Resilient VAE for Unsupervised Anomaly Detection at LCLS

R. Humble[1], W. Colocho[2], E. Darve[1], D. Ratner[2]

[1] Institute for Computational and Mathematical Engineering, Stanford University

[2] SLAC National Accelerator Laboratory

SLAC NATIONAL ACCELERATOR LABORATORY

Stanford University

Outline

1. **Overview**: Anomalies at LCLS
2. **Method**: Resilient VAE for Unsupervised Anomaly Detection

Anomalies at LCLS

What is the Linac Coherent Light Source (LCLS)?

- LCLS is a hard X-ray free electron laser (FEL)
- Delivers X-ray laser to users around the clock
- User experiments demand <u>stability</u>
- Produces over 200,000 data streams

Many LCLS failure modes – classified as two types

Type A: Downtime (beam goes offline)

- ~3% of availability lost
- >180 hours/year
- ~3 full user experiments

Type B: FEL Performance Degradation

- Experiment noise
- LCLS anomalies → User anomalies

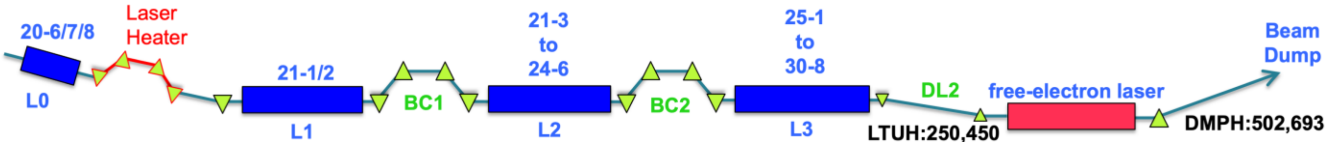
Both types of failures lead to science loss

***Response*:** Operators guard against anomalies

- Limit the operational range of accelerator
- Hold certain components in reserve

Three characteristics make finding anomalies hard

Dataset: Pulse-by-pulse readings from 151 beam position monitors (X,Y,TMIT)



Goals:

1. Find anomalies in the LCLS beam data
2. Attribute the anomaly back to particular beam readings

Challenge: Beam data is

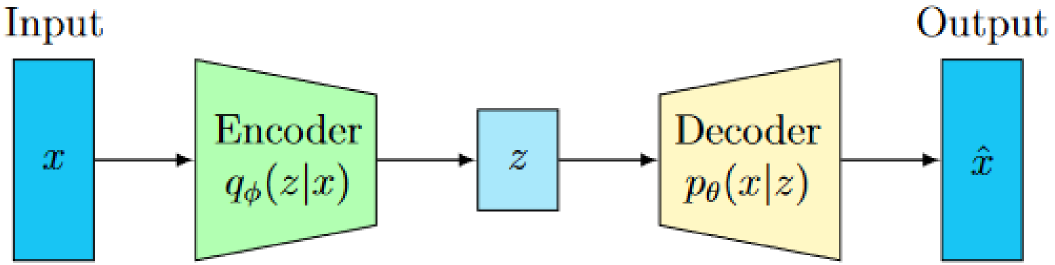
- *Unlabeled* → no supervised methods
- *High (>450) dimensional* → curse of dimensionality
- *Contaminated* → no normal training set exists

Need specialized method

Resilient VAE for Unsupervised Anomaly Detection

Background: VAEs for Anomaly Detection

Variational autoencoders (VAEs) widely used for anomaly detection



Typically trained on only normal data

Assumption: Only normal data is well-reconstructed through low-dimension compression

Poor reconstruction → Large anomaly score

VAE struggles to train on contaminated data

Reconstruction loss

Latent penalty

VAE Loss

$$\max_{\theta,\phi} L_{VAE}(x;\theta,\phi) = \mathrm{E}_{q_\phi(z|x)}[\log p_\theta(x|z)] - D_{KL}\left(q_\phi(z|x) \,||\, p(z)\right)$$

