iDark
The intelligent dark matter survey

Radboud University

netherlands eScience center

Dark Machines

LUC HENDRIKS
RADBOUD UNIVERSITY, NIJMEGEN (NL)

WHAT IS DARKMACHINES?

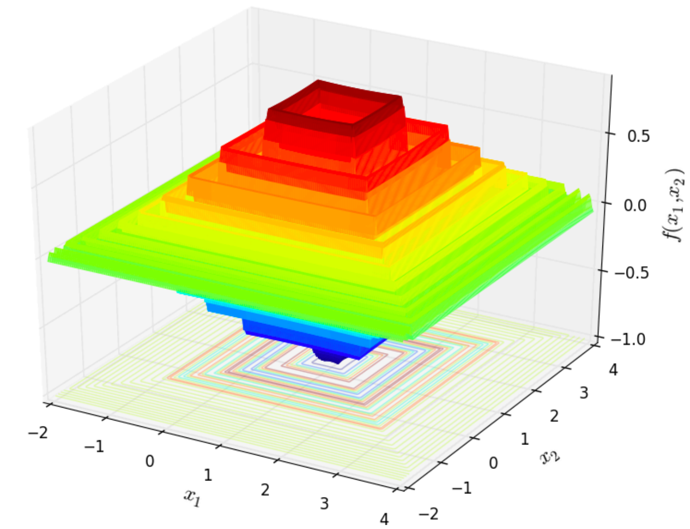
- Research collective of about 200 researchers
- ML & DM experts combining knowledge to solve hard problems
    - Multidisciplinary: eg. ML experts joined from biomedical imaging
    - Challenge based (DM experts deliver data, ML experts deliver solution)
    - Challenges are self-organised by challenge leaders
    - Each challenge produces 1+ papers
    - Anyone can join if interested

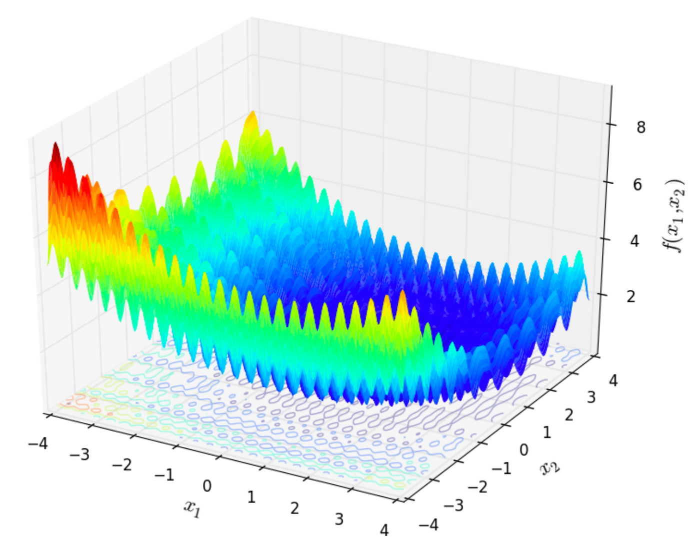
WHAT IS DARKMACHINES?

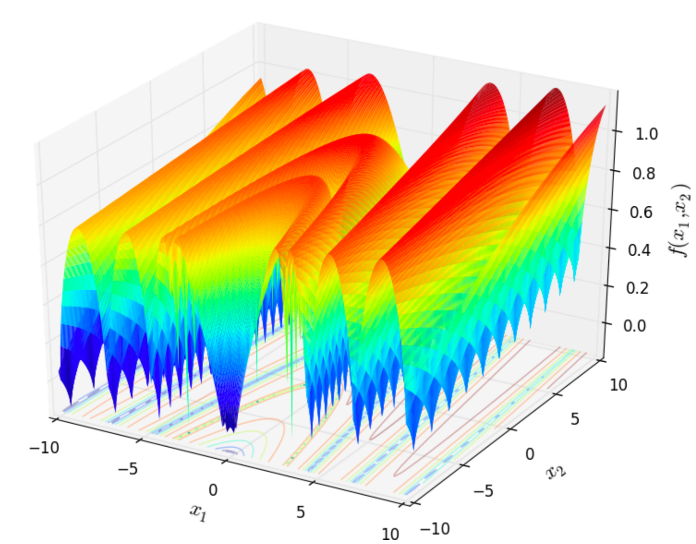
- Research collective of about 200 researchers
- ML & DM experts combining knowledge to solve hard problems
    - Multidisciplinary: eg. ML experts from medical sciences
    - Challenge based (DM experts deliver data, ML experts deliver solution)
    - Challenges are self-organised by challenge leaders
    - Each challenge produces 1+ papers
    - Anyone can join if interested
- Regular meetings
    - Monthly virtual meeting
    - Yearly in-person meeting (2018: Leiden, 2019: Trieste, 2020: CERN)

CHALLENGE: HIGH-D PARAMETER OPTIMISATION

- Goal: find best algorithm to find optima in (very complex) high-D parameter spaces
- Very typical problem in all of physics



