Anomaly aware unsupervised learning for dark matter direct detection

Andre Scaffidi

In collaboration with with Juan Herrero-Garcia and Riley Patrick

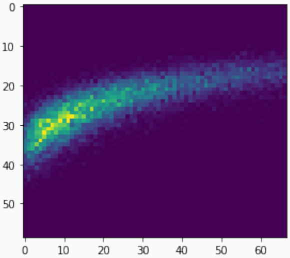
September 2, 2022

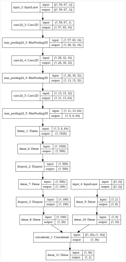
Istituto Nazionale di Fisica Nucleare, Sezione di Torino, via P. Giuria 1, I–10125 Torino, Italy

The Dark Universe:
A Synergic Multimessenger
Approach

Overview

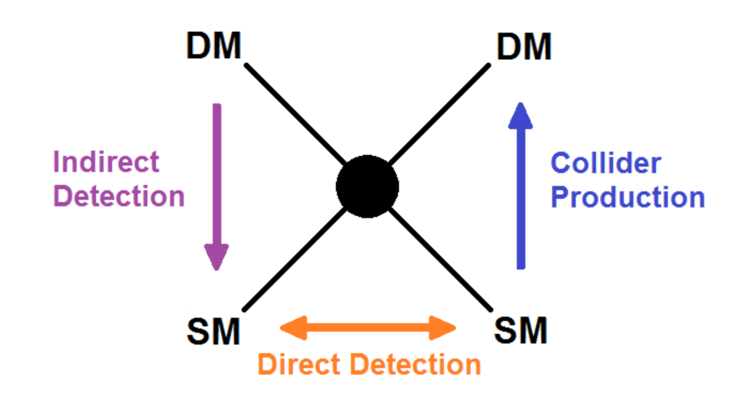
- Apply Machine Learning (ML) to direct detection of dark matter? Long thought difficult due to low statistics.



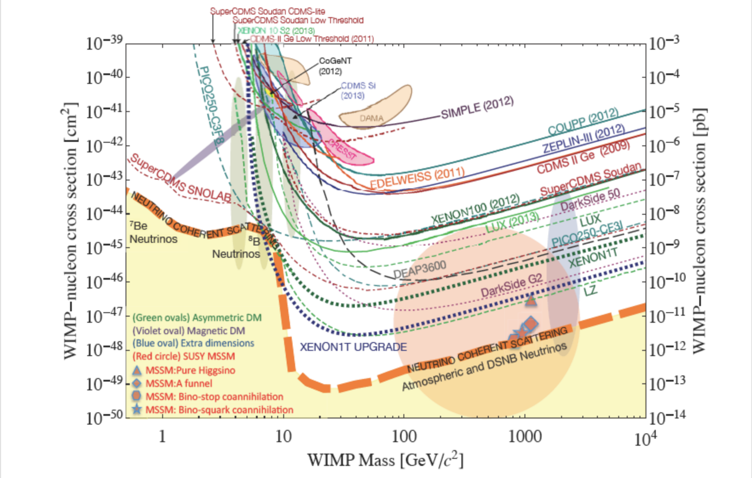


$m_\chi, \sigma,$ Bkg/Sig?

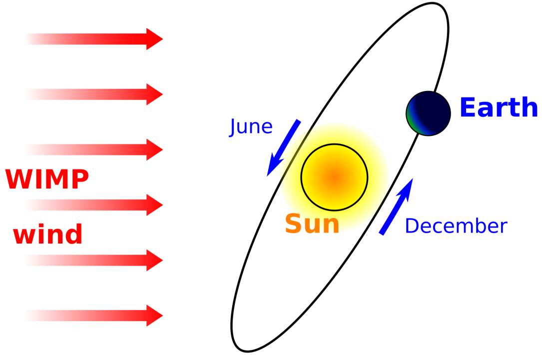
Dark Matter Direct Detection


DM
DM
Indirect
Detection
Collider
Production
SM
SM
Direct Detection

- Current and planned next-generation DD experiments are probing/will probe a very large portion of the parameter space of the WIMP (Weakly Interacting Massive Particle) model.



Direct detection: Schematic



Direct detection: Traditional likelihood-based analysis

After some exposure $\rightarrow$ collect events:

$$\mathcal{L}(s+b) \sim \frac{e^{-\mu_s(\theta)-\mu_b(\theta)}}{n!} \prod_{i=1}^{n} \frac{d\left(N_s+N_b\right)}{dE}\left(E_i \mid \theta\right)$$

Direct detection: Traditional likelihood-based analysis

After some exposure $\rightarrow$ collect events:

$$\mathcal{L}(s+b) \sim \frac{e^{-\mu_s(\theta)-\mu_b(\theta)}}{n!} \prod_{i=1}^{n} \frac{d\left(N_s+N_b\right)}{dE}\left(E_i \mid \theta\right)$$

- Expected number of signal events: $\mu_s = MT \cdot \int dN_s/dE$
- Expected number of background events: $\mu_b = MT \cdot \int dN_b/dE$
- Spectral information

Direct detection: Traditional likelihood-based analysis

After some exposure $\rightarrow$ collect events:

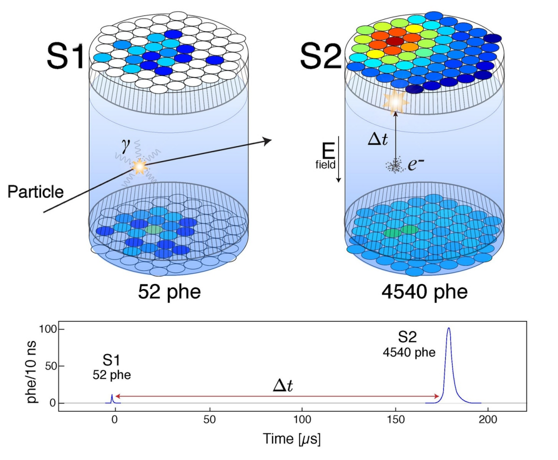
$$\mathcal{L}(s+b) \sim \frac{e^{-\mu_s(\theta)-\mu_b(\theta)}}{n!} \prod_{i=1}^{n} \frac{d\left(N_s + N_b\right)}{dE}\left(E_i \mid \theta\right)$$