Gaussian Data Model

$$p_\theta(x_d|z) \sim N(\mu_d(z;\theta); \sigma_d(z;\theta)^2)$$

Gaussian VAE Loss

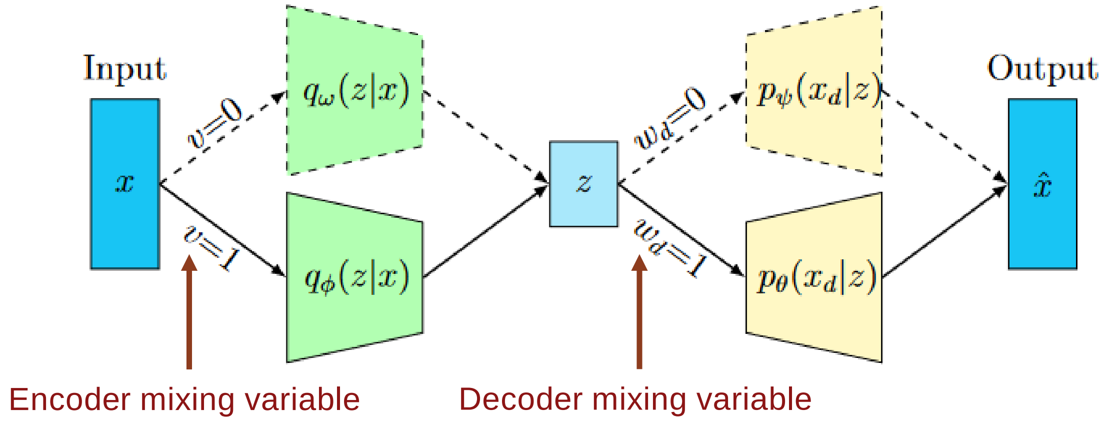
$$\max_{\theta,\phi} L_{VAE}(x;\theta,\phi) = -\frac{1}{2}\sum_d \left(\frac{x_d - \mu_d}{\sigma_d}\right)^2 + \cdots$$

Loss dominated by worst reconstructions → Abnormal examples!

Note: Gradient clipping does not help here! Only addresses magnitude of gradients.

Modify to be "resilient" to anomalies → Resilient VAE

Idea: Create "outlier" path through the network[1]



Path through the network depends on <u>mixing variables</u> v and w_d

Encoder Mixing

$$\boxed{q(z|x,v) = q_\phi(z|x)^v q_\omega(z|x)^{1-v}}$$

Decoder Mixing

$$\boxed{p(x_d|z,w_d) = p_\theta(x_d|z)^{w_d} p_\psi(x_d|z)^{1-w_d}}$$

Divert anomalous-looking examples through new path

[1] *Extends ideas first proposed in:* Eduardo, S. "*Robust Variational Autoencoders*" (2020).

ResVAE uses probabilistic inference to determine path

Add mixing variables to probabilistic model

Sample Mixing

$$\begin{array}{lll} \textit{Generation:} & p(v) & \sim Ber(\eta) \\ \textit{Inference:} & q(v|x) & \sim Ber\big(\gamma(x)\big) \end{array}$$

Feature Mixing

$$\begin{array}{lll} \textit{Generation:} & p(w_d|v) & \sim Ber(v\alpha_d) \\ \textit{Inference:} & q(w_d|x,v) & \sim Ber\big(v\pi_d(x)\big) \end{array}$$

Recover $\gamma(x)$ and $\pi_d(x)$ from probabilistic inference

Closed form solution via coordinate variational inference (CVI)

Sample Mixing

$$\gamma(x) = \mathrm{sigmoid}(\lambda_3 g + \eta)$$

"sample reconstruction ratio"

Feature Mixing

$$\pi_d(x) = \mathrm{sigmoid}\big(\lambda_{2,d} r_d + \alpha_d\big)$$

"feature reconstruction ratio"

***Intuition*: Pick the path with best reconstruction**

Loss adjusted to disregard anomalies during training

VAE Loss

$$\max_{\theta,\phi} L_{VAE}(x;\theta,\phi) = \sum_d \mathrm{E}_{q_\phi(z|x)}[\log p_\theta(x_d|z)] - D_{KL}\left(q_\phi(z|x) \,||\, p(z)\right)$$

