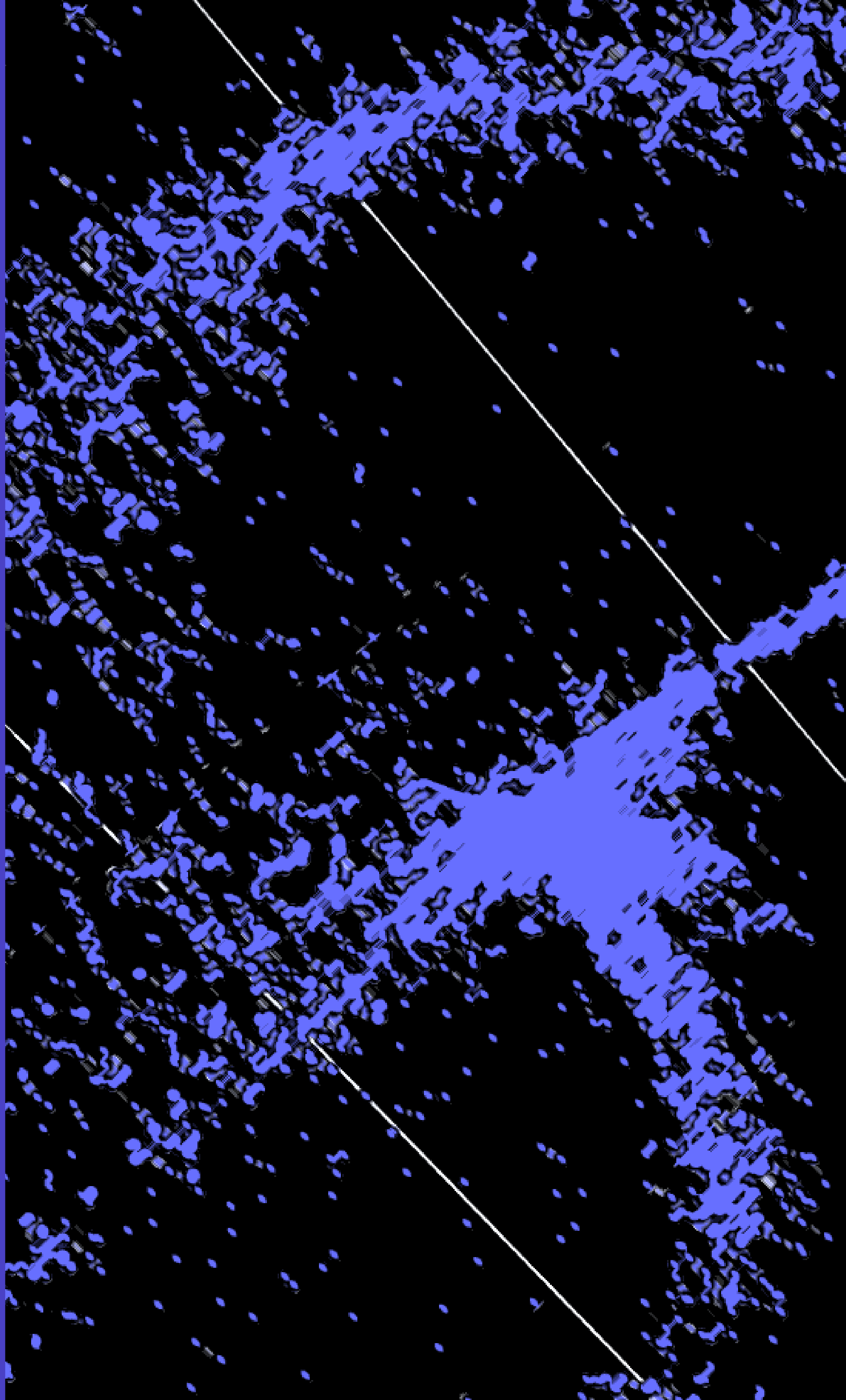
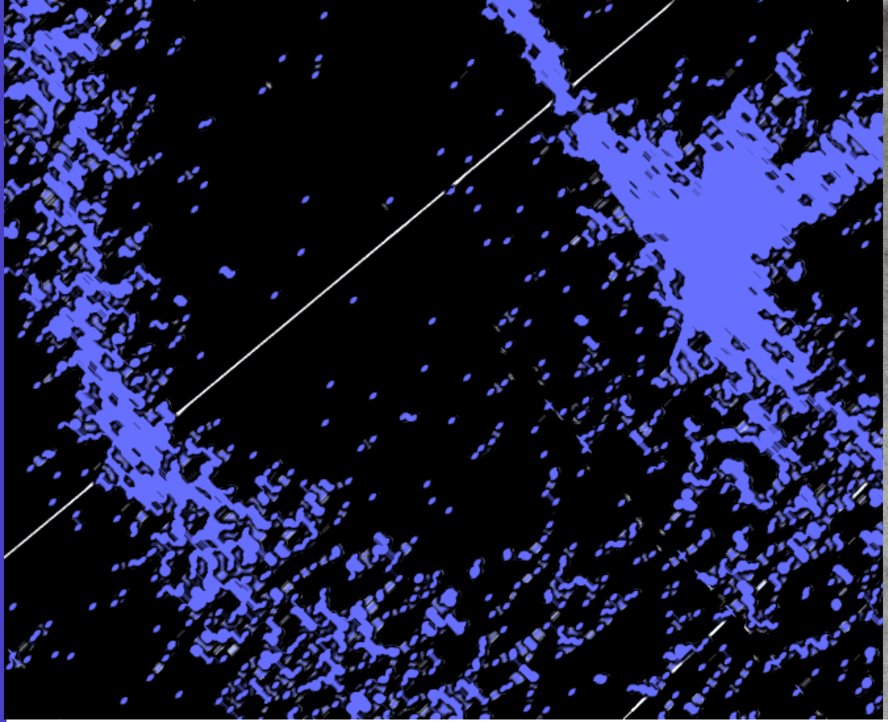
Non-Gaussian Sliced Wasserstein Autoencoders for Anomaly Detection in High Energy Physics