- Expected number of signal events: $\mu_s = MT \cdot \int dN_s/dE$
- Expected number of background events: $\mu_b = MT \cdot \int dN_b/dE$
- Spectral information

Model parameters $\theta = \{m_\chi, \sigma ...\}$ phenomenalogically determine two things:

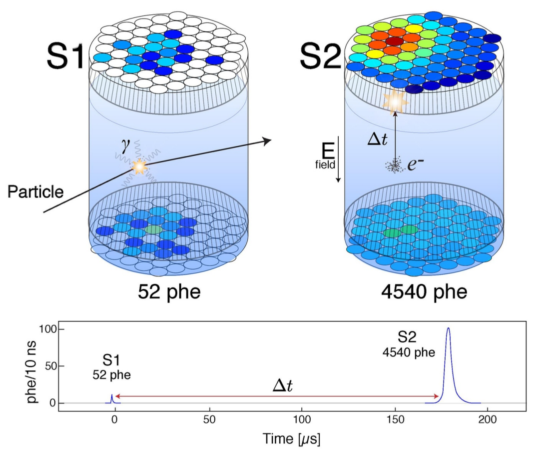
$$\left.\begin{array}{l} \bullet \text{ Number of expected events} \\ \bullet \text{ Signal spectrum `shape'} \end{array}\right\} \text{Important for ML analysis}$$

XENONnT as a test-bed: Events



- S1: Prompt scintillation signal from recoil event.
- S2: Electron charges produced during ionization drift upwards $\rightarrow$ extracted into gaseous phase creating larger scintillation.

XENONnT as a test-bed: Two types of events



- Nuclear Recoil (NR) $\rightarrow$ WIMPs
- (Dominant) Background $\rightarrow$ Electron Recoil (ER).

XENONnT as a test-bed: Training data

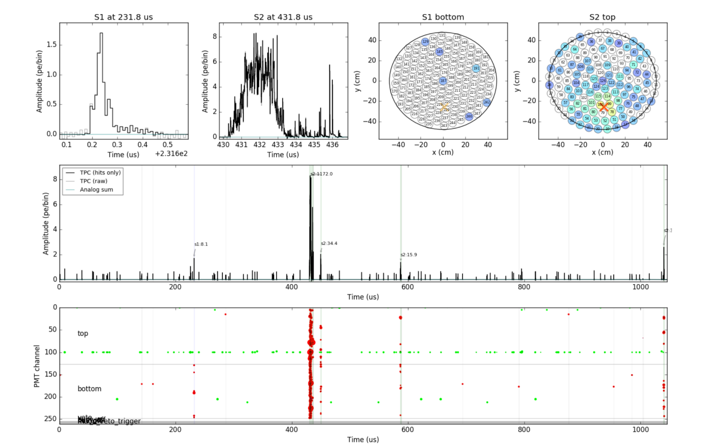


Figure 1: Nuclear recoil (NR) event example image from arxiv:1911.09210.

XENONnT as a test-bed: Training data

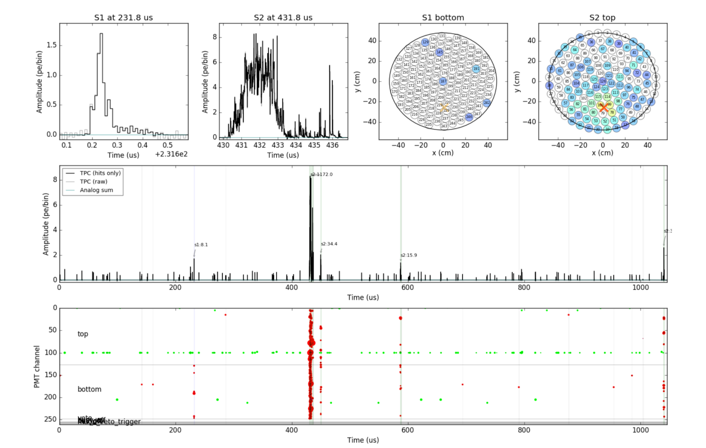


Figure 1: Nuclear recoil (NR) event example image from arxiv:1911.09210.

- Distance and ratio between S1/S2 peaks → NR vs. ER.

XENONnT as a test-bed: Training data

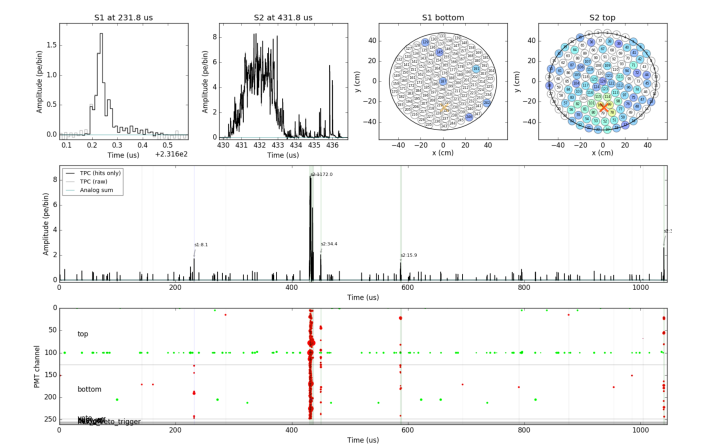
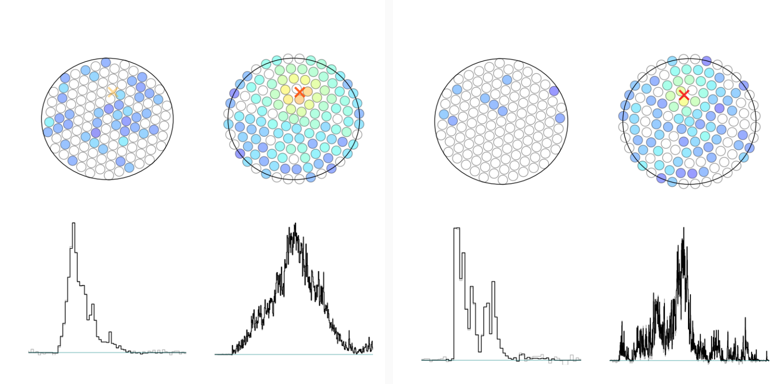


Figure 1: Nuclear recoil (NR) event example image from arxiv:1911.09210.

