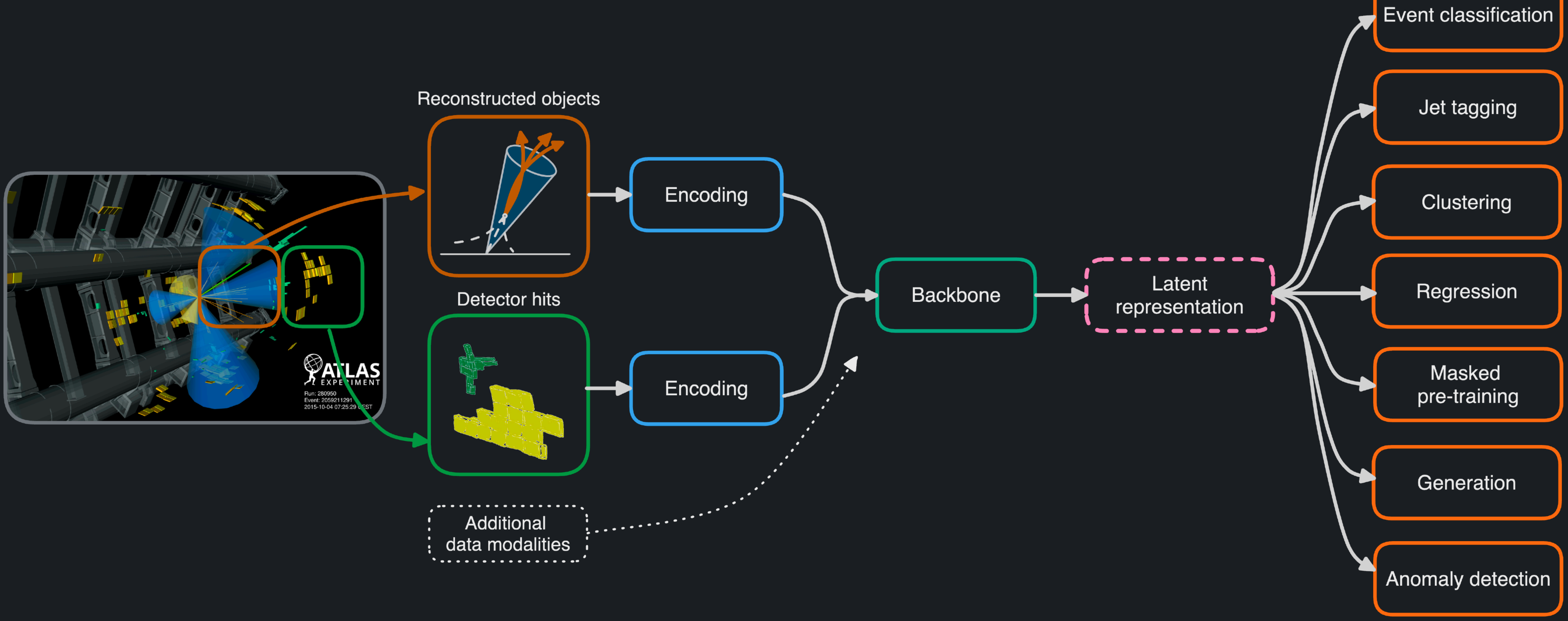
CLUSTER OF EXCELLENCE
QUANTUM UNIVERSE

joschka.birk@uni-hamburg.de

arXiv:2403.05618

Foundation models for HEP

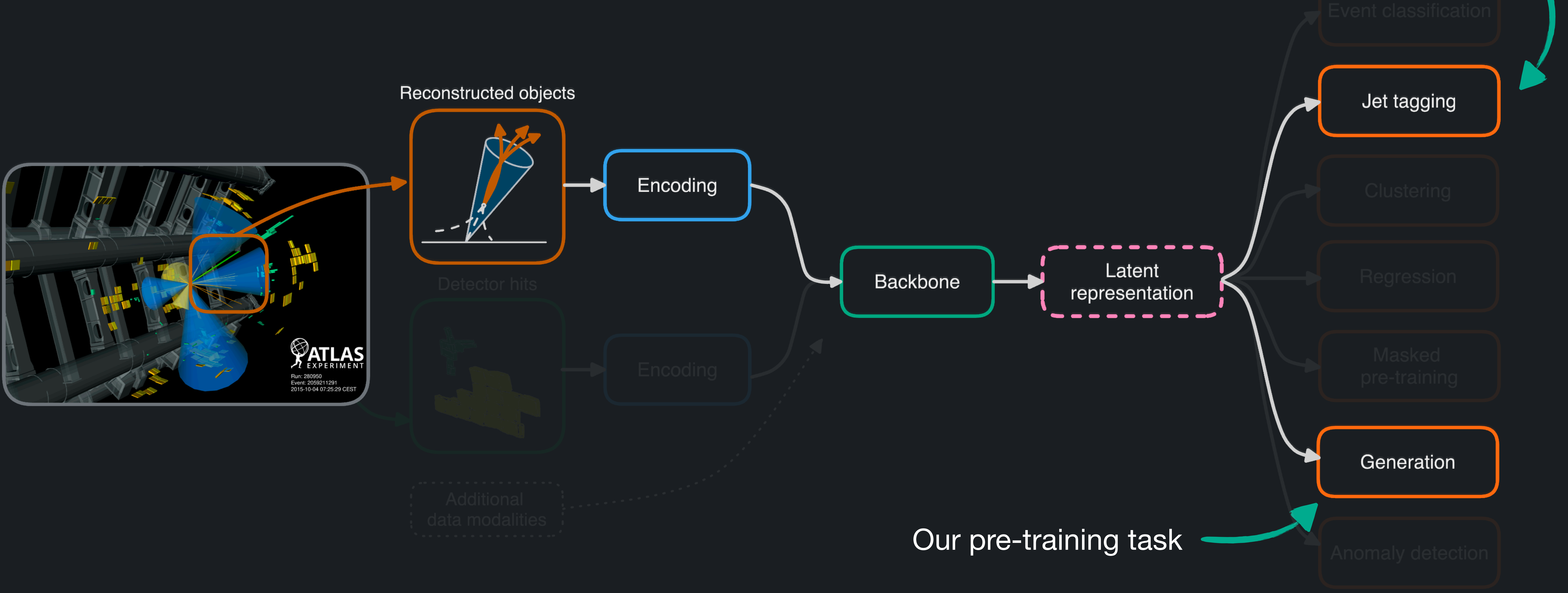
Pre-train on one task/dataset, **then fine-tune** on other task/dataset



Foundation models for HEP

Our fine-tuning task

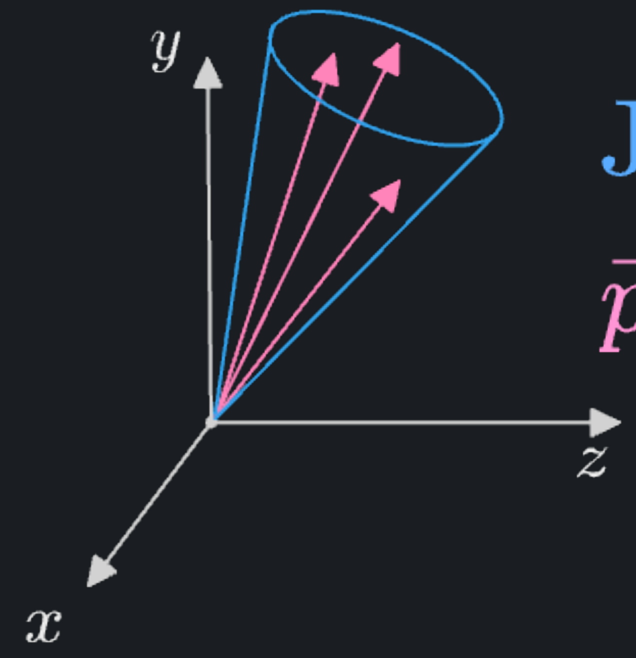
Pre-train on one task/dataset, **then fine-tune** on other task/dataset



Our pre-training task

Our approach

Jet constituents with **continuous features**



$$\text{Jet} = \{\vec{p}_1, \vec{p}_2, \ldots, \vec{p}_n\}$$
$$\vec{p}_i = (p_\mathrm{T}, \eta^\mathrm{rel}, \phi^\mathrm{rel})$$

Constituents are **tokenized with a VQ-VAE**
(using the approach presented by Sam Klein earlier)

$$\text{Jet} = \{\texttt{start-token}, \texttt{token}_1, \ldots, \texttt{token}_n, \texttt{end-token}\}$$
$$\texttt{token}_i = \text{integer value} \in [1, \ldots, 8192]$$

Unsupervised pre-training of transformer backbone on generative task (next-token prediction)

