Reweighting Adversarial Networks for Unbinned Unfolding

Umar Sohail Qureshi[1,2] **Krish Desai**[3,4] **Benjamin Nachman**[2,5] **Jesse Thaler**[6,7]

[1]Department of Physics, Stanford University
[2]Fundamental Physics Directorate, SLAC National Accelerator Laboratory
[3]Department of Physics, University of California, Berkeley
[4]Physics Division, Lawrence Berkeley National Laboratory
[5]Department of Particle Physics and Astrophysics, Stanford University
[6]Center for Theoretical Physics, Massachusetts Institute of Technology
[7]The NSF AI Institute for Artificial Intelligence and Fundamental Interactions

uqureshi@cern.ch krish.desai@berkeley.edu nachman@stanford.edu jthaler@mit.edu

Abstract

Differential cross sections are the currency of scientific exchange in particle and nuclear physics, but complex detector effects must be removed or *unfolded* in order to render them useful. Recent machine-learning methods have enabled unbinned unfolding in high dimensions, though leading approaches are iterative and can degrade when measured data have limited overlapping support with detector-level simulations. We introduce *Reweighting Adversarial Networks* (RAN), a non-iterative unbinned unfolding method that learns pre-detector-level importance weights while matching detector-level data through a Wasserstein critic. RAN avoids repeated iterations, reduces computational overhead, and remains robust when detector-level overlap is poor. Evaluated on Gaussian and synthetic proton-proton collision experiments, RAN achieves competitive or improved closure relative to the state-of-the-art OMNIFOLD method while requiring only a single adversarial training stage.

1 Introduction

Correcting detector effects is an essential step in differential cross section measurements. Traditionally, this *unfolding* problem is solved using binned techniques, but binning limits dimensionality, introduces discretization artifacts, and complicates downstream analyses of derived observables [13]. Machine-learning approaches such as OMNIFOLD have demonstrated that unfolding can be performed directly on continuous, high-dimensional event representations [1, 4]. These methods, however, are typically iterative and require training multiple networks across several rounds, increasing computational cost and introducing a stopping ambiguity.

A second challenge is support overlap. Existing reweighting-based unbinned methods rely on classifier learning at detector level, so they can struggle when simulated detector-level events do not sufficiently cover the observed data, even if the underlying particle-level distributions still overlap [1, 11]. This motivates methods that infer particle-level corrections without explicitly performing detector-level reweighting.

We introduce *Reweighting Adversarial Networks* (RAN), a non-iterative unfolding method that learns particle-level weights as a function of *Generation* (i.e. particle-level Monte-Carlo (MC)) events and optimizes these weights by comparing the induced reweighted *Simulation* (defined as detector-level MC) distribution to observed data. We demonstrate its behavior in a controlled Gaussian

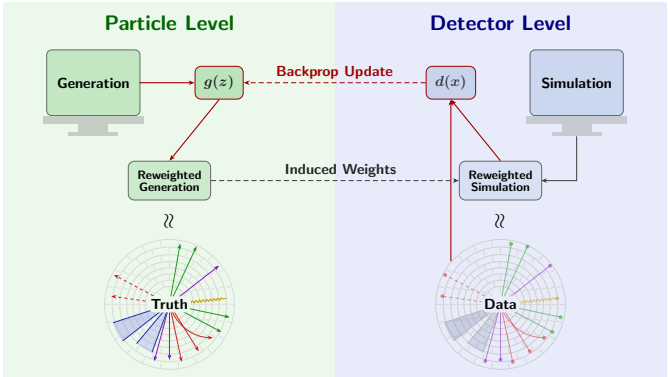

Figure 1: A high-level overview of RANs. The generator function g generates weights to reweight the particle-level simulation ('Generation'). The detector-level simulation ('Simulation') dataset inherits its weights from the matching Generation events. The discriminator d learns to classify events between Data and Reweighted Simulation.

setup designed to degrade detector-level overlap and in a realistic, synthetic proton-proton collision benchmark involving hadronic jet substructure.

2 Method

Consider paired events (z, x), where $z \in Z \subseteq \mathbb{R}^{N_T}$ denotes a particle-level event and $x \in X \subseteq \mathbb{R}^{N_D}$ its detector-level counterpart. Let natural events (Truth and Data) be distributed according to a joint probability density function $p(Z_T = z, X_D = x)$, and let Monte-Carlo events (Generation and Simulation) be distributed according to joint probability density function $q(Z_G = z, X_S = x)$. RAN, like all other unfolding methods assumes that the detector response function $r(x|z)$ is universal, such that the probability of seeing detector-level x conditioned on particle-level z is $r(x|z) = p(x|z) = q(x|z)$. The goal is to reweight the particle-level Generation's probability density $q(z)$ into an unfolded density $\widetilde{q}(z)$ such that the corresponding detector-level distribution $\widetilde{q}(x)$ is statistically indistinguishable (by a well trained Wasserstein critic) from the observed data distribution $p(x)$. When the reweighted Simulation, $\widetilde{q}(x)$, is statistically indistinguishable from the Data, $p(x)$, we demonstrate that the reweighted Generation, $\widetilde{q}(z)$, provides a good reconstruction for the Truth, $p(z)$. RAN parameterizes the particle-level weights through:

$$\widetilde{q}(z) = g(z)\, q(z), \tag{1}$$

$$g(z) = \frac{a(\mathrm{NN}(z))}{\widehat{P}}, \tag{2}$$

where $\mathrm{NN}(z)$ is a generic neural network, $\widehat{P}$ is a batch-wise normalization factor, and $a(\cdot)$ is a positive output activation to avoid potential negative weights. This is an extension of the MOMENT UNFOLDING method [8], replacing a finite order polynomial of degree N in the exponent with a generic function approximation. For Moment Unfolding, a byproduct of estimating only moments is that it regularized the inverse problem by limiting the value of N. For RANs, we use the following activation function to regulate the weight function:

$$a(t) = \log(1 + e\,\mathrm{softplus}(t)), \tag{3}$$

$$\mathrm{softplus}(t) = \log\big(1 + e^t\big), \tag{4}$$

which grows logarithmically as $t \to \infty$ thereby preserving positivity while suppressing extreme weights for large inputs more effectively than a pure exponential, or even linearly growing activations like ReLU [14]. The reweighted detector-level sample is then compared with data using a Wasserstein critic $c(x)$. The training objective is:

$$\mathcal{L}_{\mathrm{WGAN}}[g, c] = \frac{\sum_{(z,x)\in\mathrm{sim.}} g(z)\, c(x)}{\sum_{z\in\mathrm{sim.}} g(z)} - \mathbb{E}_{x\sim\mathrm{data}}[c(x)] + \lambda\, \mathbb{E}\Big[\big(\|\nabla_{\widehat{x}} c(\widehat{x})\| - 1\big)^2\Big], \tag{5}$$

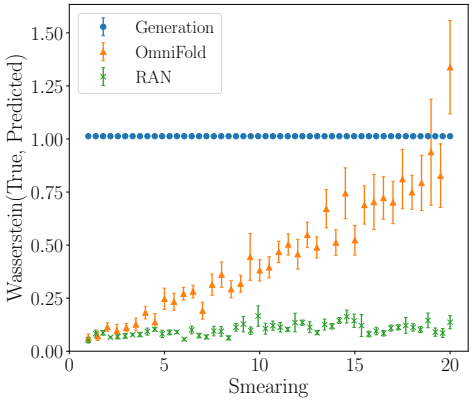

Figure 2: Wasserstein distance between each unfolded distribution and the Truth distribution as a function of the detector smearing factor for OMNIFOLD (orange triangles), RAN (green crosses), and the Generation baseline (blue circles).

with the final term enforcing the gradient penalty [10]. The generator minimizes the discrepancy between weighted Simulation and Data, while the critic estimates the Wasserstein-1 distance between the two detector-level distributions [3].

Three ingredients are important for stable training. First, the Wasserstein objective provides smoother gradients than a binary classifier when distributions are weakly overlapping [3].

Second, the generator is pretrained near the identity map so that training begins from weights close to one, reflecting the prior belief that simulation is already a useful approximation. RAN performs only a single adversarial optimization and reweights events exclusively at particle level. This avoids detector-level reweighting steps and associated failure modes when overlap is poor.

3 Experiments

3.1 Gaussian overlap stress test

To isolate the effect of detector-level support mismatch, we construct a one-dimensional Gaussian example. The target particle-level distribution is $Z_T \sim \mathcal{N}(0, 1)$ and the simulated particle-level distribution is $Z_G \sim \mathcal{N}(-1, 1)$. These are passed through a detector response that scales the variable by a multiplicative smearing factor, σ, so that $X_{\text{Sim.}} = \sigma \times Z_G \sim \mathcal{N}(-\sigma, \sigma^2)$.

As the smearing increases, detector-level data and simulation become increasingly separated, while the particle-level overlap remains fixed. As Figure 2 demonstrates, OMNIFOLD performs well when detector-level overlap is good, but performance degrades as overlap worsens because the classifier can no longer infer reliable detector-level weights [1]. RAN remains stable over the same range because it never directly attempts detector-level reweighting; instead, it adjusts particle-level weights using gradients from a Wasserstein critic and infers the corresponding detector-level distribution through the natural pairing between Simulation and Generation events. The resulting unfolded distributions maintain an approximately constant Wasserstein distance to truth even in regimes where OMNIFOLD deteriorates.

