Novelty detection meets collider physics

Based on arXiv:1807:10261
In collaboration with Tao Liu, Ying-Ying Li and He Wang

Jan Hajer

Centre for Cosmology, Particle Physics and Phenomenology — Université catholique de Louvain

IRN Terascale @ Bruxelles

High Energy Physics (HEP) is ‘Big Data Science’

Long history of applying supervised Machine Learning (ML) to data analysis

1990	Neural network (NN) used for the top quark search in D0
2004	Boosted Decision Trees (BDTs) first applied by MiniBooNE to neutrino data
today	BDT is very popular in HEP data analysis
	e.g. in TOP2018, more than 50 % of the results were based on BDT analyses

High Energy Physics (HEP) is 'Big Data Science'

Long history of applying supervised Machine Learning (ML) to data analysis

- **1990** Neural network (NN) used for the top quark search in D0
- **2004** Boosted Decision Trees (BDTs) first applied by MiniBooNE to neutrino data
- **today** BDT is very popular in HEP data analysis
  e.g. in TOP2018, more than 50 % of the results were based on BDT analyses

Problems beyond Supervised Learning

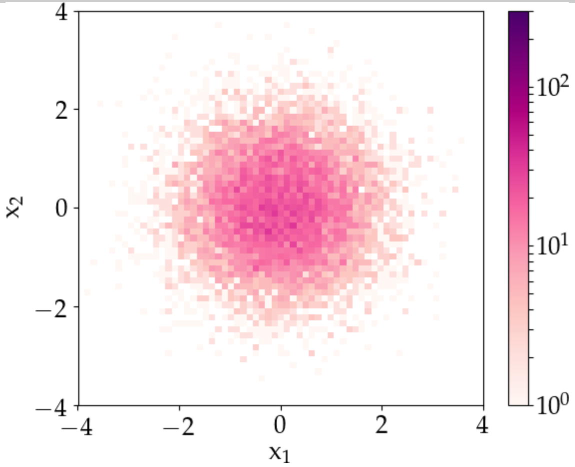
- How can we search for unexpected signals of new physics?
- How can we search model independently for NP in interesting final states?

New Physics with similar final states but different kinematics

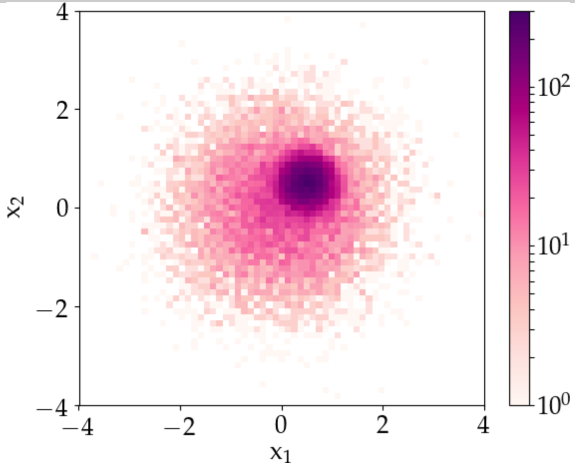
- **Case I** di-top partner T production vs. Z' production (both decaying to top pair)
- **Case II** exotic Higgs decays (rich topologies): $h \to Za$ and $h \to a + \mathrm{DM}$

Toy data samples (2d Gaussian samples)

Training data (Background)



Testing data (Signal + Background)



Dimension d of the input space

here 2

in HEP $\sim 4 \times n_{\text{physics objects}}$

Construction of a Deep Neuronal Network (DNN) for Novelty Detection

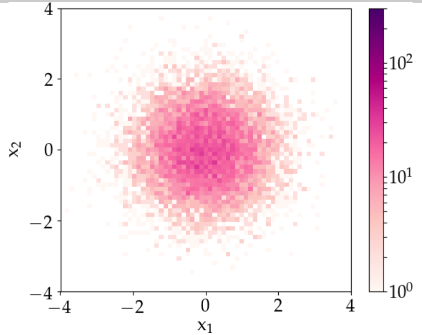
Algorithm (first part)

c_1 c_2

Supervised Classif cation

$f(c_1 \oplus c_2)$

Training data c_1, c_2, ...