Thomas Stern

Supervisors: Amit Mishra, James Keaveney






Agenda

- Anomaly detection in HEP
- Sliced Wasserstein Autoencoders
- Latent space anomaly detection
- Experiments + Results
- Conclusions + Future research

Anomaly detection in HEP

- The standard model (SM) was completed with the discovery of the Higgs boson
- There are strong motivations for physics beyond the standard model (BSM).
  - The nature of dark matter and dark energy
  - the mass of neutrinos etc.…
- The large hadron collider (LHC) at CERN can shed light on these challenges

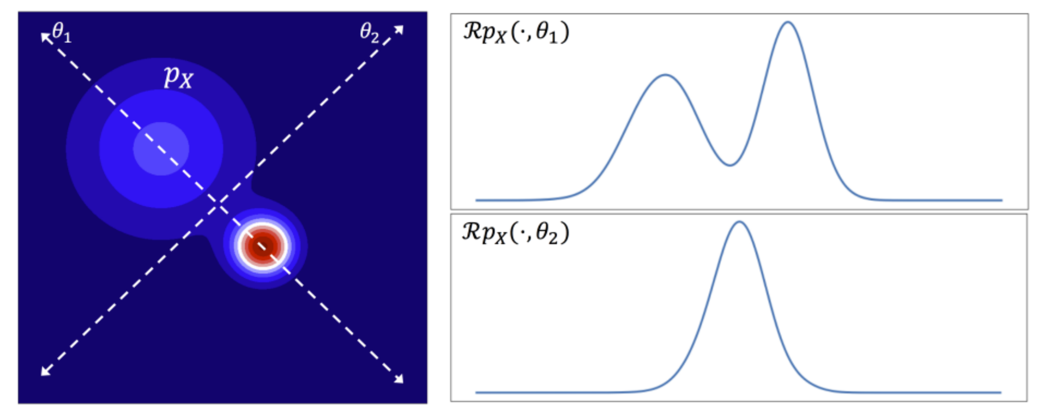
Most searches for new physics at CERN target specific experimental signatures

- The lack of a predefined target might turn this strength into a limitation
- Model dependence may have created blind spots
- Machine learning techniques have become the advocated avenue to reduce model dependence

When in the data processing pipeline the anomaly detection happens.