3.2 Jet substructure benchmark

We study hadronic jets in proton–proton collisions, following the standard OMNIFOLD benchmark setup [1, 2]. Events are generated at $\sqrt{s} = 14$ TeV using PYTHIA (v8.243) for the simulation prior [15] and HERWIG (v7.1.5) for the target truth distribution [5]. Detector effects are modeled with DELPHES (v3.4.2) [9], and jets are clustered with anti-k_T and radius parameter $R = 0.4$ [7, 6]. We consider six standard observables: jet mass m, constituent multiplicity M, width w, N-subjettiness ratio τ_{21} [16], SoftDrop momentum fraction z_g, and groomed jet mass $\ln \rho$ [12].

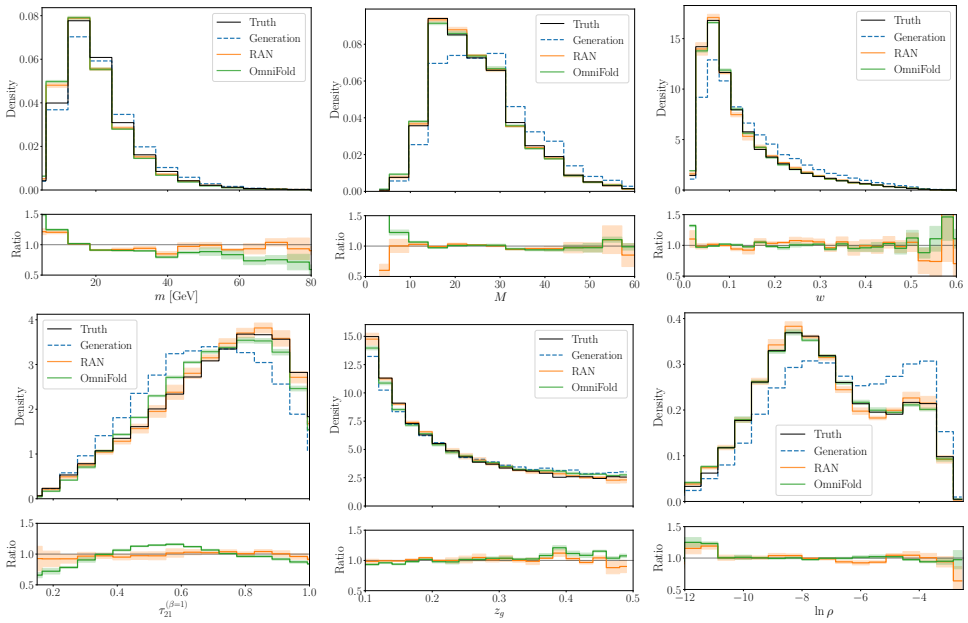

Figure 3: Distributions of jet mass m (top-left), constituent multiplicity M (top-middle), width w (top-right), the N-subjettiness ratio τ_{21} (bottom-left), SoftDrop groomed momentum fraction z_g (left), and the groomed jet mass $\ln \rho$ (right) at particle-level, comparing the generation (blue dashed), truth (solid black), OMNIFOLD (solid green) and RAN (solid orange). The error bars represent the combined uncertainty from bootstrapping the training data and variation across random initializations, estimated from 5 independent runs each using a different bootstrap resample and random seed.

Table 1: Wasserstein distance between truth and unfolded prediction for the substructure observables.

	m	M	w	τ_{21}	z_g	$\ln \rho$
RAN	$\mathbf{0.073 \pm 0.010}$	$\mathbf{0.017 \pm 0.004}$	0.018 ± 0.012	$\mathbf{0.024 \pm 0.006}$	$\mathbf{0.011 \pm 0.004}$	0.023 ± 0.010
OmniFold	0.116 ± 0.009	0.027 ± 0.006	$\mathbf{0.014 \pm 0.004}$	0.067 ± 0.004	0.060 ± 0.008	$\mathbf{0.015 \pm 0.005}$

Across all six observables, both RAN and OMNIFOLD achieve strong closure with respect to truth. RAN is generally as good as or better than OMNIFOLD, with particularly visible gains for jet mass, multiplicity, τ_{21}, and z_g. Using the Wasserstein distance as the main quantitative summary, RAN yields the best score for four of the six observables and remains competitive on the remaining two. Representative values are shown in Table 1. These results indicate that the non-iterative adversarial formulation can retain the strengths of modern unbinned unfolding while improving robustness.

4 Conclusion

We introduced Reweighting Adversarial Networks, a non-iterative method for unbinned unfolding. RAN learns particle-level importance weights but evaluates their quality through detector-level matching with a Wasserstein critic. This formulation removes the repeated iterations required by OMNIFOLD-like methods and offers improved robustness when detector-level support overlap is limited.

In a Gaussian stress test, RAN remains stable as detector-level overlap deteriorates. In a realistic jet-substructure benchmark, it achieves competitive or superior closure relative to OMNIFOLD across standard observables. These results suggest that adversarial particle-level reweighting is a promising direction for precision unfolding in high-energy physics.