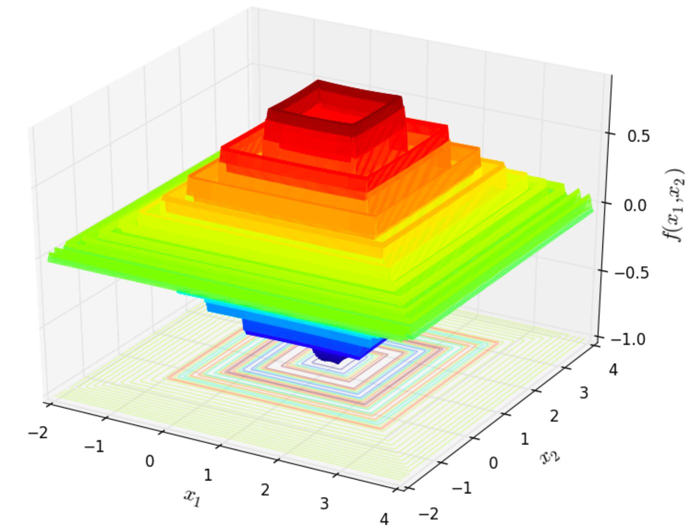


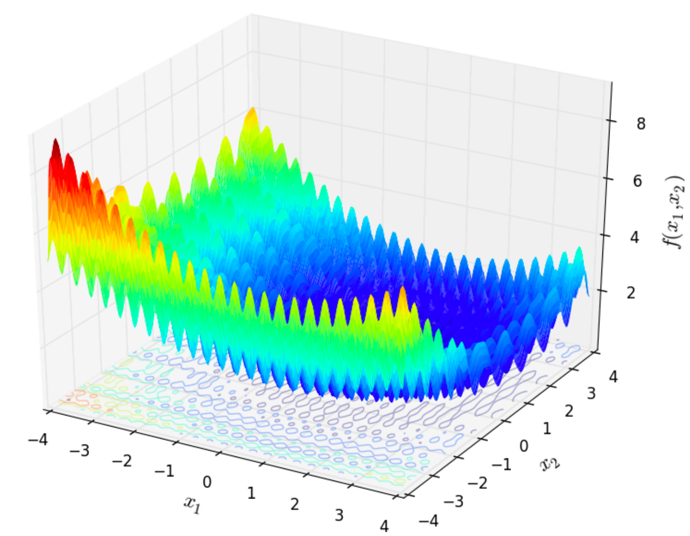


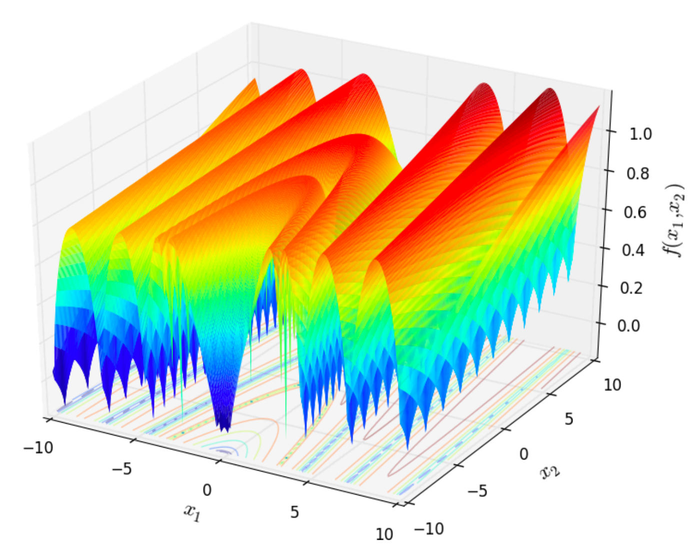
http://infinity77.net/global_optimization

CHALLENGE: HIGH-D PARAMETER OPTIMISATION

- Goal: find best algorithm to find optima in (very complex) high-D parameter spaces
- Very typical problem in all of physics
- Benchmark framework built with multiple test functions
- Everyone interested benchmarks their favourite scanner (eg MultiNest, ScannerBit, MCMC, HMC, …)
- Publish review article with comparison





