End-To-End Latent Variational Diffusion Models for Inverse Problems in High Energy Physics

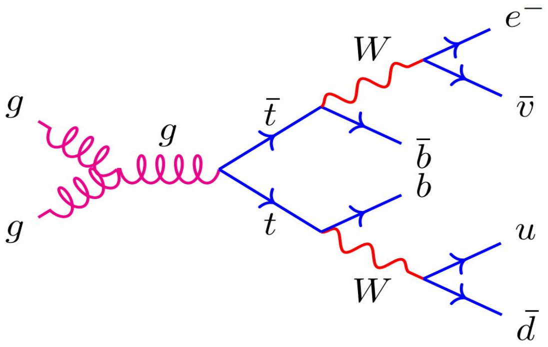
Alexander Shmakov, Kevin Greif, Michael Fenton, Aishik Ghosh, Pierre Baldi, Daniel Whiteson

November 8 2023
ML4Jets 2023
DESY

Unfolding At The LHC

- Every cross section measurement subject to detector effects
- Process of removing detector effects from measurements known as **unfolding**
    - Excellent example of an inverse problem in HEP
- Investigate **latent diffusion model** for generative unfolding

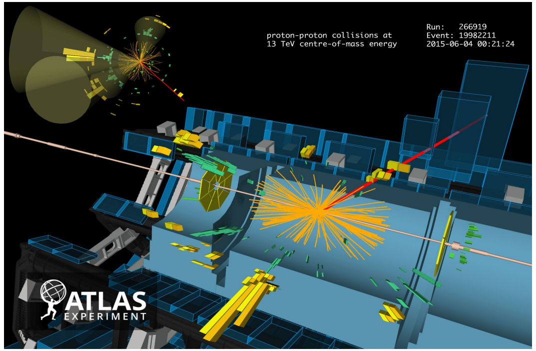
Parton



Simulation

Unfolding

Detector



Inverse problem: (noun) a problem with no solution

The Landscape of ML Based Unfolding

Discriminative

- **Omnifold** is being used to measure physics!
  - H1 collaboration substructure measurement
- By far the most mature high dimensional unfolding algorithm
- **What if we have limited data events?**

Generative

- Learn to generate unfolded data conditioned on detector observations
- Lots of interesting ideas, and open questions
- Could work even with limited data events
- But prior dependence, variable dimensions, etc…

The Landscape of ML Based Unfolding

Discriminative

- **Omnifold** is being used to measure physics!
    - [H1 collaboration substructure measurement]
- By far the most mature high dimensional unfolding algorithm
- **What if we have limited data events?**

Generative

- Learn to generate unfolded data conditioned on detector observations
- Lots of interesting ideas, and open questions
- Could work even with limited data events
- But prior dependence, variable dimensions, etc…

This Talk!

From Diffusion to Variational Latent Diffusion

- **Diffusion** models are a type of *conditional (c)* generative model which learns the reverse dynamics for a Gaussian stochastic differential equation.
- Given a *noise schedule* based on log signal-to-noise ratio γ, define our flow.

$$\sigma_t = \sqrt{\text{sigmoid}(\gamma_\phi(t))} \text{ and } \alpha_t = \sqrt{\text{sigmoid}(-\gamma_\phi(t))}$$

$$x(t) = \alpha(t)x(0) + \sigma(t)\epsilon_t \text{ where } \epsilon_t \sim \mathcal{N}(0,1)$$

- Train a network to estimate ϵ and sample according to inverse SDE.

$$\mathcal{L} = \mathbb{E}_{\epsilon\sim\mathcal{N}(0,1),t\sim\mathcal{U}(0,1)}\left[\gamma'(t)\,\|\epsilon - \hat{\epsilon}(x_t, t, c)\|_2^2\right]$$

$$\int_1^0 f(t)x(t) + \frac{g^2(t)\hat{\epsilon}(x_t, t, c)}{2\sigma}dt$$

[1] Yang Song, Jascha Sohl-Dickstein, Diederik P Kingma, Abhishek Kumar, Stefano Ermon, and Ben Poole. Score-based generative modeling through stochastic differential equations. ICLR 2021.

