

Deep image reconstruction for background subtraction in heavy-ion collisions

Umar Sohail Qureshi^{1,2,3,*} and Raghav Kunawalkam Elayavalli³

¹*Physics Department, Stanford University, Stanford, California 94305, USA*

²*SLAC National Accelerator Laboratory, Menlo Park, California 94205, USA*

³*Department of Physics and Astronomy, Vanderbilt University, Nashville, Tennessee 37235, USA*



(Received 22 July 2025; accepted 14 October 2025; published 9 December 2025)

Jet reconstruction in an ultrarelativistic heavy-ion collision suffers from a notoriously large, fluctuating thermal background. Traditional background-subtraction methods struggle to remove this soft background while preserving the jet's hard substructure. In this Letter, we present DEEPSUB, a machine-learning-based approach for full-event background-subtraction. DEEPSUB utilizes a model based on Swin Transformer layers to denoise jet images and disentangle hard jets from the heavy-ion background. DEEPSUB significantly outperforms existing subtraction techniques by reproducing key jet observables such as jet p_T and mass, and substructure observables such as girth and the energy correlation function, at the subpercent to percent level. As such, DEEPSUB paves the way for precision heavy-ion measurements in hitherto inaccessible kinematic regimes.

DOI: [10.1103/dqnx-1fqk](https://doi.org/10.1103/dqnx-1fqk)

I. INTRODUCTION

CERN's Large Hadron Collider (LHC) and Brookhaven National Laboratory's Relativistic Heavy-Ion Collider (RHIC) collide ultrarelativistic heavy ions to offer a glimpse at the quark-gluon plasma (QGP), a state of matter where quarks and gluons exist deconfined from hadronic matter. Arguably the most powerful probes of the QGP are jets: collimated sprays of particles that result from the fragmentation of high-energy quarks and gluons produced in hard QCD scatterings [1]. Interactions with the QGP particles modify the jet showering process, thereby leaving the QGP's imprint on the jet itself.

This phenomenon of jet quenching, where highly energetic partons are expected to lose energy while traversing the QGP, has been extensively studied to understand the interaction between jets and the medium [2–5]. For instance, studies of jet suppression [6–15], dijet asymmetry [16–19], and splitting functions [20–22], in addition to modifications in jet shapes [23–25], fragmentation functions [26,27], and energy-energy correlators [28,29], have provided insights into the redistribution of energy and momentum due to QGP interactions. Despite these advances, challenges remain in studying jet quenching due to the steeply falling jet spectrum and, even more so, the presence of a large fluctuating background characteristic of heavy-ion environments.

Identifying jet constituents against the background is thus a significant challenge, motivating considerable efforts to develop efficient background-subtraction methods. For instance, conventional approaches such as the area [30,31] and constituent-based subtraction techniques [32,33] estimate and subtract an average (or modulated, i.e., a function of azimuthal angle) background contribution after excluding the leading jets. However, by doing so, they often erase fine-grained jet substructure and consequently produce substantial residuals. Since one cannot identify a single particle as having been produced from either jet or a background, these methodologies often rely on ensemble-based approaches that trade background rejection and signal retention. This is particularly important for low-momentum particles in and around jets, which are expected to carry signatures of the jet-QGP interaction. These subtraction methods, originally developed for proton-proton (pp) collisions, struggle in heavy-ion environments due to much larger backgrounds and nonlinear interactions between jets and the QGP. For example, background-subtraction methods can be skewed by medium-induced recoil and other noise, distorting both jet momentum and internal structure, an effect that is particularly pronounced for low- p_T jets. This highlights the need for more sophisticated techniques capable of navigating the complex background while faithfully preserving the detailed substructure of heavy-ion jets.

The advent of machine learning (ML) has opened new avenues for analyzing heavy-ion collisions [36]. Recent efforts have leveraged ML on various representations of heavy-ion jets (i.e., sequences, images, graphs, point clouds, etc.) for quark-gluon discrimination [37,38], tagging quenched jets [39–44], predicting jet energy loss [45], and locating jet creation vertices [46]. However, applications in background-subtraction remain limited to reconstructing the transverse momenta [47–52] of jets embedded within a heavy-ion

*Contact author: uqureshi@cern.ch

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/) license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

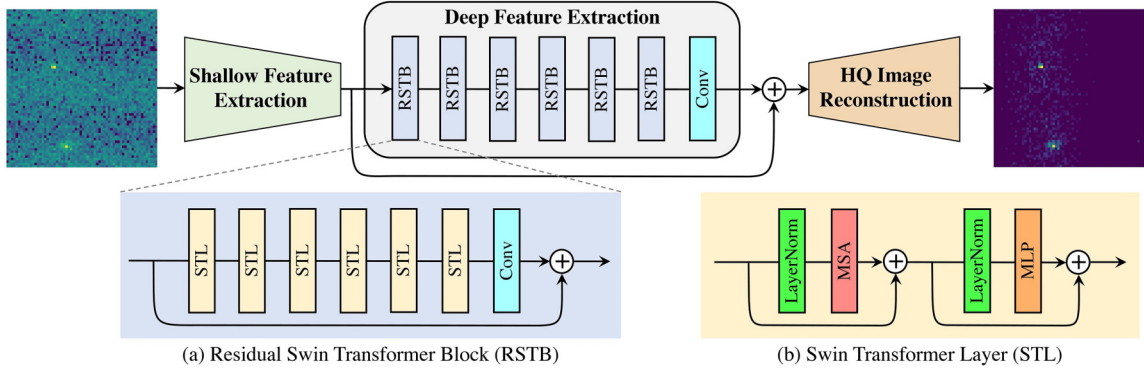


FIG. 1. The DEEPSUB architecture based on the Swin Transformer [34,35] for image reconstruction.

background. The aforementioned momentum reconstruction methods are inherently restrictive: they focus on a single one-dimensional observable and thus do not capture the full, high-dimensional structure of jets. In addition, they are often trained on generators that simulate pp collisions and only consider a small dynamic range in the space of jets (i.e., restricted jet transverse momenta). As such, these methods fail to adapt to the wide range of medium-induced modifications as shown in Refs. [50,53].