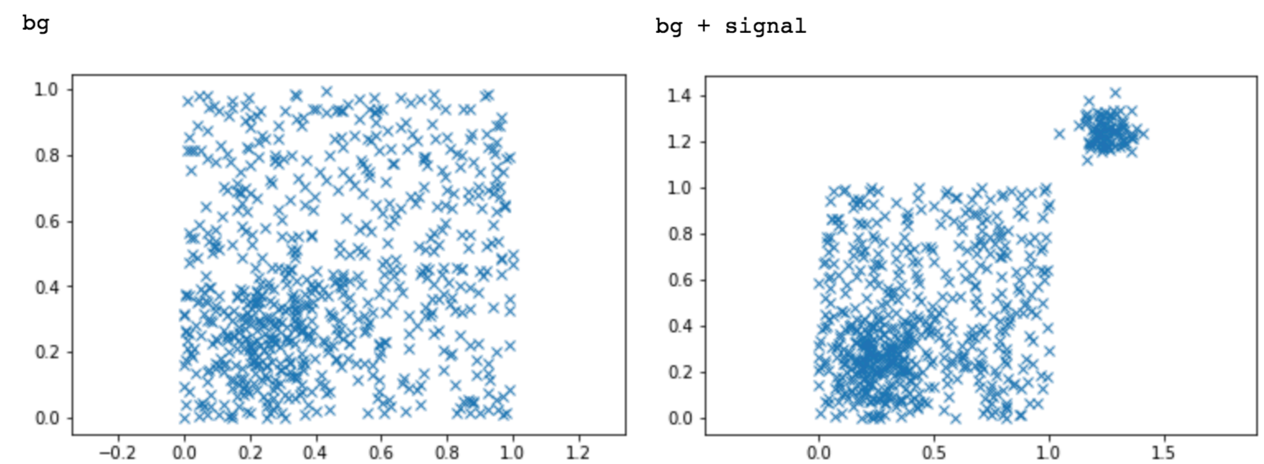


http://infinity77.net/global_optimization

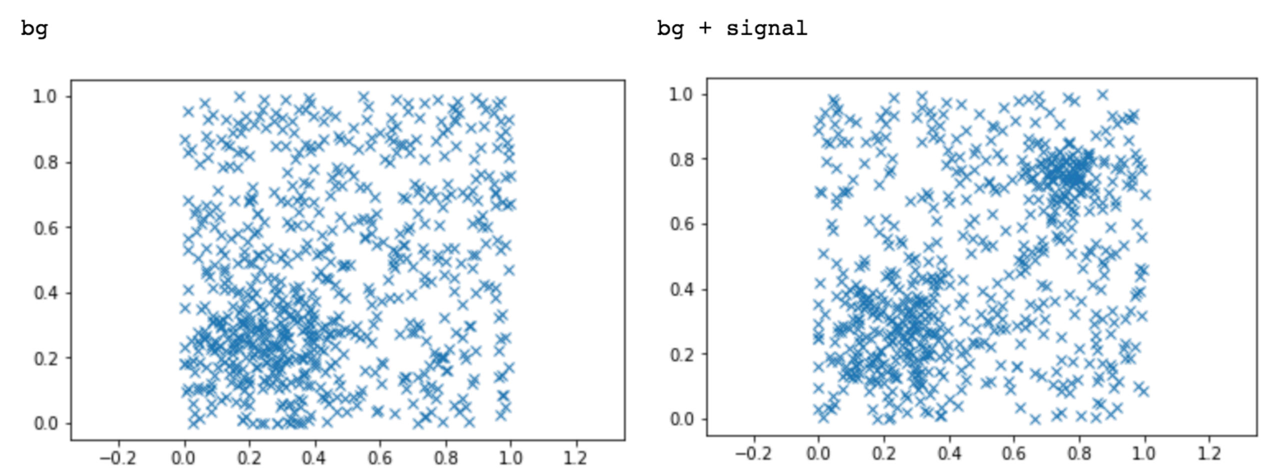
CHALLENGE: UNSUPERVISED COLLIDER SEARCHES

- Goal: find method to find new physics at colliders

Outlier detection



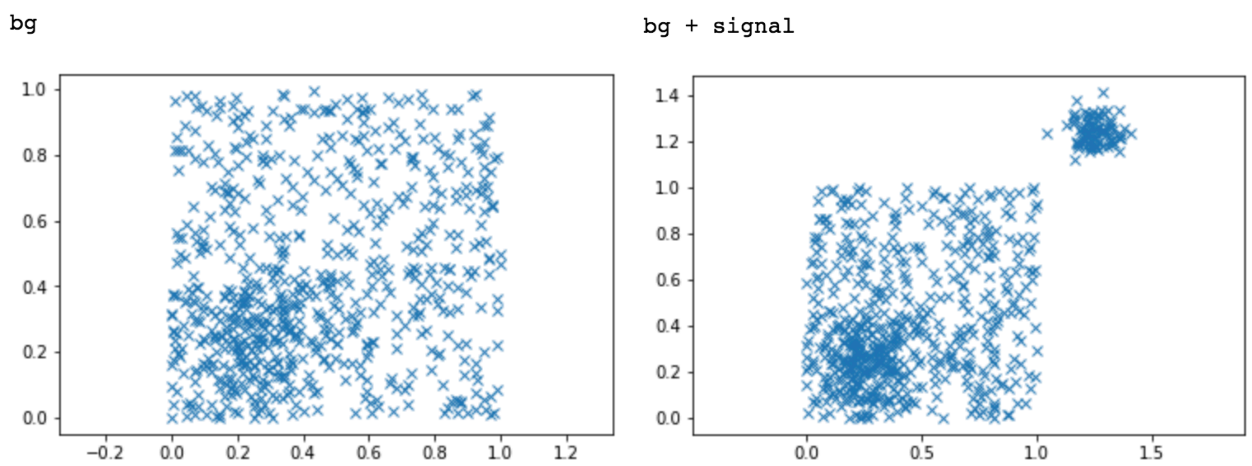
Density estimation



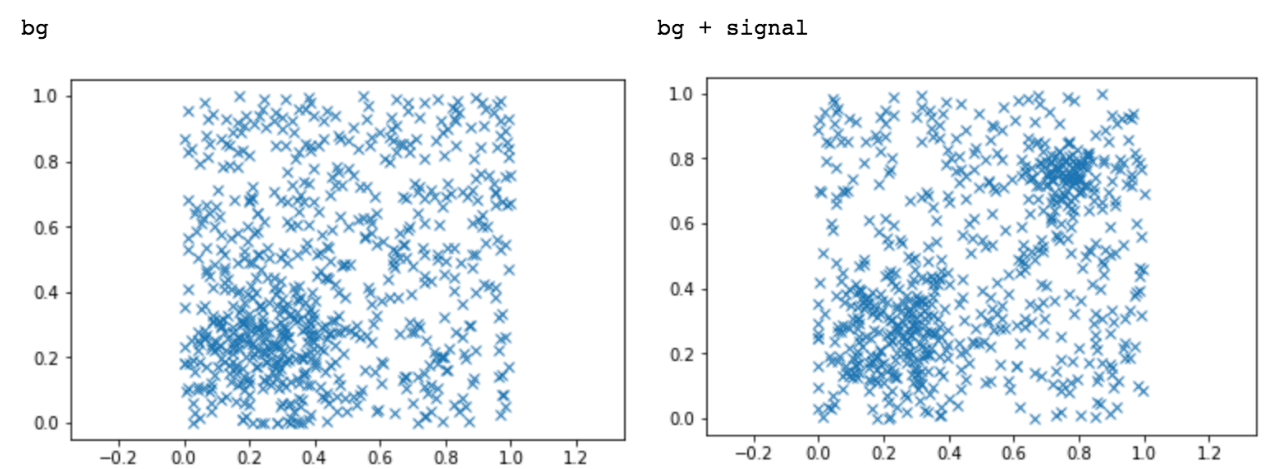
CHALLENGE: UNSUPERVISED COLLIDER SEARCHES

- Goal: find method to find new physics at colliders
- Standardised dataset published soon
- Use different ML methods for outlier detection
- Publish article with the proposed methods and their performance