- Distance and ratio between S1/S2 peaks → NR vs. ER.
- ML can learn this instead!

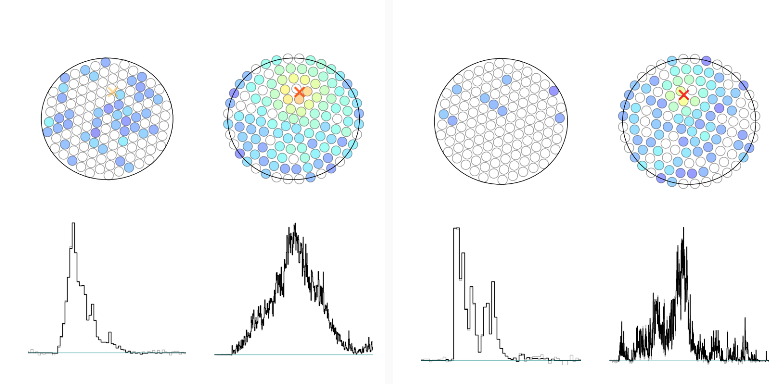
Supervised classification

Classification: Signal vs. Background



- Binary classification: ER background vs. NR signal

Classification: Signal vs. Background



- Binary classification: ER background vs. NR signal
- Sanz et. al `arXiv:1911.09210` found this image configuration optimal. `https://github.com/LucyMars/SearchForDarkMatter`

Classification: Signal vs. Background

Sanz et al. method:

1. Assume fixed WIMP mass 500 GeV and cross-section $\sigma = 10^{-45}$ cm^2 (34.2 live-days)

Classification: Signal vs. Background

Sanz et al. method:

1. Assume fixed WIMP mass 500 GeV and cross-section $\sigma = 10^{-45}$ cm^2 (34.2 live-days)
2. Use `Laidbax` (Likelihood- And Interpolated Density Based Analysis for XENON) in conjunction with `Pax` to generate graphical output of the simulated WIMP and background events

Classification: Signal vs. Background

Sanz et al. method:

1. Assume fixed WIMP mass 500 GeV and cross-section $\sigma = 10^{-45}$ cm^2 (34.2 live-days)
2. Use `Laidbax` (Likelihood- And Interpolated Density Based Analysis for XENON) in conjunction with `Pax` to generate graphical output of the simulated WIMP and background events
3. Train a **Convolutional Neural Network** (CNN) to classify WIMP vs. background

Classification: Signal vs. Background

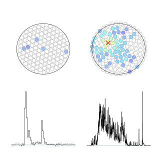
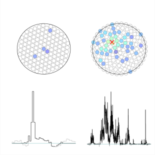
Sanz et al. method:

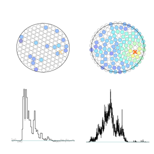
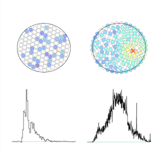
1. Assume fixed WIMP mass 500 GeV and cross-section $\sigma = 10^{-45}$ cm^2 (34.2 live-days)
2. Use `Laidbax` (Likelihood- And Interpolated Density Based Analysis for XENON) in conjunction with `Pax` to generate graphical output of the simulated WIMP and background events
3. Train a **Convolutional Neural Network** (CNN) to classify WIMP vs. background

Discovered this can work with ensemble of WIMP masses.
Cross-section irrelevant for event-by-event bkg/signal classification.

Classification: Training data generation

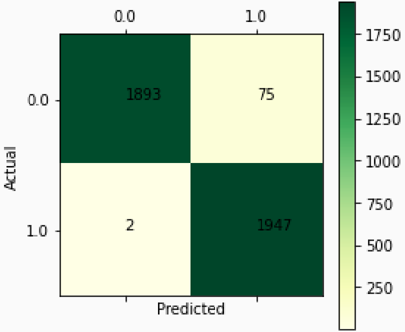
Reproducing results: generate image training set with 4×10^4 images total.
"Mixed bag" of assumed masses and cross-sections:

$\mathbf{x}_{\mathrm{NR}}^{\mathrm{Train}} = \Bigg($  ,  , $\dots \Bigg)$

$\mathbf{x}_{\mathrm{ER}}^{\mathrm{Train}} = \Bigg($  ,  , $\dots \Bigg)$

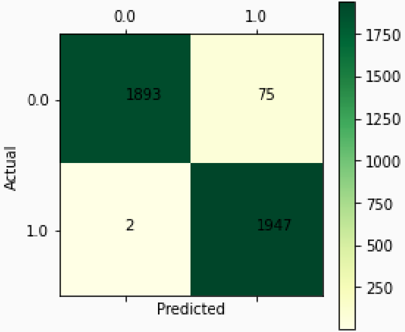
Classification: Signal vs. Background Results