A general-purpose ML-based framework is therefore needed, one capable of learning from the high-dimensional structure jets on an event-by-event basis, encoding nontrivial correlations between signal and background constituents and robustly disentangling the entirety of the signal instead of extracting a one-dimensional quantity. In this study, we present DEEPSUB, an ML-based approach for full-event background-subtraction. DEEPSUB takes as input noisy event images, where hard events are embedded in the underlying events, and outputs the signal image as shown in Fig. 1.

The remainder of this Letter is structured as follows: In Sec. II A, we describe the simulation procedure for generating hard scattering and the thermal background events. Section II B discusses the Swin-Transformer-based DEEPSUB architecture. In Sec. III, we present results on traditionally challenging jet observables and compare them between different background-subtraction methods. Finally, we conclude with a short discussion in Sec. IV.

II. METHODOLOGY

A. Samples and simulation

Dijet events are produced considering hard 2-to-2 QCD scatterings with PbPb beams at $\sqrt{s_{NN}} = 5.02$ TeV with $\hat{p}_T > 100$ GeV at 0%–10% centrality using JEWEL [54–56]. No further cutoffs are imposed to ensure that the model is trained on a large dynamic range of jets. To simulate interactions with the QGP medium, we keep recoils on and consider initial temperature $T_i = 590$ MeV and initial quenching time $\tau_i = 0.4$, which provides an adequate description of a variety of jet quenching observables. To remove the energy-momentum contributions from the recoil partons prior to interactions with the hard-scattered partons, the constituent subtraction method [57] is applied.

We simulate background events corresponding to the most-central (0%–10%) PbPb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, in accordance with ALICE data published in Refs. [58–60]. To account for fluctuations in particle multiplicity, we sample the number of particles from a Gaussian distribution with mean 15 000 and standard deviation 122 (the square root of the mean). The p_T is sampled from a Boltzmann distribution with mean $\langle p_T \rangle$ of 1.2 GeV. The pseudorapidity η is sampled uniformly over the interval $|\eta| < 3$ and the azimuth ϕ from the distribution $f(\phi) = \frac{1}{2\pi} [1 - v_2 \cos(2\phi)]$ with $v_2 = 0.05$ to account for elliptic flow modulation on ϕ . Mixed events are subsequently constructed by embedding dijet events into the underlying events.

Event images are then created by discretizing the η - ϕ plane into a 128×128 grid covering $|\eta| < 3$ and $0 \leq \phi < \pi$. Each final-state particle with coordinates (η_i, ϕ_i) and transverse momentum $p_{T,i}$ deposits its $p_{T,i}$ into the corresponding pixel, yielding a two-dimensional image of total transverse momentum per bin, similar to Refs. [61–67]. For each event, we generate a pair of images: the mixed “input” image $\mathcal{I}_{in}$, which includes both the underlying event and the dijet, and the “target” image $\mathcal{I}_{target}$, which contains only the particles from the JEWEL dijet.

B. Machine-learning architecture

DEEPSUB utilizes the Swin Transformer [34,35] architecture for event image reconstruction as shown in Fig. 1. The model processes the input as nonoverlapping local windows and applies shifted-window multihead self-attention, which enables both efficient computation and information exchange across window boundaries. The network begins with a shallow feature extractor h_{SF} (a 3×3 convolution) that preserves local, low-level features. These features are passed to a deep feature extractor composed of stacked residual Swin Transformer blocks (RSTBs), h_{RSTB} , where each RSTB bundles several Swin Transformer layers inside a residual wrapper to stabilize training and capture long- and mid-range dependencies. Convolutions inside and around RSTBs provide an inductive bias toward locality, complementing attention’s global context. A simple reconstruction head (a final 3×3 convolution in our setup) maps deep features back to the image space, yielding an output with the same spatial resolution as the input.

Given an input image $\mathcal{I}_{\text{in}} \in \mathbb{R}^{H \times W}$, a shallow feature extraction layer h_{SF} consisting of a 3×3 convolutional layer extracts shallow features $F_0 \in \mathbb{R}^{H \times W \times C}$ as $F_0 = h_{\text{SF}}(\mathcal{I}_{\text{in}})$, where C is the latent feature dimension. Deep feature representations $F_{\text{DF}} \in \mathbb{R}^{H \times W \times C}$ are then extracted from F_0 as $F_{\text{DF}} = h_{\text{DF}}(F_0)$ with h_{DF} denoting the deep feature extraction module, which consists of K RSTBs followed by a 3×3 convolution. In particular, intermediate features $F_1, F_2, \dots, F_K$ and F_{DF} are constructed sequentially as $F_i = h_{\text{RSTB}}^i(F_{i-1})$ for $i \in \{1, 2, \dots, K\}$ and $F_{\text{DF}} = h_{\text{conv}}(F_K)$ where h_{RSTB}^i and h_{conv} denote the i th RSTB and final convolution layer, respectively.

As the name suggests, the RSTB is a residual block with Swin Transformer layers (STLs) followed by a convolution. Given the input feature F_0^i of the i th RSTB, the features $F_1^i, F_2^i, \dots, F_L^i$ by L STLs are computed as $F_j^i = h_{\text{STL}_j}^i(F_{j-1}^i)$ for $j \in \{1, 2, \dots, L\}$, where $h_{\text{STL}_j}^i$ is the j th STL in the i th RSTB. Finally, the output is generated by applying a convolution before the residual connection: $F_{\text{out}}^i = h_{\text{conv}_i}(F_L^i) + F_0^i$, where h_{conv_i} is the final convolutional layer in the i th RSTB. Finally, the shallow and deep features are convolved with a single convolutional layer to produce the reconstructed output image $\mathcal{I}_{\text{reco}}$ as $\mathcal{I}_{\text{reco}} = h_{\text{conv}_{\text{IR}}}(F_0 + F_{\text{DF}})$, where $h_{\text{conv}_{\text{IR}}}$ is the convolution layer in the High-Quality (HQ) image reconstruction module.