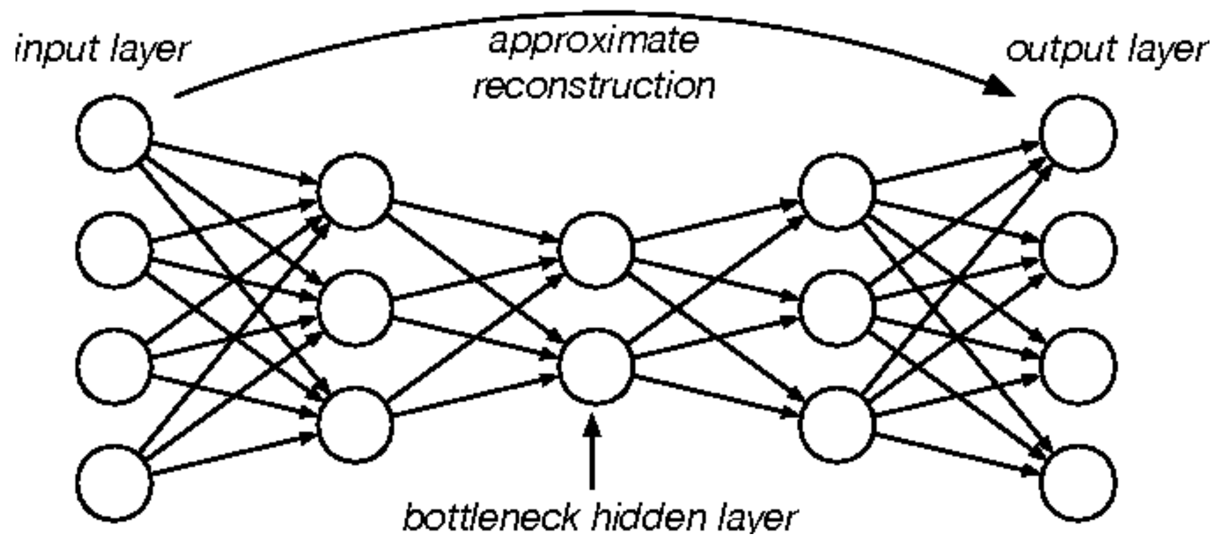
Outlier detection



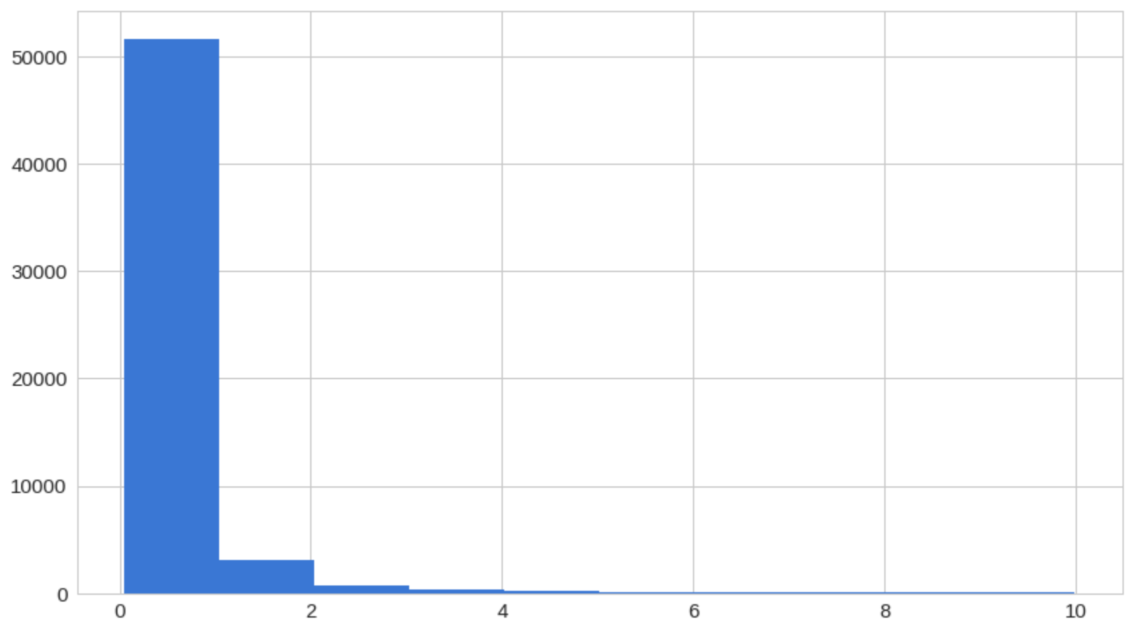
Density estimation



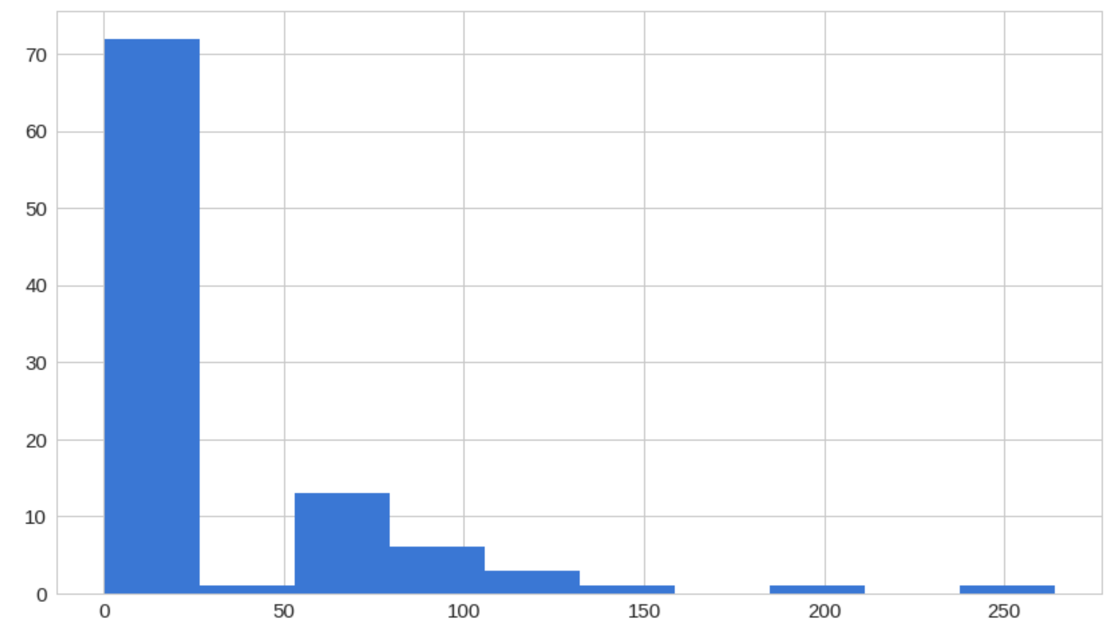
- Example: credit card fraud detection with autoencoder