From Diffusion to Variational Latent Diffusion

- **Latent Diffusion (LDM)**[2] performs the forward and reverse SDE in a *latent space* from a pre-trained VAE. This VAE is usually pre-trained in either unsupervised manner with only the data or with a contrastive objective such as CLIP.

$$x \rightarrow z \sim VAE(x)$$

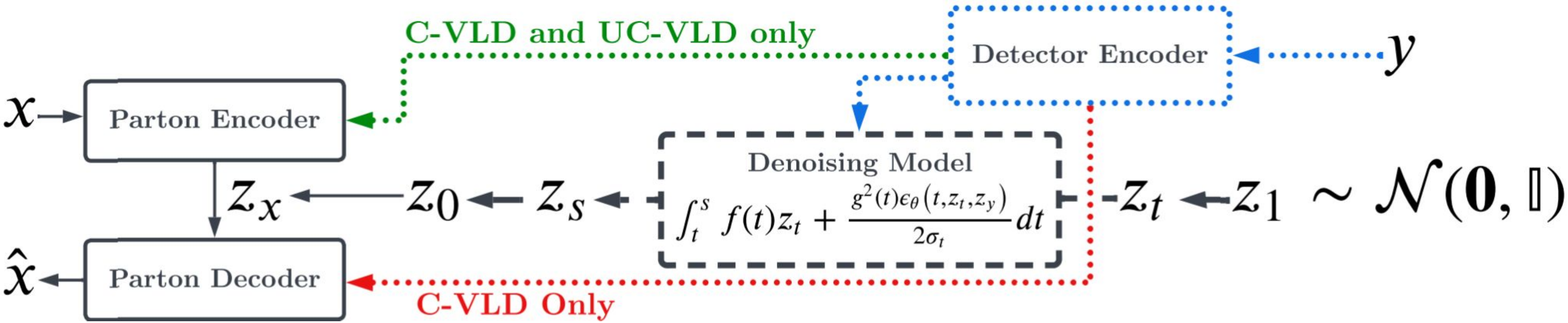
- **Variational Diffusion (VDM)**[3] Interprets the entire diffusion model as a hierarchical variational model with infinite depth. This allows us to learn an optimal noise schedule for our diffusion.

$$\gamma \rightarrow \gamma_\phi(t) \text{ with } \mathcal{L}_\phi = Var[\mathcal{L}_{\text{diffusion}}]$$

[2] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models, 2021..
[3] Diederik Kingma, Tim Salimans, Ben Poole, and Jonathan Ho. Variational diffusion models. Advances in neural information processing systems, 34:21696–21707, 2021.

From Diffusion to Variational Latent Diffusion

- We combine these ideas into a single unified end-to-end model.
- VAE is optimized to find ideal space to perform diffusion in.
  - Interpreted as another layer in the hierarchical VAE, introduce additional regularization loss.
  - Latent space may be **higher dimensional** than data space!
- Noise Schedule is optimized simultaneously as in VDM.
  - Continuous time diffusion process is used for training.
  - Inference is performed in discrete time.



Distribution-Free Metrics Results

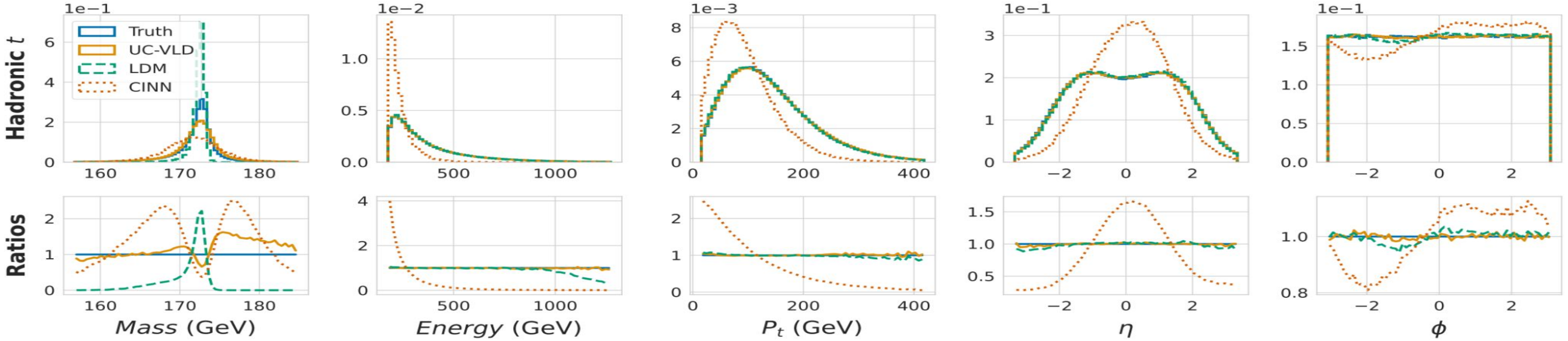
- Compare each of these models to evaluate the effect of the latent space and unified training.
- Also compare the a simple Conditional VAE (CVAE) and a normalizing flow-based model (CINN)[4].
- Notice latent space is **very important** to model performance, and unified training outperformed pre-trained LDM.