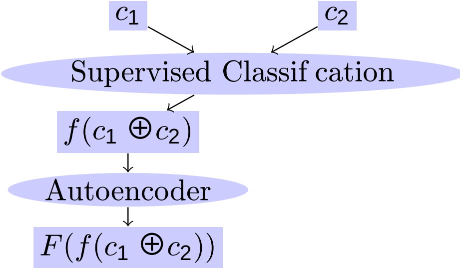


Supervised classification

- Training on the SM background features
- assume l high level variables
- NN with e.g. $N = 4n + 30 + 30 + 10 + l$ nodes
- Leads to a high dimensional feature space of e.g. $D = d + N$

Construction of a Deep Neuronal Network (DNN) for Novelty Detection

Algorithm (first part)



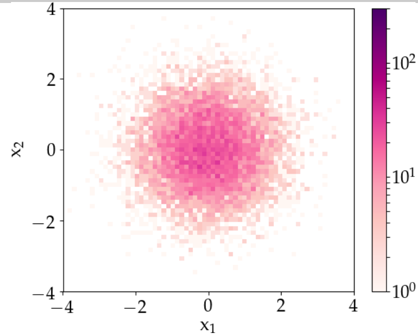
Supervised classification

- Training on the SM background features
- assume l high level variables
- NN with e.g. $N = 4n + 30 + 30 + 10 + l$ nodes
- Leads to a high dimensional feature space of e.g. $D = d + N$

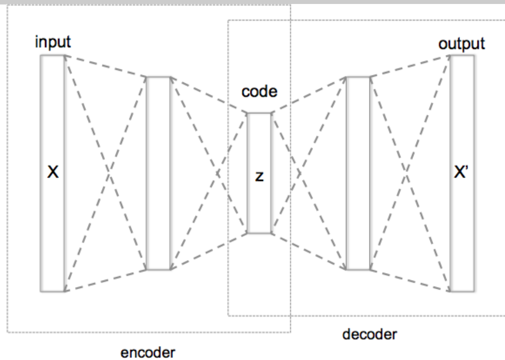
Autoencoder

- Reduces the size of the feature space to m
- Minimizes the reconstruction error $\|x - x'\|^2$
- Learns unsupervised to reconstruct its input
- It creates a submanifold in the feature space

Training data c_1, c_2, ...



Depiction of the Autoencoder DNN



Autoencoder for Novelty (Anomaly) detection at colliders

We were the first to introduce this technique but by now this idea has been picked up

M. Farina, Y. Nakai, and D. Shih. "Searching for New Physics with Deep Autoencoders". arXiv: 1808.08992 [hep-ph].

T. Heimel, G. Kasieczka, T. Plehn, and J. M. Thompson. "QCD or What?" *SciPost Phys.* 6.3, p. 030. DOI: 10.21468/SciPostPhys.6.3.030. arXiv: 1808.08979 [hep-ph].

O. Cerri, T. Q. Nguyen, M. Pierini, M. Spiropulu, and J.-R. Vlimant. "Variational Autoencoders for New Physics Mining at the Large Hadron Collider". *JHEP* 05, p. 036. DOI: 10.1007/JHEP05(2019)036. arXiv: 1811.10276 [hep-ex].

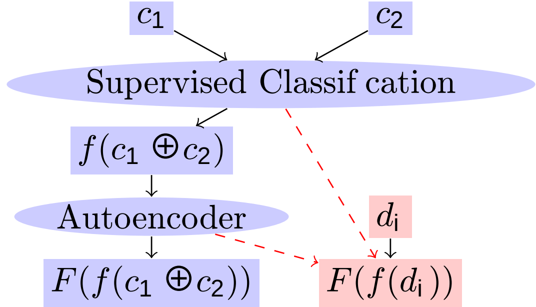
T. S. Roy and A. H. Vijay. "A robust anomaly finder based on autoencoder". arXiv: 1903.02032 [hep-ph]. №: TIFR/TH/19-4.