No fraud



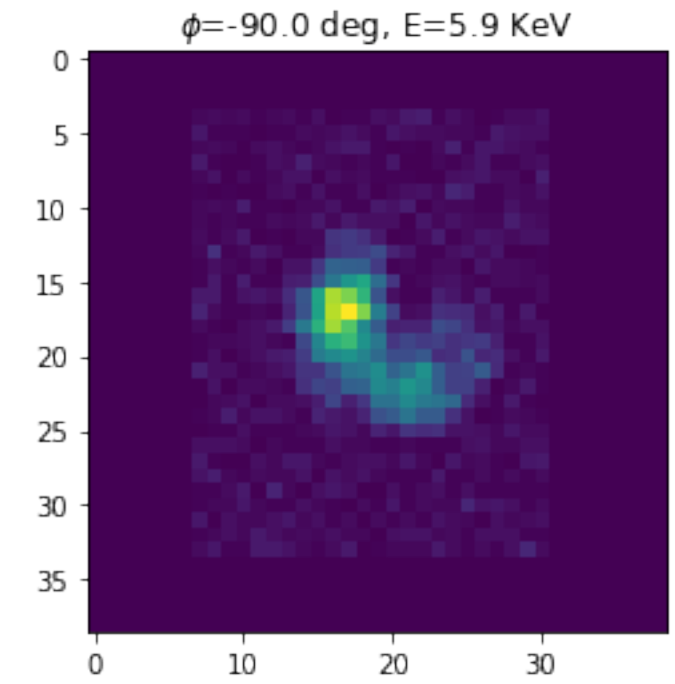
Fraud

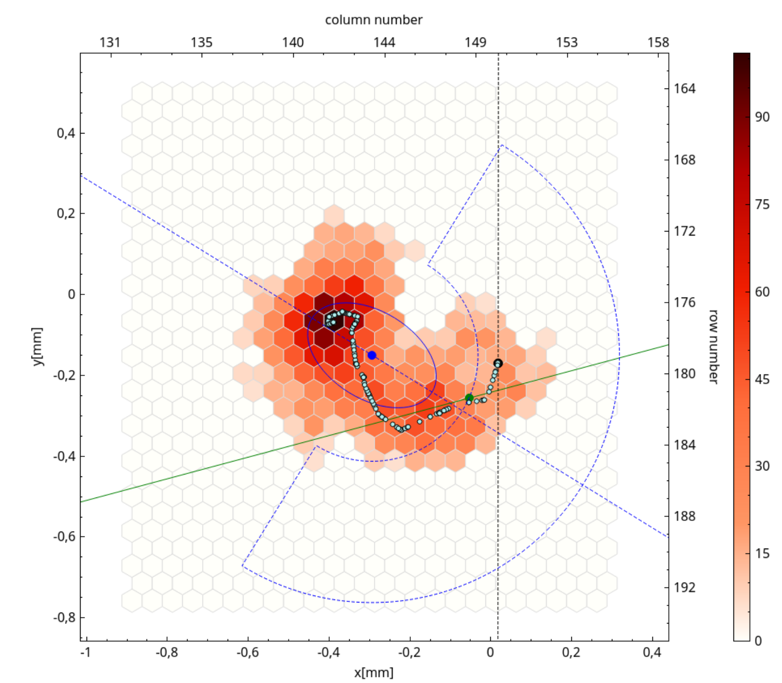


https://medium.com/@curiousily/credit-card-fraud-detection-using-autoencoders-in-keras-tensorflow-for-hackers-part-vii-20e0c85301bd

CHALLENGE: PARTICLE TRACK RECONSTRUCTION

- Particle hits detector
- Reconstruct the track of the particle
- Typical problem for many detectors
- Challenge mainly revolves around IXPE (Imaging X-ray Polarimetry Explorer, launched end 2020)



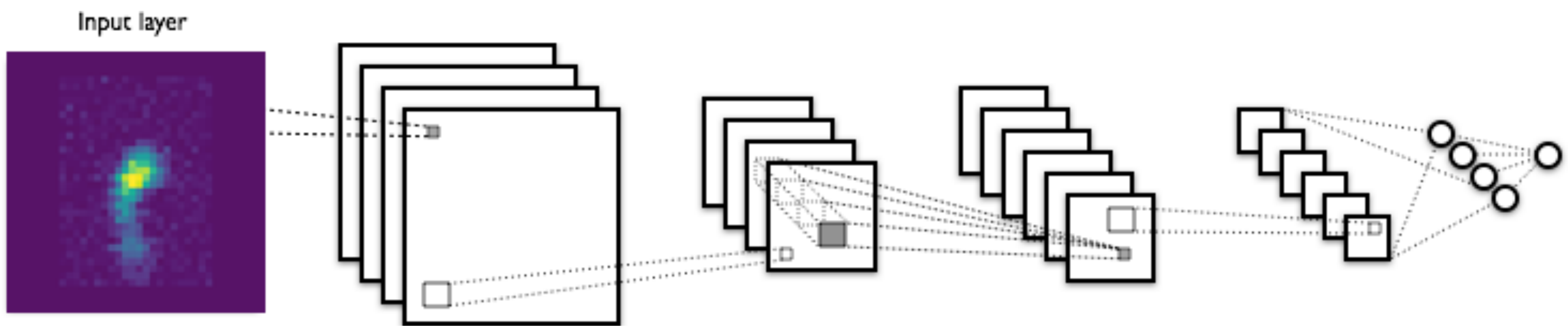


Images courtesy of Michela Negro

CHALLENGE: PARTICLE TRACK RECONSTRUCTION

- Example: CNN-based angle detection
- Difficulty: circular boundary conditions in the loss function

```
def mse_cartesian(inputs, outputs):
    x_in = tf.cos(inputs)
    y_in = tf.sin(inputs)
    x_out = tf.cos(outputs)
    y_out = tf.sin(outputs)
    dist = tf.square(x_in - x_out) + tf.square(y_in -
y_out)
```