- Train on $\sim$ 40000 images. Take testing sub-sample of $\sim$ 40%



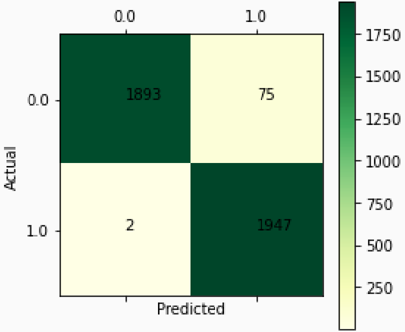
Classification: Signal vs. Background Results

- Train on $\sim$ 40000 images. Take testing sub-sample of $\sim$ 40%
- Check performance $\rightarrow$ confusion matrix:



Classification: Signal vs. Background Results

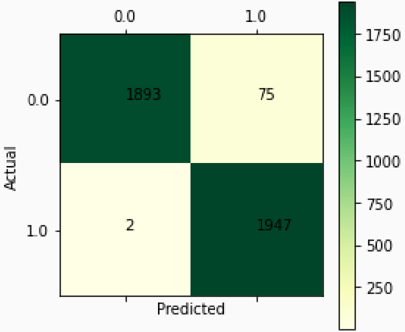
- Train on $\sim$ 40000 images. Take testing sub-sample of $\sim$ 40%
- Check performance $\rightarrow$ confusion matrix:



- Takeaway $\Rightarrow$ **98.03% accuracy**. (Recall = 98.07%, Precision = 96.39%)

Classification: Signal vs. Background Results

- Train on $\sim$ 40000 images. Take testing sub-sample of $\sim$ 40%
- Check performance $\rightarrow$ confusion matrix:

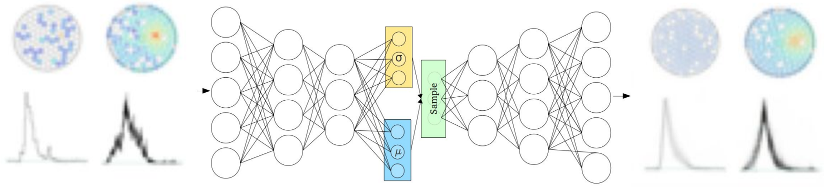


- Takeaway $\Rightarrow$ **98.03% accuracy**. (Recall = 98.07%, Precision = 96.39%)
- **This works regardless mass and cross-section: NR/ER are what matter.**

Unsupervised approach

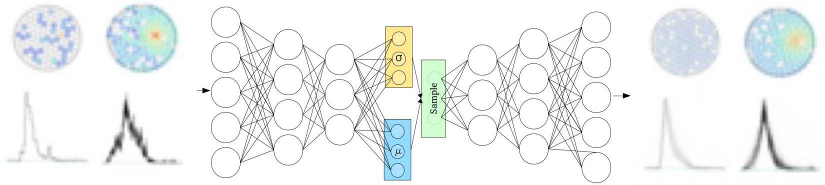
Generative Deep Learning: The Variational Auto-Encoder

- Goal: Learn low dimensional representation (encoding) of data via dimentionality reduction.



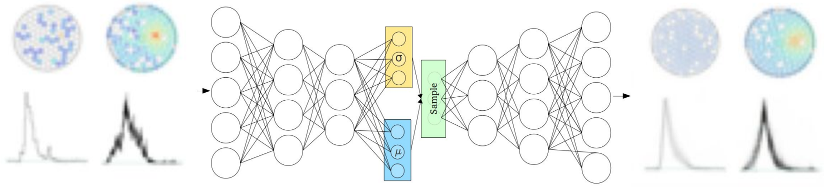
Generative Deep Learning: The Variational Auto-Encoder

- Goal: Learn low dimensional representation (encoding) of data via dimentionality reduction.
- Latent space (bottleneck) layer replaced with a bunch of normal distributions parameterized by some $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.



Generative Deep Learning: The Variational Auto-Encoder

- Goal: Learn low dimensional representation (encoding) of data via dimentionality reduction.
- Latent space (bottleneck) layer replaced with a bunch of normal distributions parameterized by some $\boldsymbol{\mu}$ and $\boldsymbol{\sigma}$.
- Our goal: Learn the latent representation of the background (ER) events.



Variational-Auto-Encoder: Training

- Use same data as with supervised CNN.

Variational-Auto-Encoder: Training

- Use same data as with supervised CNN.
- Train CVAE on just* ER background data.

Variational-Auto-Encoder: Training

- Use same data as with supervised CNN.
- Train CVAE on just* ER background data.
- Train by maximising evidence lower bound (ELBO):

$$
\begin{aligned}
\log p(x) \geq \mathrm{ELBO} &= \mathbb{E}_{q(z|x)}\left[\log \frac{p(x,z)}{q(z \mid x)}\right] \\
&\simeq \log p(x \mid z) - D_{KL}(\log q(z \mid x) || \log p(z)) \\
&= \sum_{i=1}^{\text{Batch}} x_i \log y_i + (1-x_i)\cdot \log(1-y_i) - \beta \sum_{j=0}^{K}\left[\sigma_j^2 + \mu_j^2 - \log(\sigma_j) - 1\right]
\end{aligned}
$$

K = number of latent space normal distributions.

x = Input .

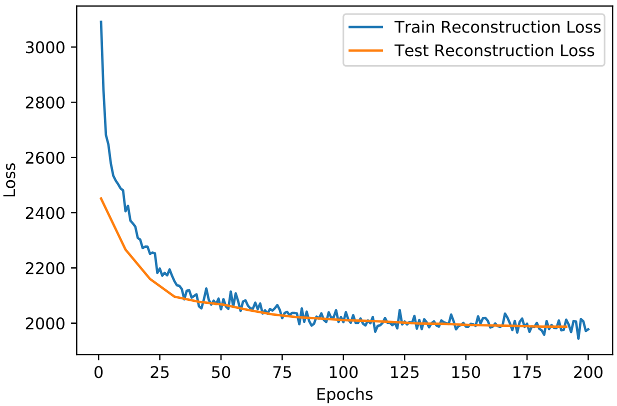
y = Output.

z = Latent vector .