Autoregressive next-token generation

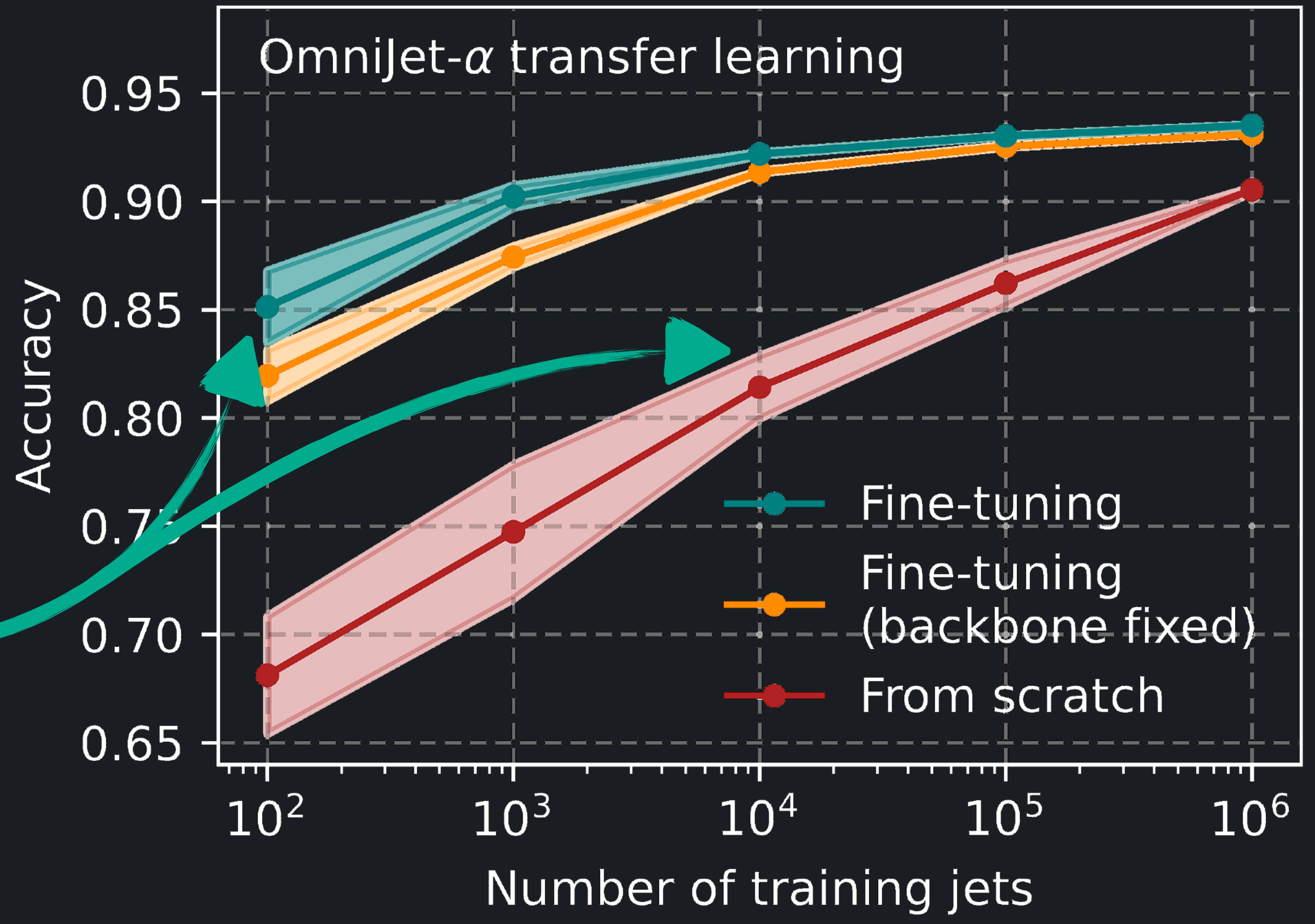
start-token
$\texttt{token}_1$
$\texttt{token}_2$
...
$\texttt{token}_n$
end-token

Fine-tuning to classification task: Swap model head and copy over the weights from the pre-trained backbone

start-token
$\texttt{token}_1$
$\texttt{token}_2$
...
$\texttt{token}_n$
end-token

Does generative pre-training help for classification?

- Classification: $t \to bqq'$ vs. q/g jets
- Generative pre-training with both jet types
- **Pre-trained / fine-tuned model** (●●) reaches same performance as from scratch training with **100-1000x less training samples**



→ ***Generative pre-training is a promising target for unsupervised pre-training in HEP***

Poster number: **54**