- Most anomaly detection is historically offline analysis
- There is a vast phase space deleted in real time
- Level-1 trigger rejects over 98% of events using algorithms implemented on custom electronic boards; optimized to accept for physics processes under study
  - Anomaly detection algorithms on L1 trigger could potentially improve event selection

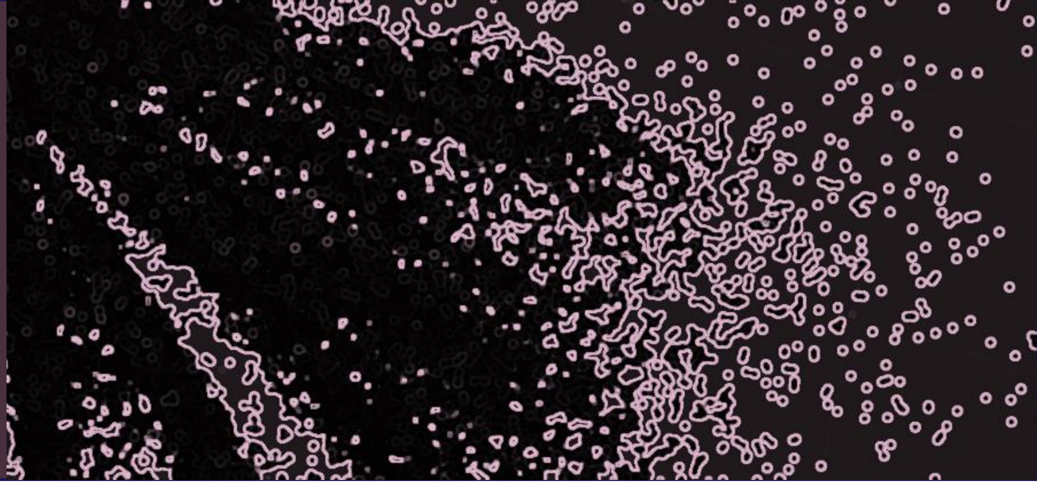
Sliced Wasserstein Autoencoder



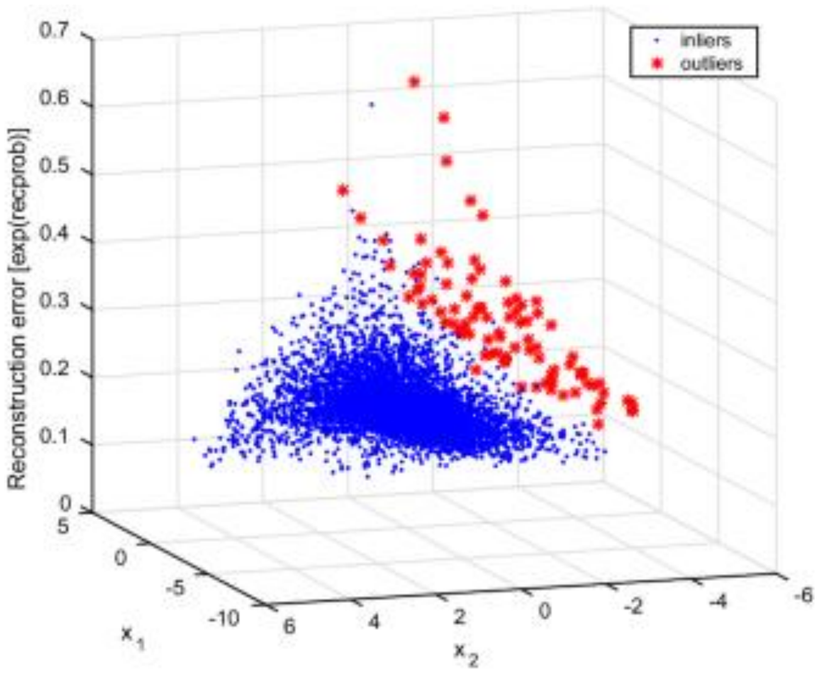
- Generative models
- Shape distribution of latent space into any sampleable probability distribution
- Autoencoder loss is regularized with the sliced-Wasserstein distance

- Easier to train:
  - parametric-free (compared to GANs);
  - almost hyperparameter-free (compared to MMD with kernels)
- Lower computational complexity
  - Good for FPGA code synthesis