β = Regularization parameter.

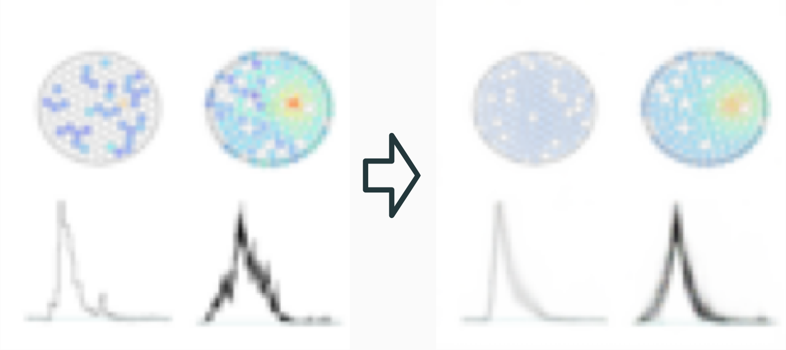
CVAE: Training

- Train the network for 200 epochs.



CVAE: Training

- Once trained, the CVAE should produce generative examples of ER images.



Anomaly detection

- **Anomaly Detection**: Once trained, run data the network has never seen before through trained network.

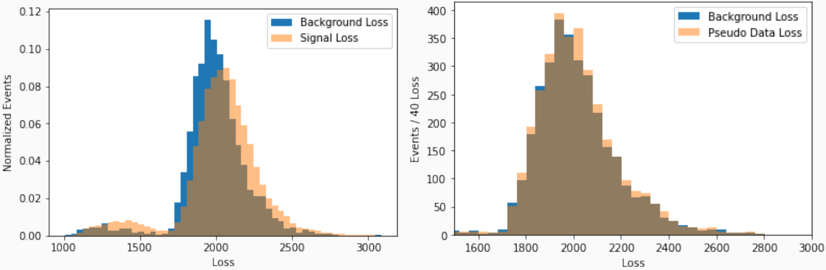
Anomaly detection

- **Anomaly Detection**: Once trained, run data the network has never seen before through trained network.
- If CVAE has learned the underlying properties of ER bkg events, any **non-background** events will in general have **higher loss**.

Anomaly detection

- **Anomaly Detection**: Once trained, run data the network has never seen before through trained network.
- If CVAE has learned the underlying properties of ER bkg events, any **non-background** events will in general have **higher loss**.
- Loss distribution of anomalous data (new physics) will show as an **excess** over background only loss distribution...

Variational-Auto-Encoder: Results



- Left: Background loss distribution + just* signal loss distribution.
- Right: Inject signal into background signal, run whole data set through network.
- Any* anomalous signal will show up as statistical deviation in (pseudo)data loss vs. (known) background loss.

Semi-unsupervised anomaly detection: New distance metric

Cool. But...

- A bit rubbish: Can we get greater separation (anomaly awareness) between these distributions?

Semi-unsupervised anomaly detection: New distance metric

Cool. But...

- A bit rubbish: Can we get greater separation (anomaly awareness) between these distributions?
- New 'anomaly function' that utilizes pre-trained supervised CNN classifier:

Semi-unsupervised anomaly detection: New distance metric

Cool. But...

- A bit rubbish: Can we get greater separation (anomaly awareness) between these distributions?
- New 'anomaly function' that utilizes pre-trained supervised CNN classifier:

$$TS = \underbrace{-ELBO}_{\text{CVAE loss}} + R\,H_B \ ,$$

where

- $H_B = -\frac{1}{N}\sum_{i=0}^{N} \log\left(1 - p\left(x_i\right)\right)$ (Binary cross-sentropy.)

Semi-unsupervised anomaly detection: New distance metric

Cool. But...

- A bit rubbish: Can we get greater separation (anomaly awareness) between these distributions?
- New ‘anomaly function’ that utilizes pre-trained supervised CNN classifier:

$$TS = \underbrace{-ELBO}_{\text{CVAE loss}} + R\,H_B\ ,$$

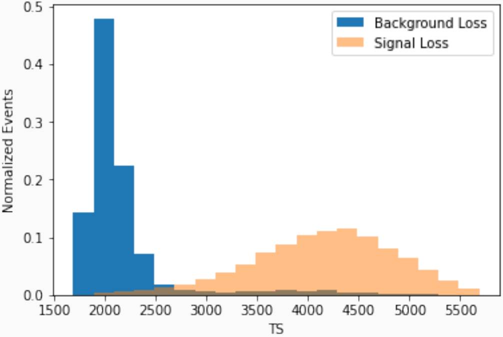
where

- $H_B = -\frac{1}{N}\sum_{i=0}^{N} \log\left(1 - p\left(x_i\right)\right)$ (Binary cross-sentropy.)
- R scales the contribution of the cross-entropy term $\rightarrow$ makes it more/less supervised.

Semi-unsupervised anomaly detection: New distance metric

$$TS = (-ELBO) + R\, H_B \ ,$$

⇒ Semi-unsupervised. Much greater anomaly awareness!