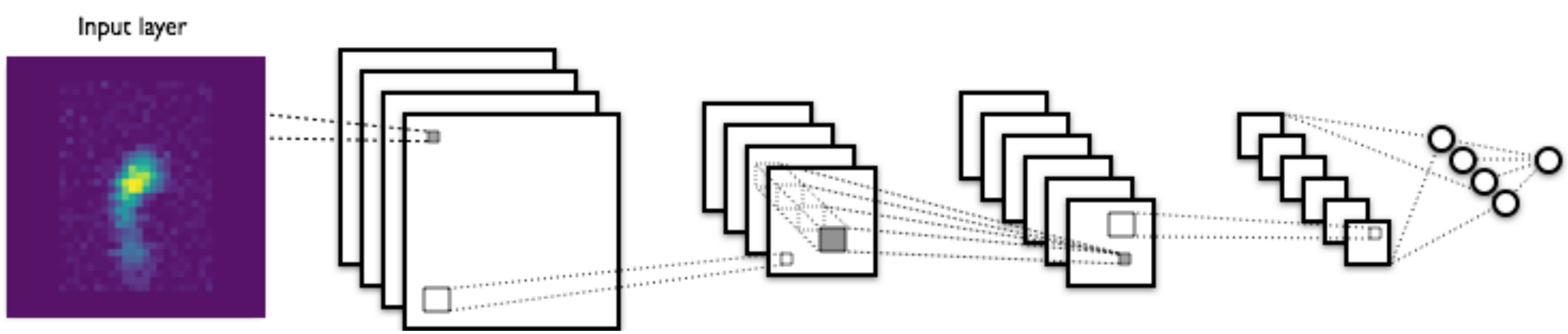


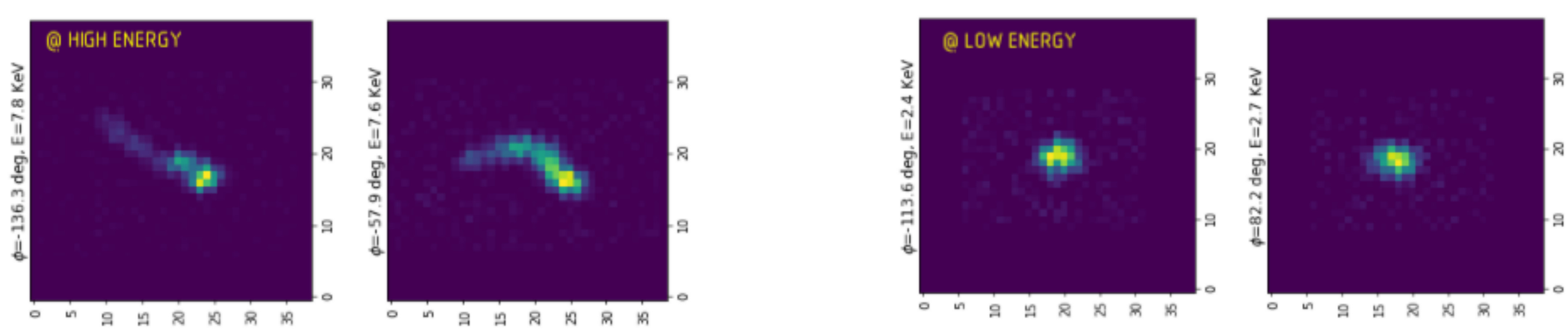
CHALLENGE: PARTICLE TRACK RECONSTRUCTION

- Example: CNN-based angle detection

- Difficulty: circular boundary conditions in the loss function

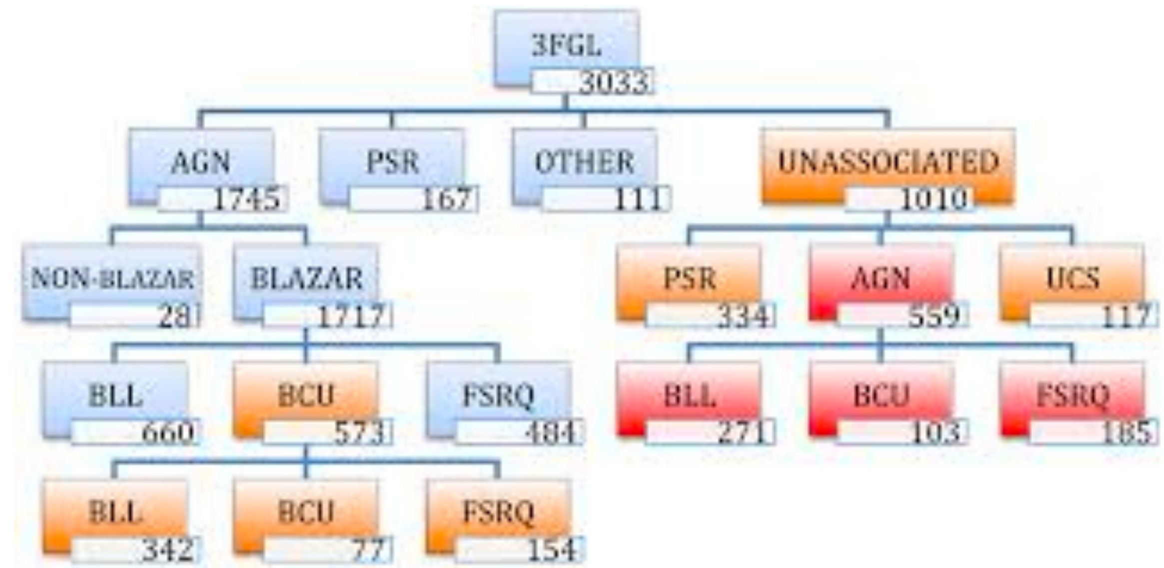
```
def mse_cartesian(inputs, outputs):
    x_in = tf.cos(inputs)
    y_in = tf.sin(inputs)
    x_out = tf.cos(outputs)
    y_out = tf.sin(outputs)
    dist = tf.square(x_in - x_out) + tf.square(y_in -
y_out)
```