The STL extends the original multihead self-attention (MSA) mechanism of the standard transformer [68] by restricting attention to local windows and alternating these windows in a shifted configuration to enable cross-window interactions. Given an image $\mathcal{I} \in \mathbb{R}^{H \times W \times C}$, it is first partitioned into HW/M^2 nonoverlapping windows of size $M \times M$, each reshaped into a feature matrix $X \in \mathbb{R}^{M^2 \times C}$. Within each window, projections $P_Q, P_K, P_V \in \mathbb{R}^{C \times d}$ produce queries $Q = XP_Q$, keys $K = XP_K$, and values $V = XP_V$ in $\mathbb{R}^{M^2 \times d}$. Local self-attention is subsequently computed as $\text{SoftMax}(\frac{QK^T}{\sqrt{d}} + B)V$ where B denotes learnable relative positional biases. Multihead self-attention concatenates h such attention outputs computed in parallel. A subsequent two-layer MLP with GELU activation is further applied to the features. Both the attention and MLP blocks employ pre-LayerNorm and residual connections, thus giving $\tilde{X} = \text{MSA}(\text{LN}(X)) + X$ and then $X' = \text{MLP}(\text{LN}(\tilde{X})) + \tilde{X}$. In contrast with the vision transformer (ViT) [69], which applies global self-attention over all tokens with $O(H^2W^2)$ complexity, the Swin Transformer's hierarchical, window-based design achieves linear $O(HWM^2)$ complexity while still capturing long-range dependencies via its shifted window scheme as well as bringing in strong inductive biases for local translational equivariance.

III. RESULTS

For this study, the number of RSTBs, STLs, and attention heads is kept to six each. The window size and latent feature dimension are kept to 8 and 180, respectively. The model in this formulation has 11.5 M parameters and uses roughly 5 GFLOPs on the forward pass. The canonical PyTorch [70] deep learning framework was employed for implementing, configuring, training, and evaluating the model on $4 \times$ Nvidia RTX A100 GPUs. The ADAM [71] optimizer with a learning

rate of 7×10^{-5} is used in conjunction with the mean-squared error (MSE) loss function between $\mathcal{I}_{\text{reco}}$ and $\mathcal{I}_{\text{target}}$. A train-validation-test split consisting of 360, 40, and 80 k events, respectively, is used with a batch size of 128, and the model is trained for 15 epochs until convergence. The best model based on the validation set MSE loss is used for all subsequent testing.

The output space of events generated by DEEPSUB is too high-dimensional to be examined holistically. Therefore, to evaluate its performance, we compute a few observables and compare their truth values to the reconstructed values. We compare DEEPSUB to the powerful and widely used event-wide iterative constituent subtraction (ICS) method [33] with two iterations using distance parameters $\Delta R_{\text{max}} = \{0.2, 0.1\}$ between signal and background particles, and a transverse momentum weight $\alpha = 1$. For the ICS baseline, we adopt the same η - ϕ discretization prior to subtraction, in line with prior image-based strategies [66]. Jets are clustered after background-subtraction with the anti- k_T algorithm [72] requiring $R = 0.4$, $p_T^{\text{cons}} > 0.3$ GeV, $p_T^{\text{jet}} > 100$ GeV, and $|\eta| < 2.5$ using FASTJET [30,73,74]. All particles are given the charged-pion $\pi^\pm$ mass of 139.5 MeV. Our comparisons focus on the following four jet observables: (i) jet p_T , i.e., the transverse momentum of the jet; (ii) jet mass, i.e., the Lorentz-invariant mass of the four-momentum sum of all particles (grid cells from the image); (iii) the jet girth g defined as

$$g = \frac{1}{p_T^{\text{jet}}} \sum_{i \in \text{jet}} p_T^i \Delta R(i, \text{jet}), \quad (1)$$

where p_T^i is the transverse momentum of constituent i and $\Delta R(i, \text{jet})$ is the distance to the jet axis; and (iv) the energy correlation function (ECF) [75] with $N = 2$ and $\beta = 2$ defined as

$$\text{ECF}_{N=2}^{\beta=2} = \sum_{i \in \text{jet}} \sum_{j < i} p_T^i p_T^j \Delta R(i, j)^2. \quad (2)$$

The top panels of Fig. 2 illustrate the distributions of these observables after applying the two background-subtraction methods. A more useful comparison of performance is shown in the bottom panels of each plot, i.e., the ratio of the reconstructed value of each observable to their truth values. From a quick glance at the plots, it is clear that DEEPSUB more closely matches, by a significant margin, the true distributions of more sensitive substructure observables like mass, girth, and the energy correlation functions as compared with the event-wide ICS method.

The top-left plot of Fig. 2 shows jet p_T , which falls by nearly four orders of magnitude between 100 and 1200 GeV in a characteristic power law. Overlaid are the background-subtracted spectra: DEEPSUB (red-dashed) and the event-wide ICS method (green-dashed). Both closely follow the true distribution, but with small deviations. In the 100–200 GeV region, ICS overshoots the truth by $\approx 5\%$, while DEEPSUB stays essentially on top of it. From 200 to 800 GeV, ICS continues to exceed the truth by up to $\approx 10\%$, whereas DEEPSUB remains within 1% of unity. Above ≈ 1 TeV, both methods fluctuate about the true yield, ICS by as much as $\approx 20\%$ and DEEPSUB by $\approx 5\%$, although much of these can be attributed