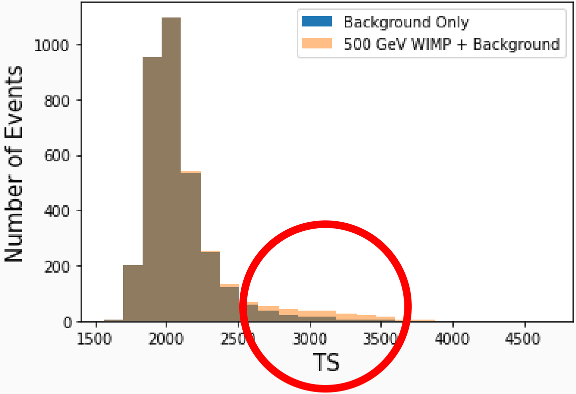


Semi-unsupervised anomaly detection: Results

Calculate $p-$value for reject background-only hypothesis:

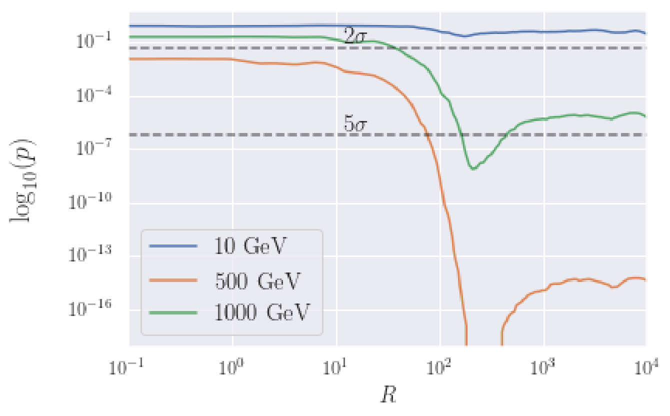
$$\chi_p^2 = \sum_{\text{bins}} \frac{(TS_{\text{ER}} - TS_{\text{ER+NR}})^2}{TS_{\text{ER}} + TS_{\text{ER+NR}}} \,,$$

$$p = 1 - \text{CDF}(\chi^2_{1\text{d.o.f}}) \,.$$



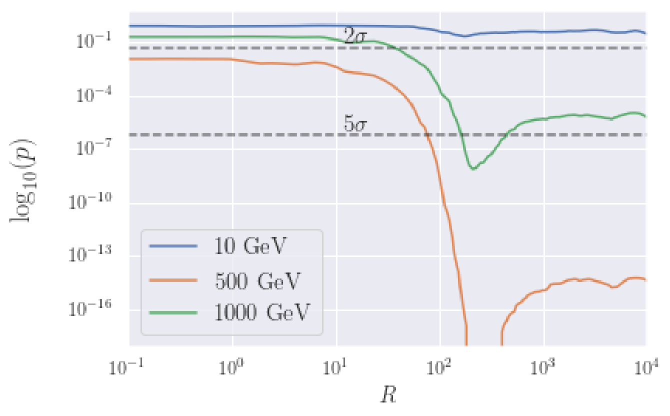
Semi-unsupervised anomaly detection: Effect of R

- Explore effect of the R parameter.



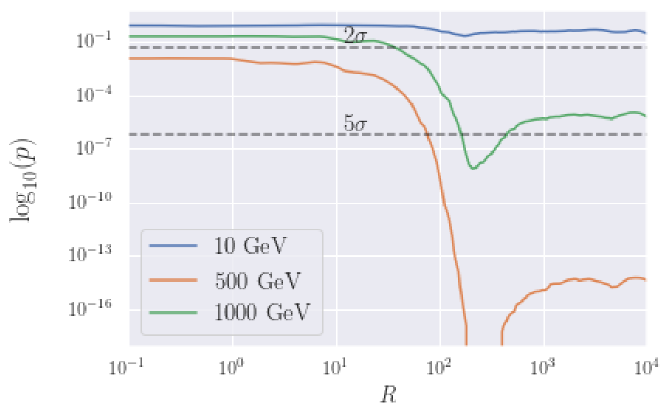
Semi-unsupervised anomaly detection: Effect of R

- Explore effect of the R parameter.
- Three mock data sets corresponding to 10, 500 and 1000 GeV at fixed $\sigma = 10^{-45}\text{cm}^2$, 5 $t\cdot$yr exposure.



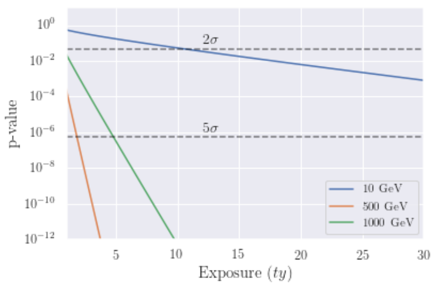
Semi-unsupervised anomaly detection: Effect of R

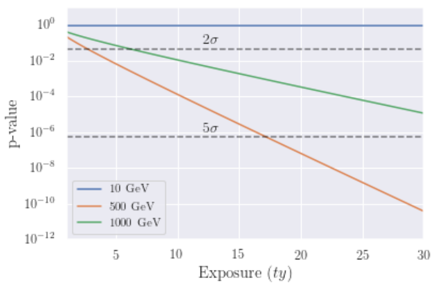
- Explore effect of the R parameter.
- Three mock data sets corresponding to 10, 500 and 1000 GeV at fixed $\sigma = 10^{-45}\text{cm}^2$, 5 $t{\cdot}$yr exposure.
- Best result for $R \sim 170$, but generally free to choose!



Semi-unsupervised anomaly detection: Forecasting

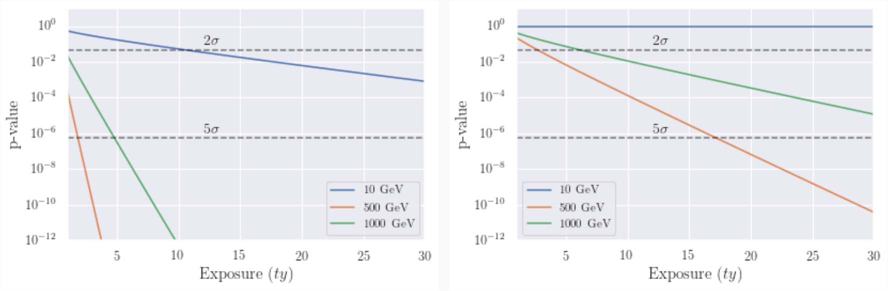
- Get anomaly sensitivity as function of exposure:





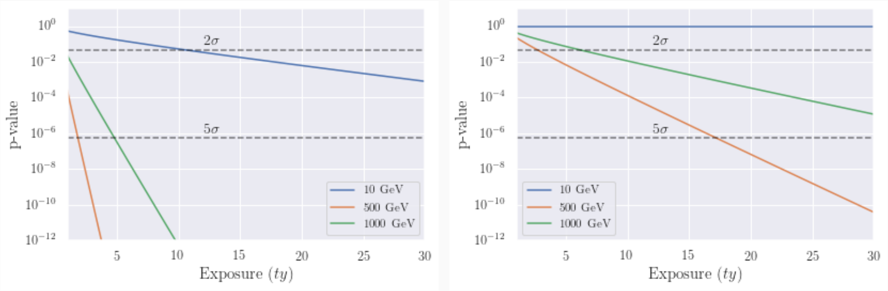
Semi-unsupervised anomaly detection: Forecasting

- Get anomaly sensitivity as function of exposure:
- $\sigma = 10^{-45}$ cm^2 (Left) and $\sigma = 10^{-46}$ cm^2 (Right). [$R = 170$].

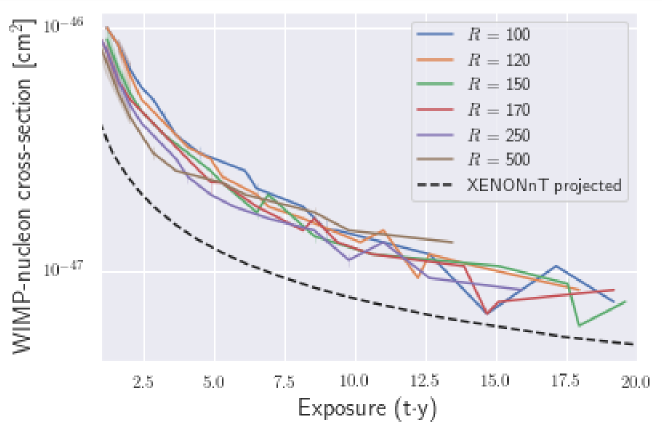


Semi-unsupervised anomaly detection: Forecasting

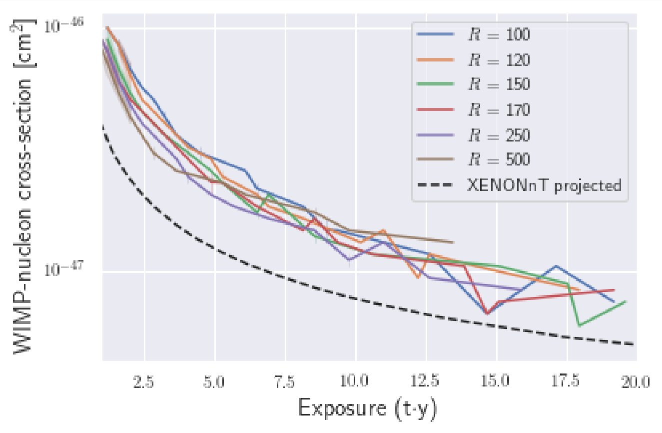
- Get anomaly sensitivity as function of exposure:
- $\sigma = 10^{-45}$ cm^2 (Left) and $\sigma = 10^{-46}$ cm^2 (Right). [$R = 170$].
- Compare to collaboration?



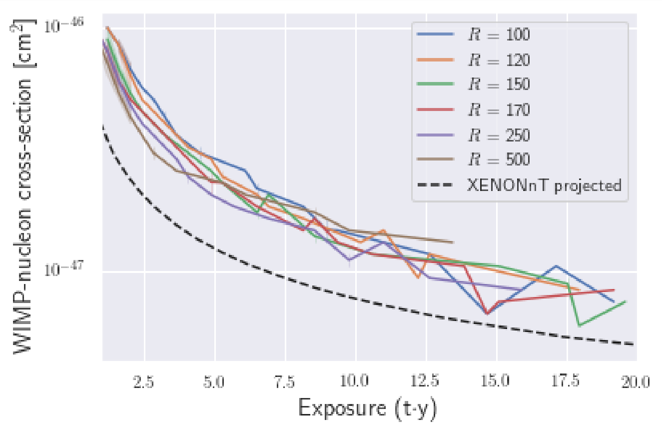
- Official XENONnT 5σ discover sensitivity for 50 GeV WIMP.



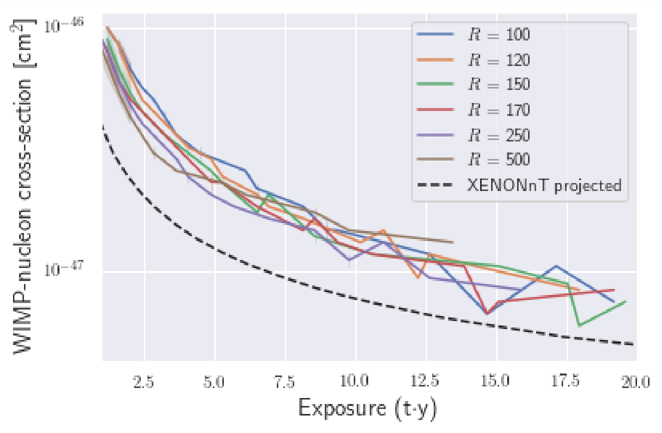
- Official XENONnT 5σ discover sensitivity for 50 GeV WIMP.
- Do for a variety of R values.



- Official XENONnT 5σ discover sensitivity for 50 GeV WIMP.
- Do for a variety of R values.
- Collaboration has very sophisticated methods (auxiliary data etc.)



- Official XENONnT 5σ discover sensitivity for 50 GeV WIMP.
- Do for a variety of R values.
- Collaboration has very sophisticated methods (auxiliary data etc.)
- Not exactly* comparing apples with apples...



Quick summary

- Can use ML methods to supplement traditional likelihood based statistical techniques in dark matter searches.