	Wasserstein	Energy	K-S	KL_{64}	KL_{128}	KL_{256}
VLD	108.76	7.59	4.08	**3.47**	**3.74**	**4.53**
UC-VLD	**73.56**	**6.35**	**3.41**	5.77	7.10	8.48
C-VLD	389.62	25.39	4.65	9.54	10.09	10.79
LDM	402.32	24.09	5.91	14.71	16.34	17.92
VDM	2478.35	181.35	17.14	29.28	32.29	35.60
CVAE	484.56	32.29	6.37	7.79	9.17	10.60
CINN	3009.08	185.13	15.74	28.55	30.19	32.37

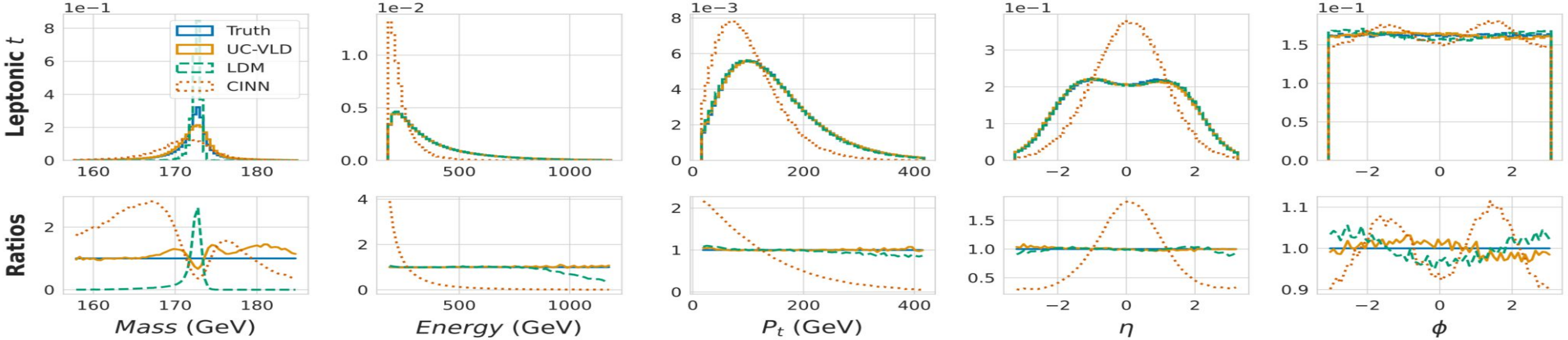
[4] Marco Bellagente, Anja Butter, Gregor Kasieczka, Tilman Plehn, Armand Rousselot, Ramon Winterhalder, Lynton Ardizzone, and Ullrich Köthe. Invertible networks or partons to detector and back again. SciPost Phys., 9:074, 2020.

Testing Dataset Kinematics Distributions

Hadronic Top Quark Kinematics



Leptonic Top Quark Kinematics



Future Work

- Want to perform **particle** level unfolding
- We lose the definite fixed-length encoding available due to the parton's feynman diagram. **Particle events are variable length!**
- Extend this method to be able to unfold an arbitrary number of objects simultaneously.
- Investigate **prior dependence** of generative unfolding
- Acceptance effects, inefficiencies, systematics, etc. etc. etc.

"Unfolding is a complicated business and one is well advised to ask in each problem if it can be avoided"
- G. Cowan, A Survey of Unfolding Methods for Particle Physics

Thank you

Posterior Distribution Examples

- Compare posterior distributions produced by the LVD for individual example events to an empirical estimate of the posterior from the testing dataset.
- Notice that the LVD has very tight posteriors for challenging kinematics including the Mass and Pt.
- Notice that the LVD managed to discover a bi-modal posterior for neutrino eta!