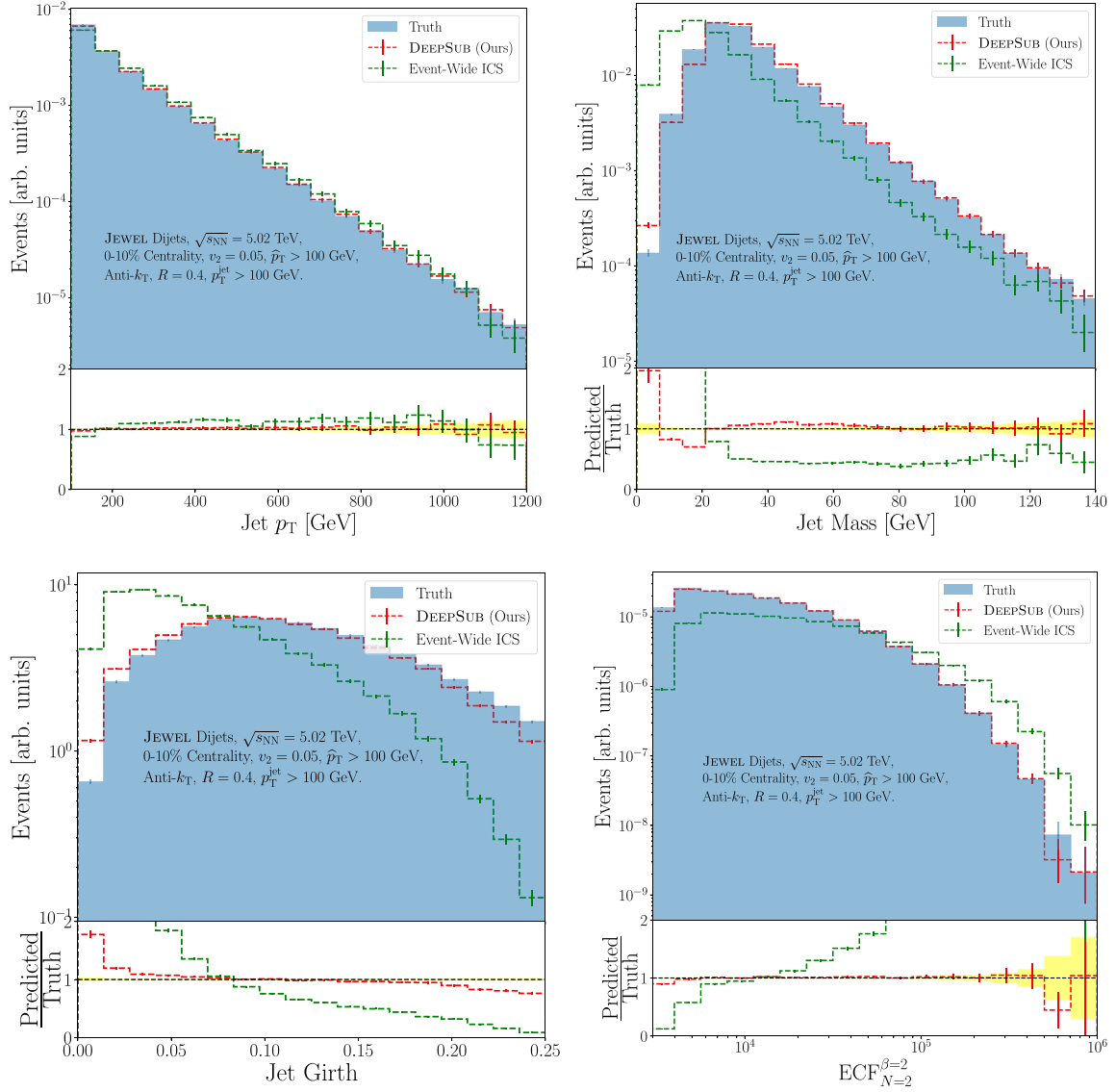


FIG. 2. Distributions of jet p_T (top-left), mass (top-right), girth (bottom-left), and the energy correlation function (bottom-right) for DEEPSUB (red-dotted), event-wide ICS (green-dashed), and the ground truth (blue-filled) on JEWEL dijet events. The error bars represent the statistical uncertainty in each bin. The bottom panels in each plot show the ratio of the reconstructed value of each observable to their truth value, and the yellow-shaded region indicates the statistical uncertainty in the truth.

to statistical uncertainties. Overall, DEEPSUB delivers a more faithful reconstruction of the hard-jet spectrum with smaller residual deviations compared with the traditional event-wide ICS approach.

The top-right panel of Fig. 2 shows the reconstructed jet mass spectrum, which falls by roughly three orders of magnitude. Both methods struggle at the very lowest mass bins $\approx 3\text{--}20$ GeV, here ICS overshoots the truth by as much as an order of magnitude, while DEEPSUB overpredicts by $\approx 40\%$. In the intermediate mass region (20–60 GeV), ICS underestimates the yield by a factor of $\approx 2\text{--}3$, whereas DEEPSUB remains within 1%–5% of the true spectrum. At high masses (>80 GeV), ICS still runs low by a factor of 2 and DEEPSUB by $\approx 1\%$ –5% except in the very last bins where statistical uncertainties dominate. Since jet mass is largely governed by large-angle structure, this behavior is indicative

of over-subtraction and destruction of genuine large-angle soft radiation within the jet by the ICS method, a notion that is supported by Refs. [32,33]. In contrast, DEEPSUB is able to preserve this softer wide-angle information and delivers a markedly more accurate reconstruction of the jet mass spectrum than ICS.

The bottom-left panel shows the reconstructed jet girth distribution. At the smallest girth values $\lesssim 0.075$, ICS overshoots the truth by roughly 80%–100%, while DEEPSUB remains much closer, overpredicting by only 1%–5%. In the mid-girth region ($0.08 \lesssim g \lesssim 0.12$), ICS transitions from overpredicting to underpredicting, while DEEPSUB remains within $\approx 1\%$ –5% of the truth. Beyond $g \approx 0.12$, ICS increasingly underestimates the yield, falling to $\approx 50\%$ of the truth by $g \approx 0.018$ and down to 10% by $g \approx 0.25$, whereas DEEPSUB holds within $\approx 5\%$ –10% across the entire high-girth tail.