Resilient VAE Loss

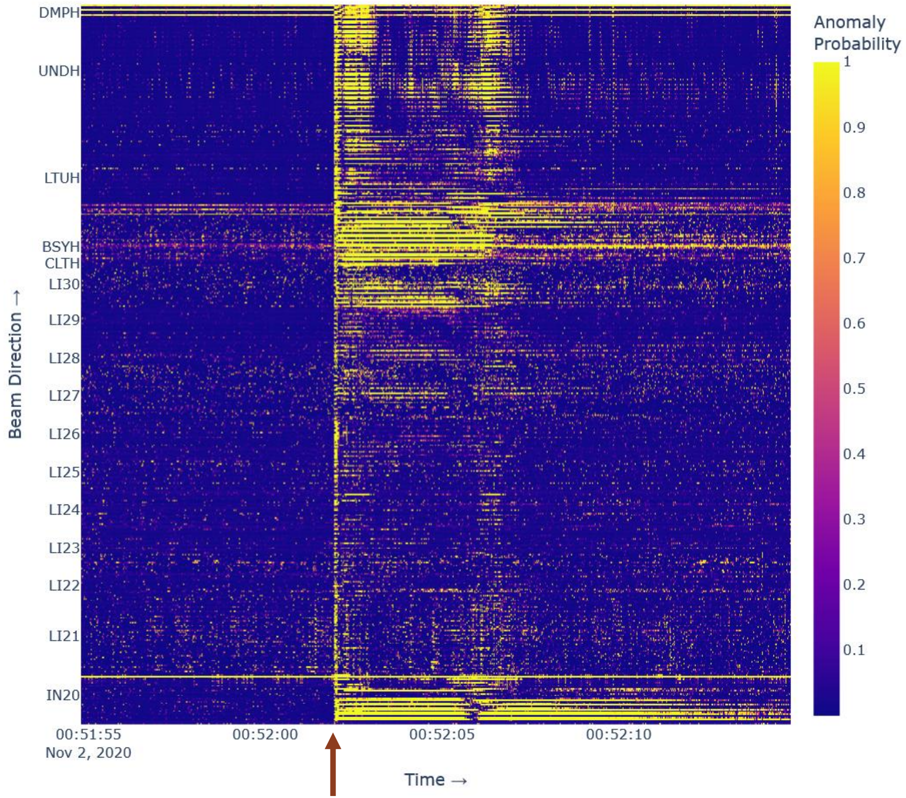
Feature weighting

$$\max_{\theta,\phi} L_{ResVAE}(x;\theta,\phi) = \boxed{\gamma(x)} \sum_d \boxed{\pi_d(x)} \mathrm{E}_{q_\phi(z|x)}[\log p_\theta(x_d|z)] - \boxed{\gamma(x)} D_{KL}\left(q_\phi(z|x) \,||\, p(z)\right)$$

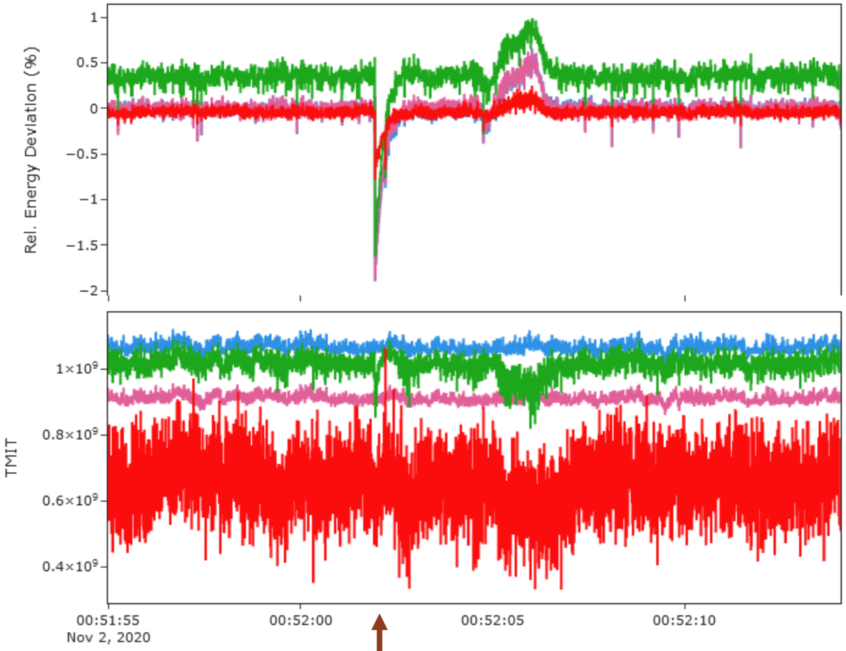
Sample weighting

Focus only on normal-looking examples!