Latent space anomaly detection



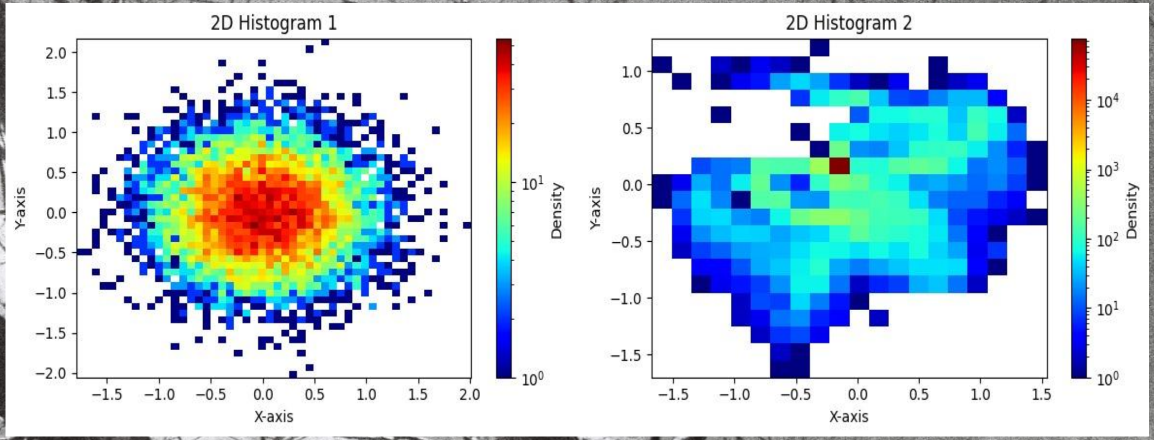
- Models generalize so well that they can also well reconstruct anomalies.
- Use the latent space distribution with the associated reconstruction error