A. Blance, M. Spannowsky, and P. Waite. "Adversarially-trained autoencoders for robust unsupervised new physics searches". arXiv: 1905.10384 [hep-ph]. №: IPPP/19/41.

and others

Construction of a Deep Neuronal Network (DNN) for Novelty Detection

Algorithm (second part)



Including the signal data

Application of

- f classification
- F autoencoder

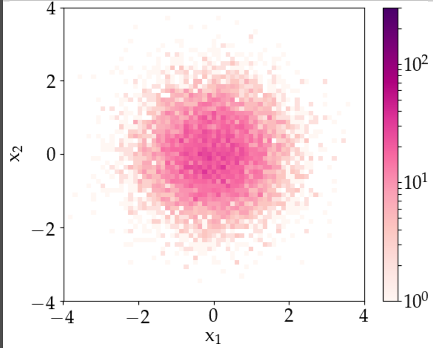
to testing data d_i
gives the relation between

signal	testing data
background	training feature manifold

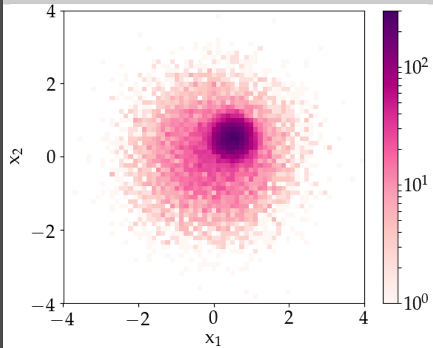
Final step

Novelty Evaluation

Training data c_1, c_2, ...



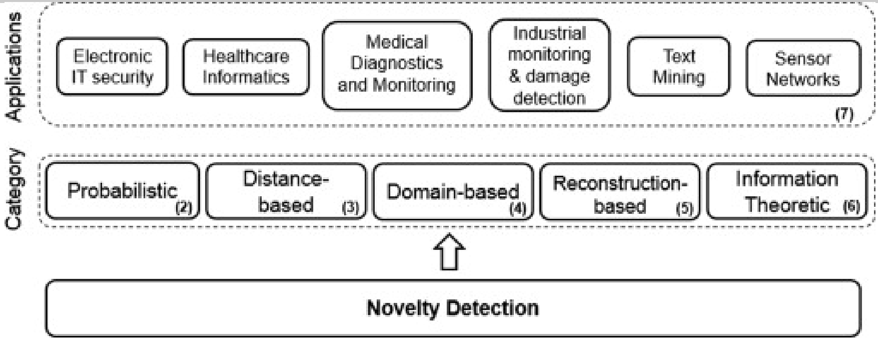
Testing data d_i



Novelty detection

- "The task of detecting novel events without prior knowledge"
- No signal data available for model training
- Model independent and complementary to supervised learning

The history of novelty detection can be told as a history of developing novelty evaluators



Traditional novelty evaluator

- Large distance results in a high score
- Short distance results in a low score
- Therefore it is measure of isolation
- Successfully applied to recognize, e.g., anomalous finger prints or faces

Traditional Novelty Evaluator

Traditional novelty measure

[Kriegel et al. 2009; Socher et al. 2013]

$$\Delta_{\mathrm{trad}} = \frac{d_{\mathrm{train}} - \langle d'_{\mathrm{train}} \rangle}{\langle d'^2_{\mathrm{train}} \rangle^{1/2}}$$

d_{train} mean distance of a testing data point to the k nearest neighbours

$\langle d'_{\mathrm{train}} \rangle$ average of the mean distances of k nearest neighbours

$\langle d'^2_{\mathrm{train}} \rangle^{1/2}$ standard deviation of $\langle d'_{\mathrm{train}} \rangle$

Purpose of this evaluator

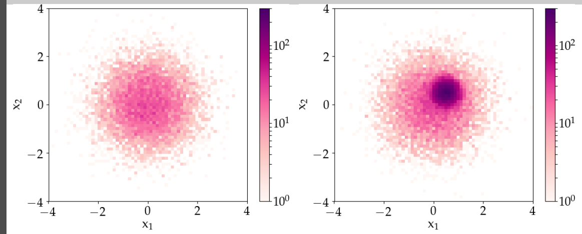
- It measures the isolation of training data from testing data
- Training data points at and beyond the tail of the testing distribution scores high