ResVAE used to identify and diagnose LCLS anomalies

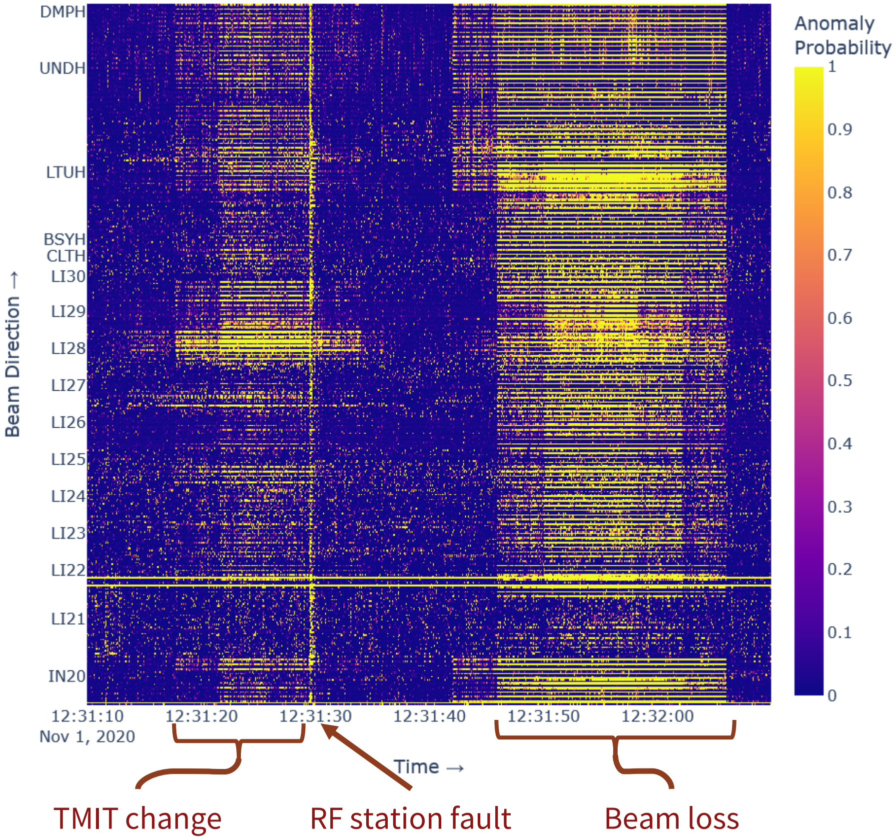


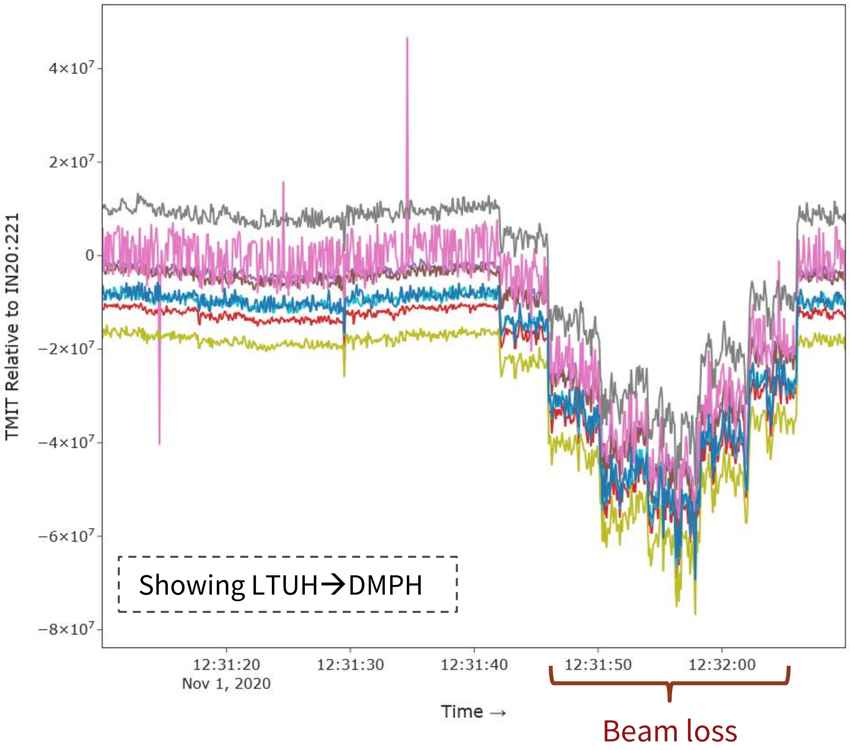
RF station fault



RF station fault

ResVAE used to identify and diagnose LCLS anomalies





Open questions

How to characterize the different anomaly types?

Can the root cause be identified?

How to present this information to operators?

Can the model learn from operator feedback?

Thank you!

Appendix

Resilient VAE: Full generative model

Generation / Decoder

$$
\begin{aligned}
p(v) &= \mathrm{Ber}(\beta) \\
p(w_d|v) &= \mathrm{Ber}(\alpha_d^v 0^{1-v}) \\
p(z|v) &= p(z)^v \boxed{p_o}(z)^{1-v} \\
p(x_d|z, w_d) &= p_\theta(x_d|z)^{w_d} \boxed{p_\psi}(x_d|z)^{1-w_d}
\end{aligned}
$$

Inference / Encoder

$$
\begin{aligned}
q(v|x) &= \mathrm{Ber}(\gamma(x)) \\
q(w_d|v, x) &= \mathrm{Ber}(\pi_d(x)^v 0^{1-v}) \\
q(z|v, x) &= q_\phi(z|x)^v \boxed{q_\omega}(z|x)^{1-v}
\end{aligned}
$$

Outlier distributions

Limiting to “good” beam conditions

Limit our dataset to periods when machine is running at “good” conditions

1. Stopper = inactive
2. Rate = 120Hz
3. Split = 120 Hz HXR/0 Hz SXR
4. Beam = actively logged
5. Charge > min. threshold

- Above conditions must be met at least 5 minutes
- Allow temporary (1 minute) violation of conditions 2-5
  - Machine will automatically respond to “catastrophic” errors
  - Without this condition, we filter out many of the worst faults