Quick summary

- Can use ML methods to supplement traditional likelihood based statistical techniques in dark matter searches.
- Depending on the experimental premise, can utilize semi-supervised methods to amplify the analysis power.

Quick summary

- Can use ML methods to supplement traditional likelihood based statistical techniques in dark matter searches.
- Depending on the experimental premise, can utilize semi-supervised methods to amplify the analysis power.
- More work needed to explore the range of optimal anomaly functions to use for anomaly detection.

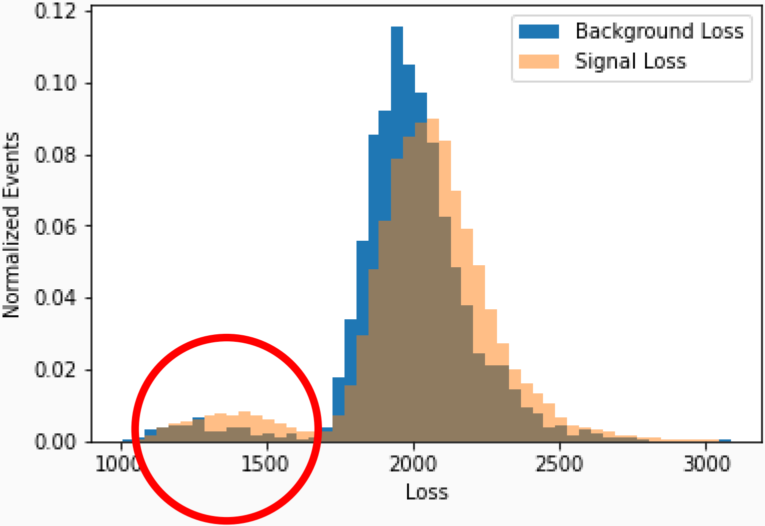
Quick summary

- Can use ML methods to supplement traditional likelihood based statistical techniques in dark matter searches.
- Depending on the experimental premise, can utilize semi-supervised methods to amplify the analysis power.
- More work needed to explore the range of optimal anomaly functions to use for anomaly detection.
- Continue work to use SBI (flow based methods) for parametric regression and posterior estimation.

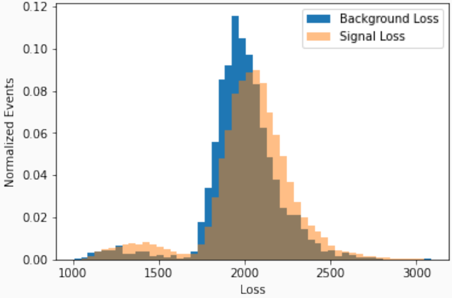
Thanks for listening!

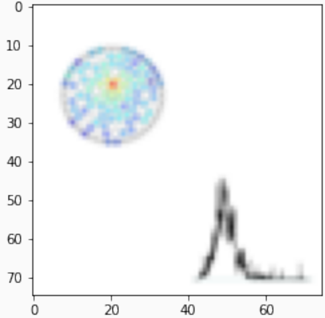
Backup slides

Variational-Auto-Encoder: Interesting discovery...



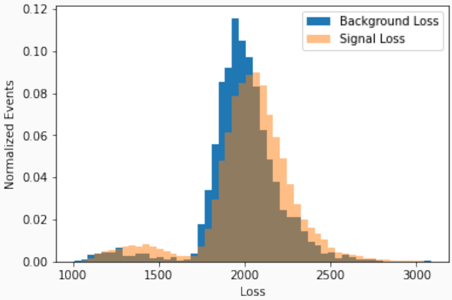
Variational-Auto-Encoder: Interesting discovery...

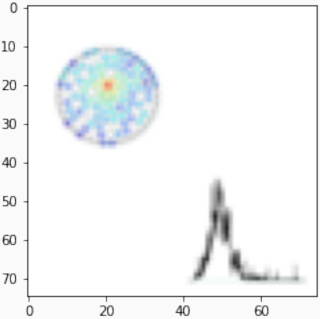




- Actually events with no S1 (S2 only)

Variational-Auto-Encoder: Interesting discovery...





- Actually events with no S1 (S2 only)
- S2-only analyses been done before. Could extract more power from such an analysis with ML?

Direct detection: Traditional likelihood-based analysis

After some exposure → collect events:

$$\mathcal{L}(s+b) \sim \frac{e^{-\mu_s(\theta)-\mu_b(\theta)}}{n!} \prod_{i=1}^{n} \frac{d\left(N_s + N_b\right)}{dE}\left(E_i \mid \theta\right)$$

Direct detection: Traditional likelihood-based analysis

After some exposure $\rightarrow$ collect events:

$$\mathcal{L}(s+b) \sim \frac{e^{-\mu_s(\theta)-\mu_b(\theta)}}{n!} \prod_{i=1}^{n} \frac{d\left(N_s+N_b\right)}{dE}\left(E_i \mid \theta\right)$$

- Expected number of signal events: $\mu_s = MT \cdot \int dN_s/dE$
- Expected number of background events: $\mu_b = MT \cdot \int dN_b/dE$
- Spectral information

Direct detection: Traditional likelihood-based analysis

After some exposure $\rightarrow$ collect events:

$$\mathcal{L}(s+b) \sim \frac{e^{-\mu_s(\theta)-\mu_b(\theta)}}{n!} \prod_{i=1}^{n} \frac{d\left(N_s+N_b\right)}{dE}\left(E_i \mid \theta\right)$$

- Expected number of signal events: $\mu_s = MT \cdot \int dN_s/dE$
- Expected number of background events: $\mu_b = MT \cdot \int dN_b/dE$
- Spectral information

Model parameters $\theta = \{m_\chi, \sigma ...\}$ phenomenalogically determine two things:

$$\left.\begin{array}{l}\bullet \text{ Number of expected events} \\ \bullet \text{ Signal spectrum `shape'}\end{array}\right\} \text{Important for ML analysis}$$