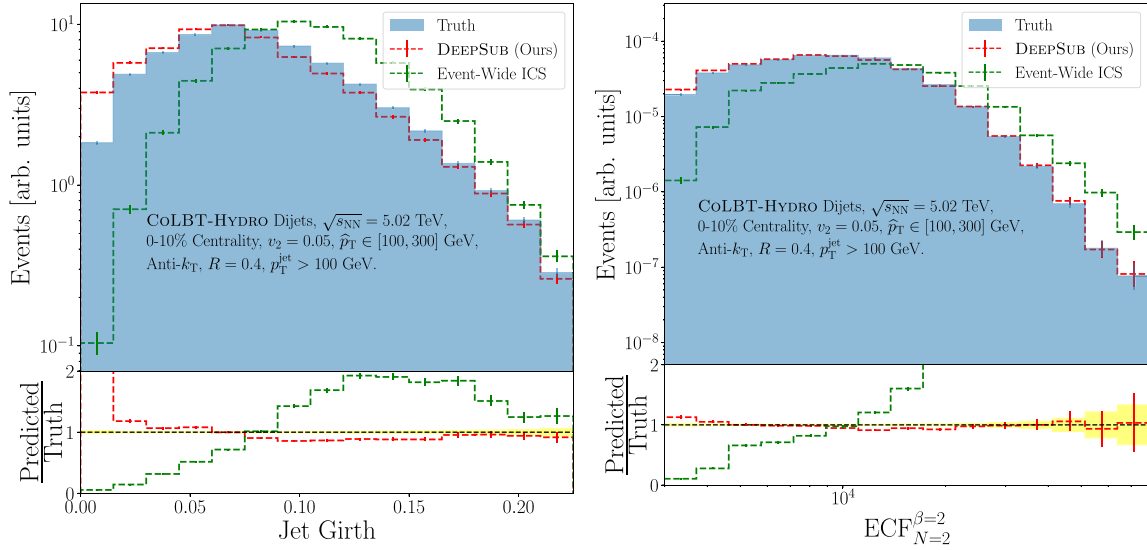


FIG. 3. Distributions of jet girth (left), and the energy correlation function (right) for DEEPSUB (red-dotted), event-wide ICS (green-dashed), and the ground truth (blue-filled) on COLBT-HYDRO dijet events. The error bars represent the statistical uncertainty in each bin. The bottom panels in each plot show the ratio of the reconstructed value of each observable to their truth value, and the yellow-shaded region indicates the statistical uncertainty in the truth.

Finally, the bottom-right panels show the ECF distribution, which falls by more than six orders of magnitude. At the smallest ECF values $\approx 10^4$, ICS severely underpredicts the yield (ratio ≈ 0.15) while DEEPSUB remains within $\approx 5\%$ of the truth. In the midrange ($\text{ECF} \in [2 \times 10^4, 10^5]$), ICS overshoots by up to 80%, whereas DEEPSUB tracks the truth within $\approx 1\%$ – 5% . At higher ECF (10^5 – 5×10^5), ICS again overestimates by as much as an order of magnitude, while DEEPSUB continues to stay within $\approx 1\%$ – 5% . In the extreme tail ($\text{ECF} > 5 \times 10^5$), both methods exhibit statistical fluctuations and DEEPSUB shows a large dip (to 20%), indicative of limited counts. Once again, DEEPSUB reproduces the ECF with significantly higher accuracy than ICS.

While it is encouraging that DEEPSUB performs so well in Fig. 2 relative to ICS, it is perhaps not surprising given that it was fine-tuned on the (statistically) same dataset, while ICS was generically designed to subtract a variety of underlying events from a signal. Furthermore, it is also conceivable that the model may have over-fitted to model-dependent radiation patterns, which is especially problematic since, in this case, the performance of DEEPSUB trained on simulation may not translate to real data due to mismodeling. Thus, as a second set of results, we investigate the performance of DEEPSUB on event distributions from a generator that it did not encounter during training. For this out-of-distribution validation, we use events from COLBT-HYDRO [76] simulation with $\hat{p}_T \in [100, 300]$ GeV and test them identically to the JEWEL events. For brevity, we omit distributions of jet p_T and mass, and present the more challenging girth and ECF distributions in Fig. 3.

The performance of DEEPSUB for these distributions continues to inspire confidence, remaining within 1%–5% of the truth, with the exception of tails, despite significant shifts in the spectra themselves. Meanwhile, ICS continues to show large order-of-magnitude deviations from the truth. Although

DEEPSUB was never trained on COLBT-HYDRO, it still outperforms the generic ICS approach by a wide margin, thus demonstrating its robustness to generator shifts, although further systematic studies may be needed to ensure reliable transfer to real data. The main reason one can expect DEEPSUB to extrapolate to this dataset is that while the models use very different mechanisms to implement jet energy loss and other quenching effects, large overlaps exist in radiation patterns and other particle correlations. Therefore, the same latent representations DEEPSUB learns to distinguish signal from background remain effective.

Another key advantage of DEEPSUB is speed. For 80 k JEWEL events parallelized across a 32-core Intel Xeon (2500 per core), it takes approximately 20 minutes for ICS to run. In comparison, DEEPSUB running on a fairly modest Nvidia RTX A6000 GPU takes only 3 minutes.

IV. DISCUSSION AND FUTURE WORK

In summary, we have presented a paradigm for constructing a data-driven background-subtraction method in heavy-ion environments. This framework is based on a deep learning model that takes as input an event image (a noisy image consisting of signal and underlying event) and outputs another image (the reconstructed signal particles). While we have used dijet events at LHC energies as an example, this approach is generalizable to a variety of processes at past, current, and future accelerators. It is also worth reminding that while we chose a handful of observables for their sensitivity to jet substructure, the goal is that whatever observables are used for an analysis, DEEPSUB should be able to provide an accurate reconstruction.