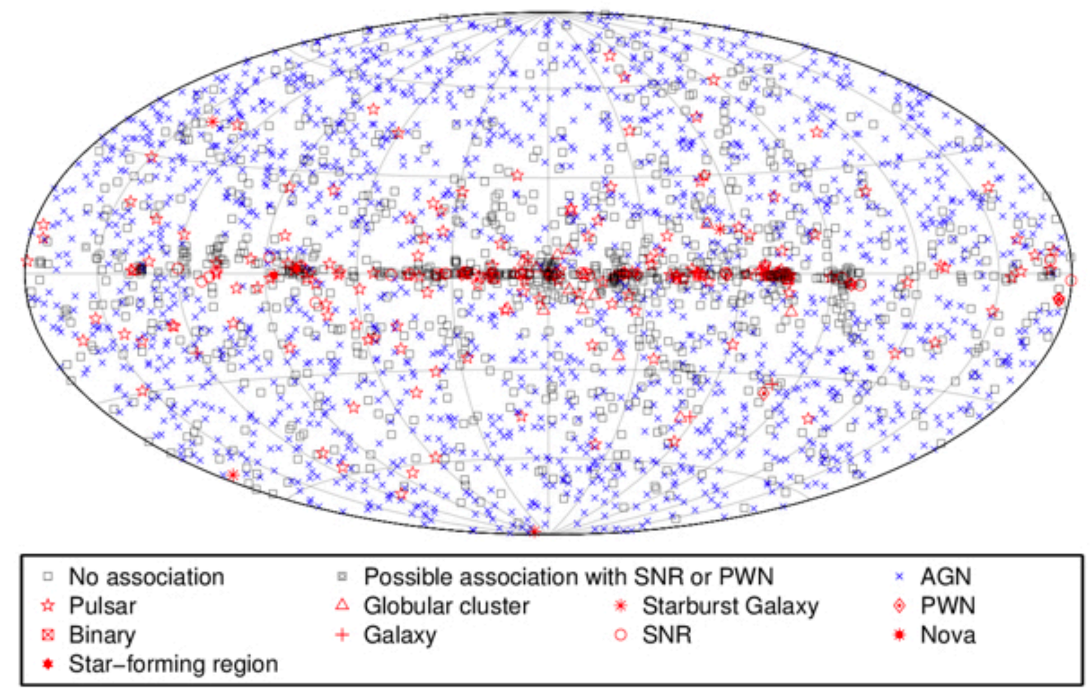




CHALLENGE: INCLUSIVE ANALYSIS OF FERMI POINT SOURCES

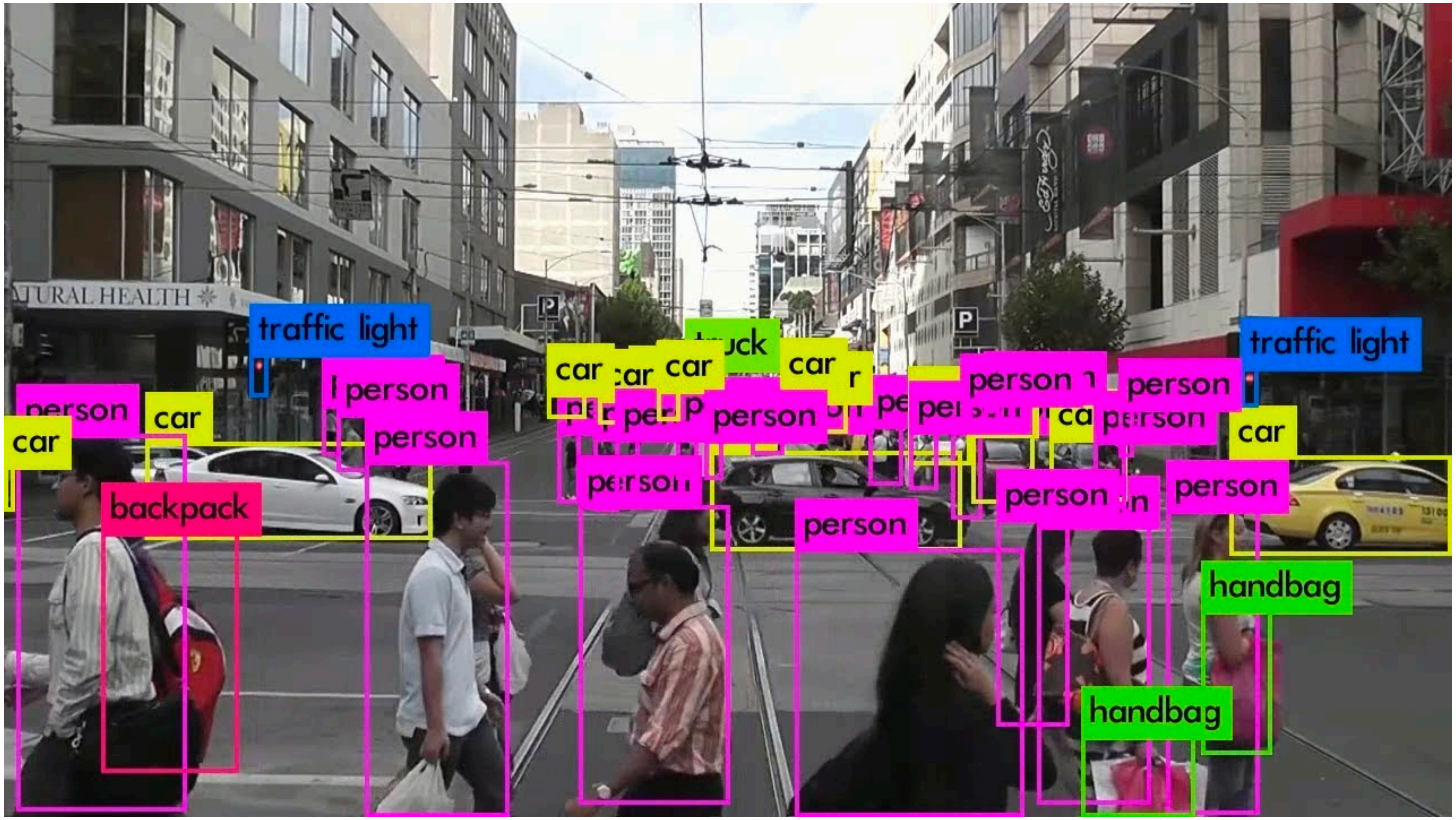
- About 33% of points sources are unassociated
- Can ML both localise and classify point sources better than current state-of-the-art techniques?
- How to deal with uncertainty
- Add radio data to create a multimessenger approach