Future work will incorporate nuisance parameters [17] and systematic uncertainties in fully data-ready analyses, as well as extend the method to richer full phase space event representations such as point clouds.

References

[1] Anders Andreassen et al. "OmniFold: A Method to Simultaneously Unfold All Observables". In: *Phys. Rev. Lett.* 124 (2020), p. 182001. arXiv: 1911.09107 [hep-ph].

[2] Anders Andreassen et al. "OmniFold: A Method to Simultaneously Unfold All Observables". In: *Phys. Rev. Lett.* 124.18 (2020), p. 182001. DOI: 10.1103/PhysRevLett.124.182001. arXiv: 1911.09107 [hep-ph].

[3] Martin Arjovsky, Soumith Chintala, and Léon Bottou. "Wasserstein Generative Adversarial Networks". In: *Proceedings of the 34th International Conference on Machine Learning*. Ed. by Doina Precup and Yee Whye Teh. Vol. 70. Proceedings of Machine Learning Research. PMLR, 2017, pp. 214–223. URL: https://proceedings.mlr.press/v70/arjovsky17a.html.

[4] Miguel Arratia et al. "The Landscape of Machine Learning in HEP Unfolding". In: *JINST* 17 (2022), P10018. arXiv: 2107.08979 [hep-ph].

[5] Johannes Bellm et al. "Herwig 7.2 release note". In: *Eur. Phys. J. C* 80.5 (2020), p. 452. DOI: 10.1140/epjc/s10052-020-8011-x. arXiv: 1912.06509 [hep-ph].

[6] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "FastJet User Manual". In: *Eur. Phys. J. C* 72 (2012), p. 1896. DOI: 10.1140/epjc/s10052-012-1896-2. arXiv: 1111.6097 [hep-ph].

[7] Matteo Cacciari, Gavin P. Salam, and Gregory Soyez. "The anti-k_t jet clustering algorithm". In: *JHEP* 04 (2008), p. 063. DOI: 10.1088/1126-6708/2008/04/063. arXiv: 0802.1189 [hep-ph].

[8] Krish Desai, Benjamin Nachman, and Jesse Thaler. "Moment Extraction Using an Unfolding Protocol Without Binning". In: *Physical Review D* 110.11, 116013 (Dec. 13, 2024). DOI: 10.1103/PhysRevD.110.116013. arXiv: 2407.11284 [hep-ph].

[9] J. de Favereau et al. "DELPHES 3, A modular framework for fast simulation of a generic collider experiment". In: *JHEP* 02 (2014), p. 057. DOI: 10.1007/JHEP02(2014)057. arXiv: 1307.6346 [hep-ex].

[10] Ishaan Gulrajani et al. *Improved Training of Wasserstein GANs*. 2017. arXiv: 1704.00028 [cs.LG]. URL: https://arxiv.org/abs/1704.00028.

[11] Nathan Huetsch et al. "The landscape of unfolding with machine learning". In: *SciPost Phys.* 18.2 (2025), p. 070. DOI: 10.21468/SciPostPhys.18.2.070. arXiv: 2404.18807 [hep-ph].

[12] Andrew J. Larkoski et al. "Soft Drop". In: *JHEP* 05 (2014), p. 146. DOI: 10.1007/JHEP05(2014)146. arXiv: 1402.2657 [hep-ph].

[13] Eamonn Maguire, Lukas Heinrich, and Graeme Watt. "HEPData: a repository for high energy physics data". In: *J. Phys. Conf. Ser.* 898.10 (2017). Ed. by Richard Mount and Craig Tull, p. 102006. DOI: 10.1088/1742-6596/898/10/102006. arXiv: 1704.05473 [hep-ex].

[14] Vinod Nair and Geoffrey E. Hinton. "Rectified Linear Units Improve Restricted Boltzmann Machines". In: *Proceedings of the 27th International Conference on Machine Learning*. International Conference on Machine Learning (Haifa, Israel). Omnipress, 2010, pp. 807–814.

[15] Torbjörn Sjöstrand et al. "An introduction to PYTHIA 8.2". In: *Comput. Phys. Commun.* 191 (2015), pp. 159–177. DOI: 10.1016/j.cpc.2015.01.024. arXiv: 1410.3012 [hep-ph].

[16] Jesse Thaler and Ken Van Tilburg. "Identifying Boosted Objects with N-subjettiness". In: *JHEP* 03 (2011), p. 015. DOI: 10.1007/JHEP03(2011)015. arXiv: 1011.2268 [hep-ph].

[17] Huanbiao Zhu et al. "Machine Learning-based Unfolding for Cross Section Measurements in the Presence of Nuisance Parameters". In: (Dec. 2025). arXiv: 2512.07074 [stat.AP].