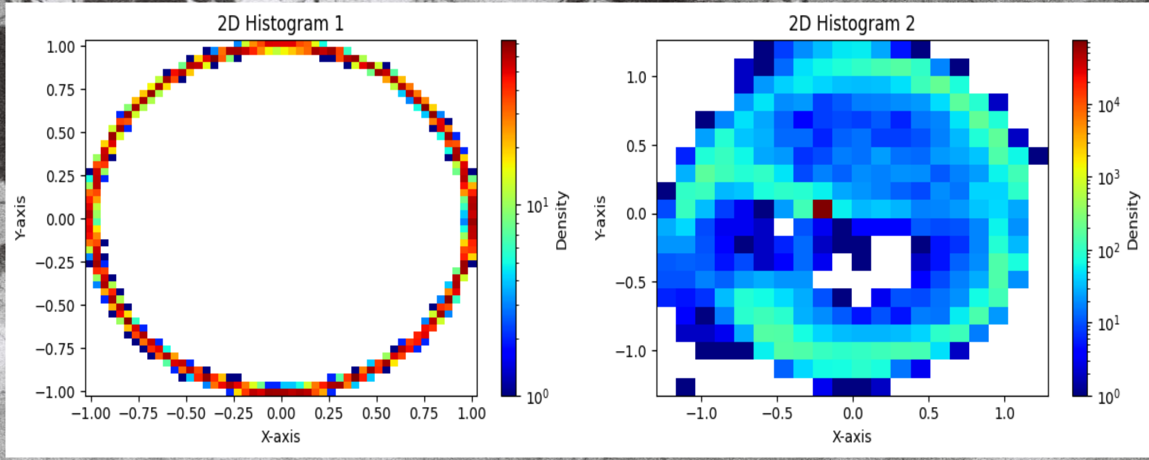


Experiments

- Using the ML4JETS dataset. We attempted to distinguish signal from background

- Different prior distributions were implemented

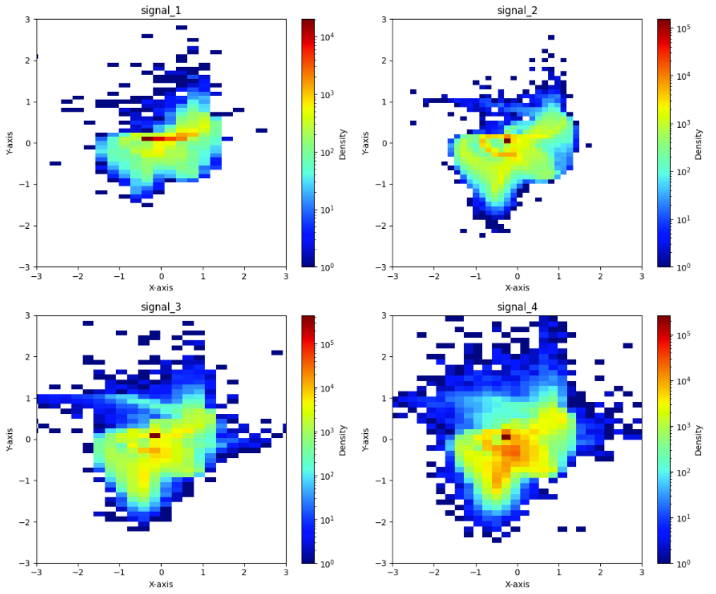




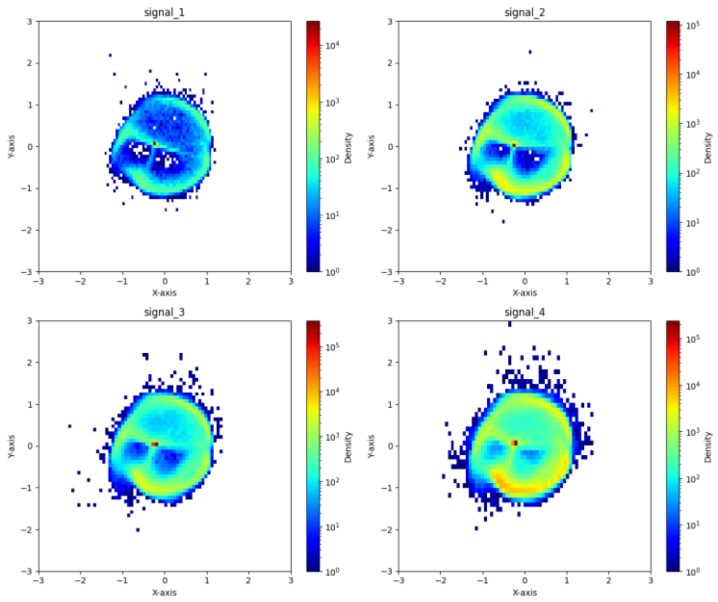
Choice of latent prior

- Showed no improvement in the AUC for anomaly detection based on the choice of prior
- The signal data is forced into the same distribution as the background encoding
- The MSE as an anomaly metric is oblivious to how the data is distributed