The long-term objective of DEEPSUB is to solve the challenges encountered by experimental collaborations in mitigating the heavy-ion background and to provide a unified

analysis framework for both experimentalists and phenomenologists. Nonclosure due to background-subtraction is often an $O(1)$ fraction of all systematic uncertainty in a heavy-ion jet substructure analysis. Thus, if we can reduce it to the extent shown in this study, then we could pave the way for precision QCD measurements in heavy-ion collisions. Furthermore, since DEEPSUB, once trained, is portable, it would be a straightforward one-time exercise for experiments to share their tuned models for background-subtraction for anyone to use.

Although this work represents a major advancement toward this long-term vision, several critical steps remain. First, we must evaluate DEEPSUB's performance on other processes such as γ -jet, Z-jet, or minimum-bias events, and also include additional output features such as particle ID information. Next, establishing true universality will require validating the method across a variety of energies, background parameters, and observables. Moreover, for widespread adoption, DEEPSUB should also provide a flexible, user-friendly interface, on par with that of FASTJET, for instance. Lastly, we need to work with experiments to develop models and facilitate the sharing of these models with the broader community.

With that said, we believe that the progress presented in this Letter serves as proof of concept for DEEPSUB in establishing a new paradigm for fast, accurate background-subtraction in ultrarelativistic heavy-ion environments.

ACKNOWLEDGMENTS

The authors express their gratitude to Yi (Luna) Chen (Vanderbilt) for insightful discussions and Michael Kagan (SLAC) for useful feedback on the paper. The authors also appreciate the support of the Vanderbilt ACCRE computing facility and the Data Science Institute at Vanderbilt University for providing computing resources, including access to the DGX A100 AI Computing Server. R.K.E. acknowledges funding from the U.S. Department of Energy, Office of Science, Office of Nuclear Physics, Grant No. DE-SC0024660.

DATA AVAILABILITY

The data that support the findings of this article are openly available [77,78].

-
- [1] Y. Mehtar-Tani, J. G. Milhano, and K. Tywoniuk, Jet physics in heavy-ion collisions, *Int. J. Mod. Phys. A* **28**, 1340013 (2013).
 - [2] J. D. Bjorken, *Energy Loss of Energetic Partons in Quark-Gluon Plasma: Possible Extinction of High p_T Jets in Hadron-Hadron Collisions*, Technical Report (FERMILAB, Batavia, 1982).
 - [3] D. d'Enterria, Jet quenching, *Landolt-Bornstein* **23**, 471 (2010).
 - [4] A. Majumder and M. Van Leeuwen, The theory and phenomenology of perturbative QCD based jet quenching, *Prog. Part. Nucl. Phys.* **66**, 41 (2011).
 - [5] G.-Y. Qin and X.-N. Wang, Jet quenching in high-energy heavy-ion collisions, *Int. J. Mod. Phys. E* **24**, 1530014 (2015).
 - [6] K. Adcox *et al.* (PHENIX Collaboration), Suppression of hadrons with large transverse momentum in central Au + Au collisions at $\sqrt{s_{NN}} = 130$ -GeV, *Phys. Rev. Lett.* **88**, 022301 (2001).
 - [7] C. Adler *et al.* (STAR Collaboration), Centrality dependence of high p_T hadron suppression in Au + Au collisions at $\sqrt{s_{NN}} = 130$ -GeV, *Phys. Rev. Lett.* **89**, 202301 (2002).
 - [8] K. Aamodt *et al.* (ALICE Collaboration), Suppression of charged particle production at large transverse momentum in central Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, *Phys. Lett. B* **696**, 30 (2011).
 - [9] S. S. Adler *et al.* (PHENIX Collaboration), Suppressed π^0 production at large transverse momentum in central Au + Au collisions at $\sqrt{s_{NN}} = 200$ GeV, *Phys. Rev. Lett.* **91**, 072301 (2003).
 - [10] J. Adams *et al.* (STAR Collaboration), Transverse momentum and collision energy dependence of high p_T hadron suppression in Au + Au collisions at ultrarelativistic energies, *Phys. Rev. Lett.* **91**, 172302 (2003).
 - [11] C. Adler *et al.* (STAR Collaboration), Disappearance of back-to-back high p_T hadron correlations in central Au + Au collisions at $\sqrt{s_{NN}} = 200$ -GeV, *Phys. Rev. Lett.* **90**, 082302 (2003).
 - [12] K. Aamodt *et al.* (ALICE Collaboration), Particle-yield modification in jet-like azimuthal di-hadron correlations in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, *Phys. Rev. Lett.* **108**, 092301 (2012).
 - [13] S. Chatrchyan *et al.* (CMS Collaboration), Study of high- p_T charged particle suppression in PbPb compared to pp collisions at $\sqrt{s_{NN}} = 2.76$ TeV, *Eur. Phys. J. C* **72**, 1945 (2012).
 - [14] G. Aad *et al.* (ATLAS Collaboration), Measurement of charged-particle spectra in Pb + Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS detector at the LHC, *J. High Energy Phys.* **09** (2015) 050.
 - [15] V. Khachatryan *et al.* (CMS Collaboration), Charged-particle nuclear modification factors in PbPb and p Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, *J. High Energy Phys.* **04** (2017) 039.
 - [16] G. Aad *et al.* (ATLAS Collaboration), Observation of a centrality-dependent dijet asymmetry in lead-lead collisions at $\sqrt{s_{NN}} = 2.77$ TeV with the ATLAS detector at the LHC, *Phys. Rev. Lett.* **105**, 252303 (2010).
 - [17] S. Chatrchyan *et al.* (CMS Collaboration), Observation and studies of jet quenching in PbPb collisions at nucleon-nucleon center-of-mass energy = 2.76 TeV, *Phys. Rev. C* **84**, 024906 (2011).
 - [18] S. Chatrchyan *et al.* (CMS Collaboration), Jet momentum dependence of jet quenching in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, *Phys. Lett. B* **712**, 176 (2012).
 - [19] V. Khachatryan *et al.* (CMS Collaboration), Measurement of transverse momentum relative to dijet systems in PbPb and pp collisions at $\sqrt{s_{NN}} = 2.76$ TeV, *J. High Energy Phys.* **01** (2016) 006.
 - [20] A. M. Sirunyan *et al.* (CMS Collaboration), Measurement of the splitting function in pp and Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, *Phys. Rev. Lett.* **120**, 142302 (2018).
 - [21] M. S. Abdallah *et al.* (STAR Collaboration), Differential measurements of jet substructure and partonic energy loss in