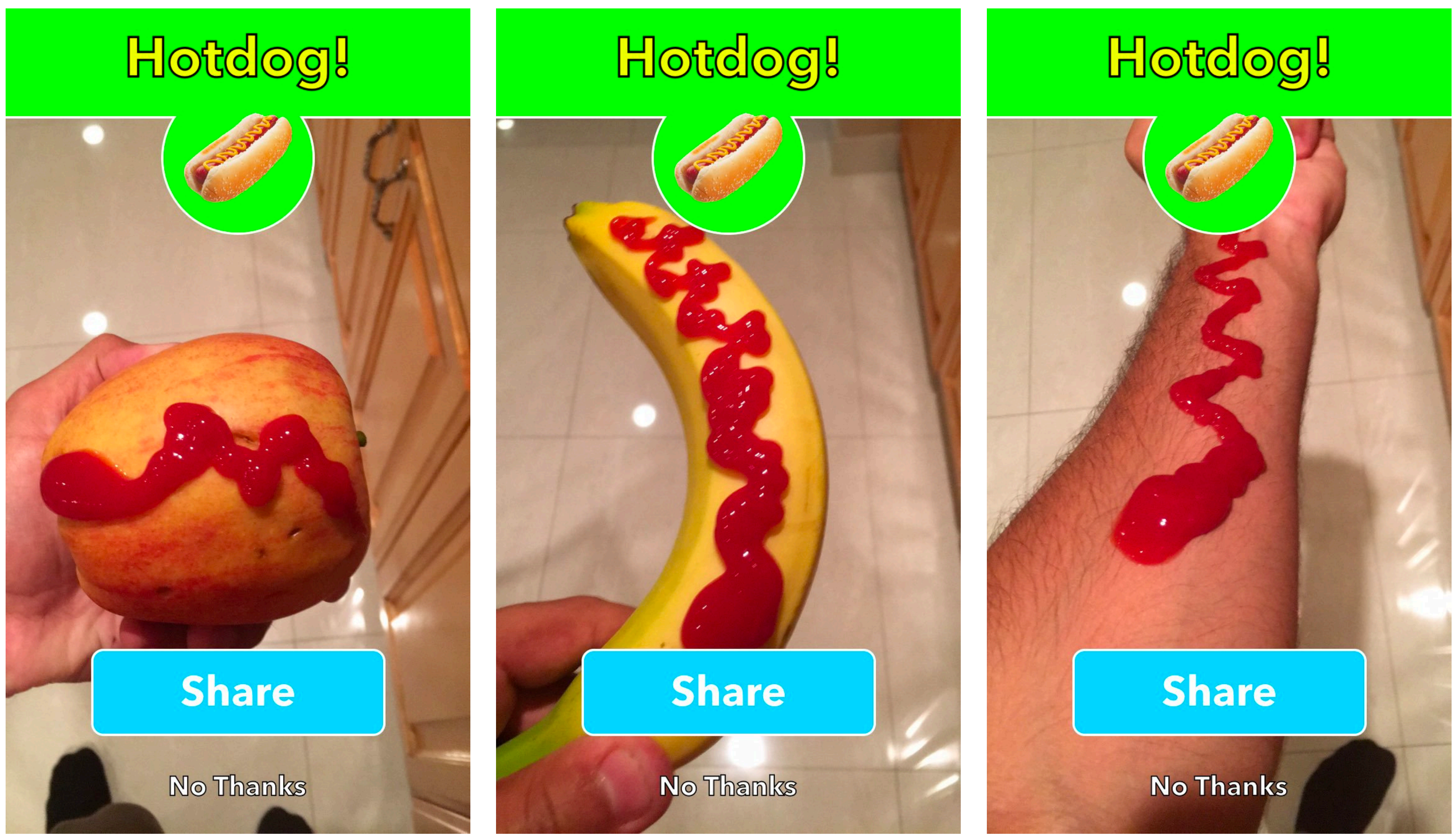
CHALLENGE: INCLUSIVE ANALYSIS OF FERMI POINT SOURCES

- Example: retrain YOLO on gamma-ray data



CHALLENGE: INCLUSIVE ANALYSIS OF FERMI POINT SOURCES

- Example: use Bayesian neural networks to quantify uncertainties
    - Typically quantify using epistemic and aleatoric loss (1703.04977)
    - No uncertainty: your hotdog classifier is actually a ketchup classifier and you don’t know



https://twitter.com/david__kha/status/865093285886304256

CHALLENGE: INCLUSIVE ANALYSIS OF FERMI POINT SOURCES

- Example: use Bayesian neural networks to quantify uncertainties
  - Aleatoric uncertainty: “noise in the data”

$$\mathcal{L}_{\rm NN}(\theta) = \frac{1}{N}\sum_{i=1}^{N} \frac{1}{2\sigma(\mathbf{x}_i)^2}||\mathbf{y}_i - \mathbf{f}(\mathbf{x}_i)||^2 + \frac{1}{2}\log\sigma(\mathbf{x}_i)^2$$

  - Epistemic uncertainty: “NN uncertainty – imperfect training”
    - Monte Carlo dropout

http://mlg.eng.cam.ac.uk/yarin/blog_3d801aa532c1ce.html

MORE CHALLENGES

- Can we detect DM subhalos in gravitational lenses?
  – deep probabilistic programming

- Can we generate events following any particular distribution using generative models?
  – VAEs / GANs

- Can we publish limits and Xs as ML models?
  – SusyAI / DeepXs

…

https://github.com/SydneyOtten/DeepXS
https://susyai.hepforge.org/

TOOLING

- High dimensional data visualisation
- Interactive plotting, shareable urls
- idarksurvey.com – almost production-ready

JOIN?

- Many different DM applications (HEP, astro, detectors, theoretical)
- Many different ML approaches (regression, classification, generative modelling, outlier detection, ...)
- For ML everything is just data – learn from each other and from other fields!
- **Peaked your interested? Join: darkmachines.org**
  Challenges can be joined via CERN mailing lists
  or contacting the challenge coordinators