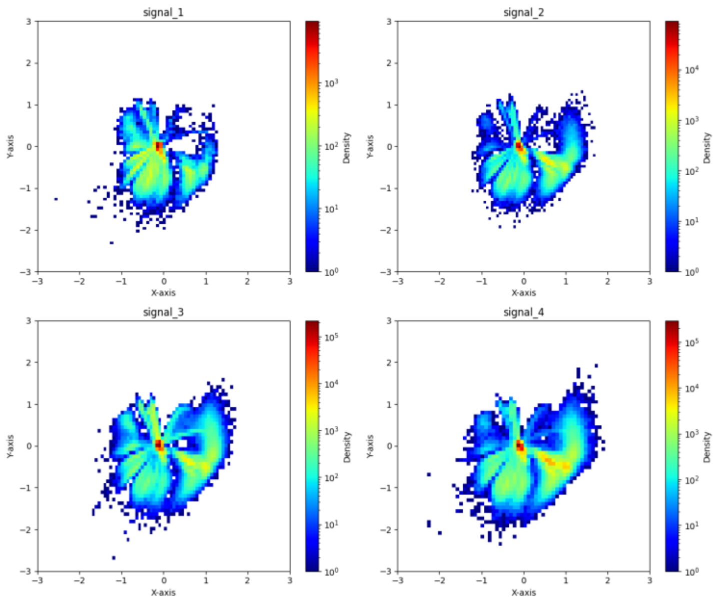
Gaussian



Circular

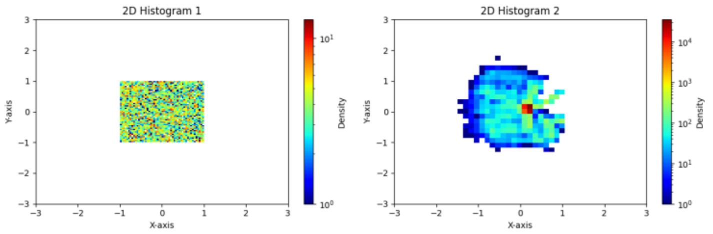


Uniform

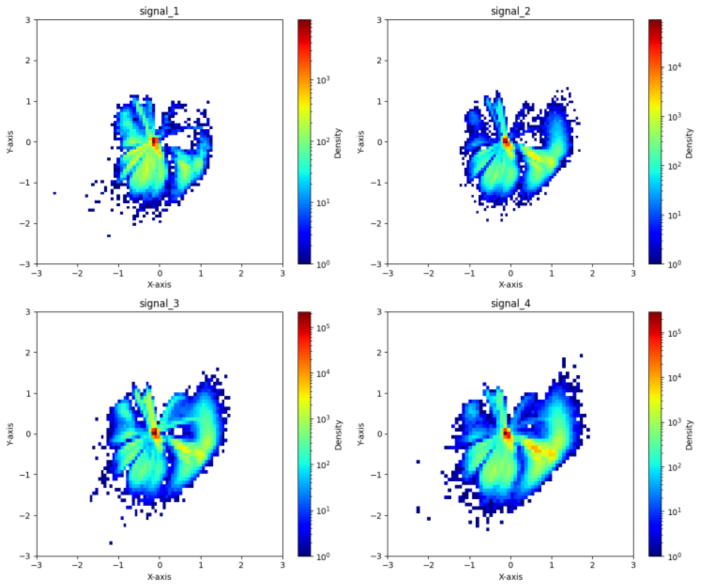


Uniform prior

Background encoding



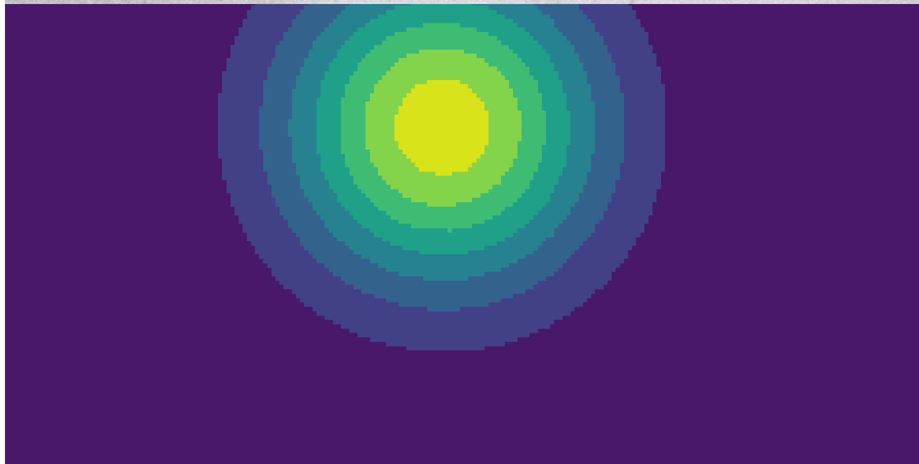
Signal encoding



- **Using a uniform prior:**
  - Encoding space is not overly constrained and naturally took on a gaussian distribution.
- **Allowing for anomaly detection using the encoding space**
  - MSE reconstruction + Mahalanobis distance
  - Identify anomalies the MSE would miss
- **Results:**
  - Showed 20% improvement in AUC for signal_1
  - 5% improvement in AUC overall

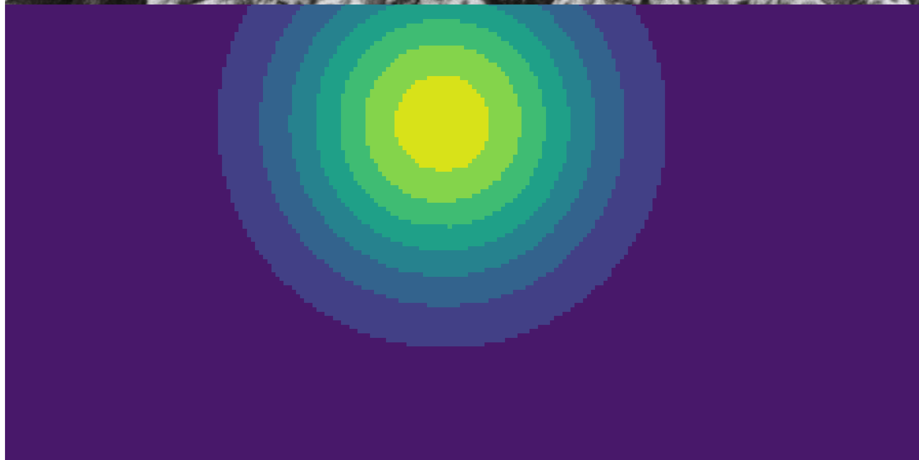
Conclusions

- Choice of prior doesn't inherently impact model's ability to identify anomalies
- Choice of latent prior impacts potential for latent space anomaly detection
- Latent anomaly detection shows improvements in anomaly detection



Future Research

- Gromov Wasserstein Autoencoders
- More sophisticated latent space anomaly detection measures
- Use the distribution to identify the difference between genuine and superficial anomaly



Thank You