- Au + Au collisions at $\sqrt{s_{NN}} = 200$ GeV, *Phys. Rev. C* **105**, 044906 (2022).
- [22] S. Acharya *et al.* (A Large Ion Collider Experiment, ALICE Collaboration), Measurement of the groomed jet radius and momentum splitting fraction in pp and Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, *Phys. Rev. Lett.* **128**, 102001 (2022).
- [23] S. Chatrchyan *et al.* (CMS Collaboration), Modification of jet shapes in PbPb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, *Phys. Lett. B* **730**, 243 (2014).
- [24] S. Acharya *et al.* (ALICE Collaboration), Measurement of jet radial profiles in Pb – Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, *Phys. Lett. B* **796**, 204 (2019).
- [25] G. Aad *et al.* (ATLAS Collaboration), Measurement of angular and momentum distributions of charged particles within and around jets in Pb + Pb and pp collisions at $\sqrt{s_{NN}} = 5.02$ TeV with the ATLAS detector, *Phys. Rev. C* **100**, 064901 (2019); **101**, 059903(E) (2020).
- [26] S. Chatrchyan *et al.* (CMS Collaboration), Measurement of jet fragmentation in PbPb and pp collisions at $\sqrt{s_{NN}} = 2.76$ TeV, *Phys. Rev. C* **90**, 024908 (2014).
- [27] M. Aaboud *et al.* (ATLAS Collaboration), Measurement of jet fragmentation in Pb + Pb and pp collisions at $\sqrt{s_{NN}} = 2.76$ TeV with the ATLAS detector at the LHC, *Eur. Phys. J. C* **77**, 379 (2017).
- [28] S.-Y. Chen, K.-M. Shen, X.-F. Xue, W. Dai, B.-W. Zhang, and E.-K. Wang, Study of EEC discrimination power on quark and gluon quenching effects in heavy-ion collisions at $\sqrt{s} = 5.02$ TeV, *arXiv:2409.13996*.
- [29] V. Chekhovsky *et al.* (CMS Collaboration), Observation of nuclear modification of energy-energy correlators inside jets in heavy ion collisions, *Phys. Lett. B* **866**, 139556 (2025).
- [30] M. Cacciari, G. P. Salam, and G. Soyez, Fastjet user manual, *Eur. Phys. J. C* **72**, 1896 (2012).
- [31] B. Abelev *et al.* (ALICE Collaboration), Measurement of charged jet suppression in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, *J. High Energy Phys.* **03** (2014) 013.
- [32] P. Berta, L. Masetti, D. Miller, and M. Spousta, Pileup and underlying event mitigation with iterative constituent subtraction, *J. High Energy Phys.* **08** (2019) 175.
- [33] P. Berta, M. Spousta, D. W. Miller, and R. Leitner, Particle-level pileup subtraction for jets and jet shapes, *J. High Energy Phys.* **06** (2014) 092.
- [34] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, Swin transformer: Hierarchical vision transformer using shifted windows, *arXiv:2103.14030*.
- [35] J. Liang, J. Cao, G. Sun, K. Zhang, L. V. Gool, and R. Timofte, Swinir: Image restoration using swin transformer, *arXiv:2108.10257*.
- [36] HEP ML Community, A living review of machine learning for particle physics.
- [37] Y.-T. Chien and R. K. Elayavalli, Probing heavy ion collisions using quark and gluon jet substructure, *arXiv:1803.03589*.
- [38] Y. Du, D. Pablos, and K. Tywoniuk, Classification of quark and gluon jets in hot QCD medium with deep learning, in *Proceedings of Particles and Nuclei International Conference 2021 — PoS(PANIC2021)*, PANIC2021 (Sissa Medialab, Lisbon, Portugal, 2022), p. 224.
- [39] L. Liu, J. Velkovska, Y. Wu, and M. Verweij, Identifying quenched jets in heavy ion collisions with machine learning, *J. High Energy Phys.* **04** (2023) 140.
- [40] L. Apolinário, N. F. Castro, M. Crispim Romão, J. G. Milhano, R. Pedro, and F. C. R. Peres, Deep learning for the classification of quenched jets, *J. High Energy Phys.* **11** (2021) 219.
- [41] Y. S. Lai, J. Mulligan, M. Płoskoń, and F. Ringer, The information content of jet quenching and machine learning assisted observable design, *J. High Energy Phys.* **10** (2022) 011.
- [42] M. Crispim Romão, J. G. Milhano, and M. van Leeuwen, Jet substructure observables for jet quenching in quark gluon plasma: A machine learning driven analysis, *SciPost Phys.* **16**, 015 (2024).
- [43] Y.-L. Du, Overview: Jet quenching with machine learning, *arXiv:2308.10035*.
- [44] U. S. Qureshi and R. Kunawalkam Elayavalli, Model-agnostic tagging of quenched jets in heavy-ion collisions, *arXiv:2411.19389*.
- [45] Y.-L. Du, D. Pablos, and K. Tywoniuk, Deep learning jet modifications in heavy-ion collisions, *J. High Energy Phys.* **03** (2021) 206.
- [46] Z. Yang, Y. He, W. Chen, W.-Y. Ke, L.-G. Pang, and X.-N. Wang, Deep learning assisted jet tomography for the study of mach cones in QGP, *Eur. Phys. J. C* **83**, (2023).
- [47] R. Haake and C. Loizides, Machine-learning-based jet momentum reconstruction in heavy-ion collisions, *Phys. Rev. C* **99**, 064904 (2019).
- [48] H. Bossi, r -dependence of inclusive jet suppression and groomed jet splittings in heavy-ion collisions with ALICE, *arXiv:2208.14492*.
- [49] T. Mengel, P. Steffanic, C. Hughes, A. C. O. Da Silva, and C. Nattrass, Multiplicity-based background subtraction for jets in heavy ion collisions, *arXiv:2402.10945*.
- [50] R. Li, Y.-L. Du, and S. Cao, Jet momentum reconstruction in the QGP background with machine learning, *arXiv:2412.06466*.
- [51] T. Mengel, P. Steffanic, C. Hughes, A. C. O. D. Silva, and C. Nattrass, Multiplicity-based background subtraction for jets in heavy ion collisions, *arXiv:2402.10945*.
- [52] Y.-L. Du, Overview: Jet quenching with machine learning, *arXiv:2308.10035*.
- [53] D. Stewart and J. Putschke, Neural network biased corrections: Cautionary study in background corrections for quenched jets, *Phys. Rev. C* **111**, 054902 (2025).
- [54] K. C. Zapp, F. Krauss, and U. A. Wiedemann, A perturbative framework for jet quenching, *J. High Energy Phys.* **03** (2013) 080.
- [55] R. K. Elayavalli and K. C. Zapp, Simulating v +jet processes in heavy ion collisions with jewel, *Eur. Phys. J. C* **76**, 695 (2016).
- [56] R. Kunawalkam Elayavalli and K. C. Zapp, Medium response in jewel and its impact on jet shape observables in heavy ion collisions, *J. High Energy Phys.* **07** (2017) 141.
- [57] J. G. Milhano and K. Zapp, Improved background subtraction and a fresh look at jet sub-structure in JEWEL, *Eur. Phys. J. C* **82**, 1010 (2022).
- [58] J. Adam *et al.* (ALICE Collaboration), Centrality dependence of the pseudorapidity density distribution for charged particles in Pb-Pb collisions at $\sqrt{s_{NN}} = 5.02$ TeV, *Phys. Lett. B* **772**, 567 (2017).
- [59] B. Abelev *et al.* (ALICE Collaboration), Measurement of event background fluctuations for charged particle jet reconstruction in Pb-Pb collisions at $\sqrt{s_{NN}} = 2.76$ TeV, *J. High Energy Phys.* **03** (2012) 053.
- [60] S. Acharya *et al.* (ALICE Collaboration), Transverse momentum spectra and nuclear modification factors of charged

- particles in pp , p -Pb and Pb-Pb collisions at the LHC, *J. High Energy Phys.* **11** (2018) 013.
- [61] J. Cogan, M. Kagan, E. Strauss, and A. Schwartzman, Jet-images: Computer vision inspired techniques for jet tagging, *J. High Energy Phys.* **02** (2015) 118.
- [62] L. de Oliveira, M. Kagan, L. Mackey, B. Nachman, and A. Schwartzman, Jet-images — deep learning edition, *J. High Energy Phys.* **07** (2016) 069.
- [63] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, Deep learning in color: Towards automated quark/gluon jet discrimination, *J. High Energy Phys.* **01** (2017) 110.
- [64] J. Lin, M. Freytsis, I. Moulton, and B. Nachman, Boosting $h \rightarrow b\bar{b}$ with machine learning, *J. High Energy Phys.* **10** (2018) 101.
- [65] S. Macaluso and D. Shih, Pulling out all the tops with computer vision and deep learning, *J. High Energy Phys.* **10** (2018) 121.
- [66] P. T. Komiske, E. M. Metodiev, B. Nachman, and M. D. Schwartz, Pileup mitigation with machine learning (pumml), *J. High Energy Phys.* **12** (2017) 051.
- [67] M. Kagan, Image-based jet analysis, [arXiv:2012.09719](https://arxiv.org/abs/2012.09719).
- [68] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, Attention is all you need, [arXiv:1706.03762](https://arxiv.org/abs/1706.03762).
- [69] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, An image is worth 16×16 words: Transformers for image recognition at scale, [arXiv:2010.11929](https://arxiv.org/abs/2010.11929).
- [70] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Köpf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai *et al.*, Pytorch: An imperative style, high-performance deep learning library, [arXiv:1912.01703](https://arxiv.org/abs/1912.01703).
- [71] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [72] M. Cacciari, G. P. Salam, and G. Soyez, The anti- k_t jet clustering algorithm, *J. High Energy Phys.* **04** (2008) 063.
- [73] M. Cacciari and G. P. Salam, Dispelling the N^3 myth for the k_t jet-finder, *Phys. Lett. B* **641**, 57 (2006).
- [74] A. Roy, J. Pivarski, and C. W. Freer, An array-oriented Python interface for FastJet, *J. Phys.: Conf. Ser.* **2438**, 012011 (2023).
- [75] A. J. Larkoski, G. P. Salam, and J. Thaler, Energy correlation functions for jet substructure, *J. High Energy Phys.* **06** (2013) 108.
- [76] W. Chen, S. Cao, T. Luo, L.-G. Pang, and X.-N. Wang, Effects of jet-induced medium excitation in γ -hadron correlation in $A + A$ collisions, *Phys. Lett. B* **777**, 86 (2018).
- [77] U. S. Qureshi and R. Kunnawalkam Elayavalli, Dataset from deepsub: Deep image reconstruction for background subtraction in heavy-ion collisions, [arXiv:2507.14036](https://arxiv.org/abs/2507.14036).
- [78] U. S. Qureshi, Deepsub: Deep image reconstruction for background subtraction in heavy-ion collisions, <https://github.com/umarsqureshi/DeepSub> (2025).