Cumulative distribution function

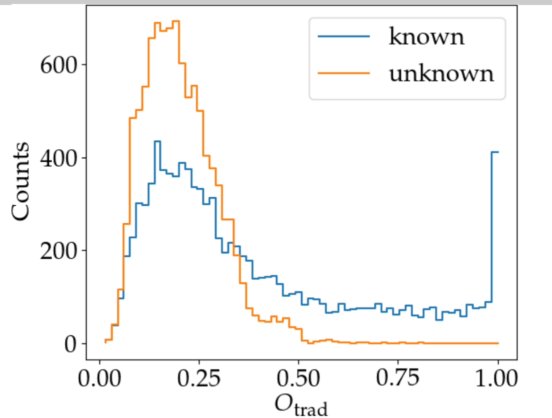
$$\mathcal{O}_{\mathrm{trad}} = \frac{1}{2}\left(1 + \mathrm{erf}\,\frac{\Delta_{\mathrm{trad}}}{c\sqrt{2}}\right) \in [0, 1]$$

c normalization constant.

Data

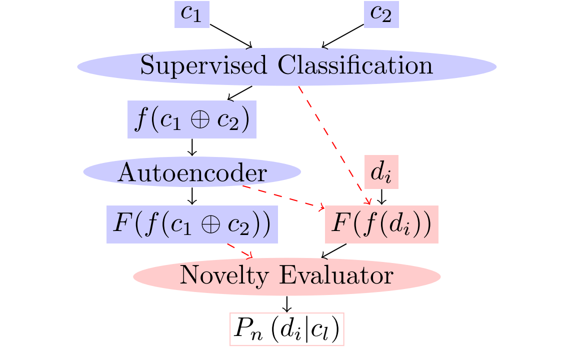


Result



Novelty Evaluation

Our DNN algorithm for novelty detection

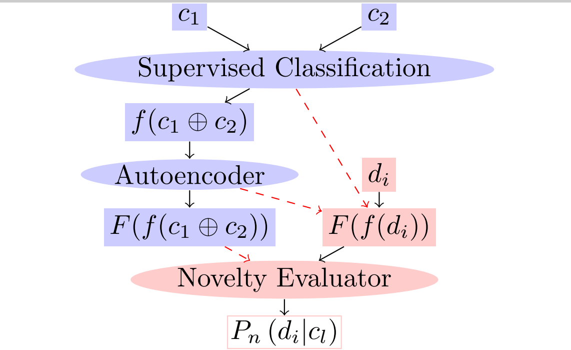


Take home message

This algorithm is able to discover New Physics without prior knowledge of the signal!

Novelty Evaluation

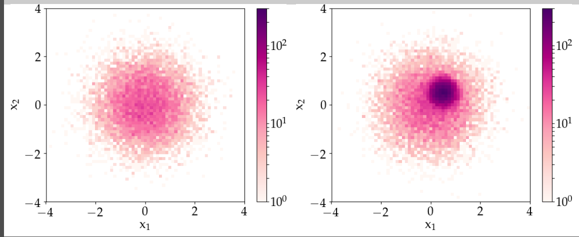
Our DNN algorithm for novelty detection



Take home message

This algorithm is able to discover New Physics without prior knowledge of the signal!

Data

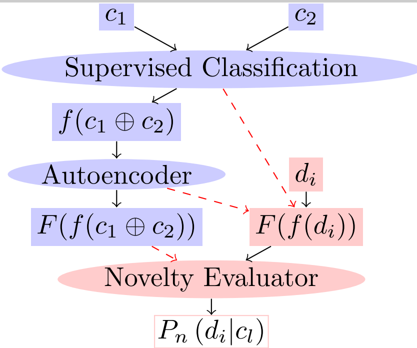


Problems of the traditional evaluator

- Insensitive to clustering in testing data
- Resonances, shape, etc. are clusters and very important for BSM physics detection

Novelty Evaluation

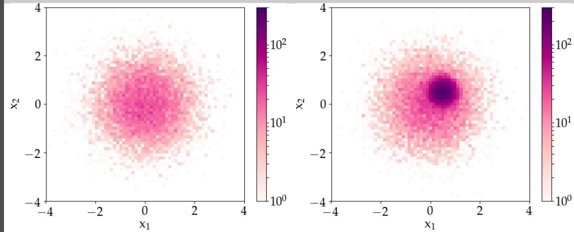
Our DNN algorithm for novelty detection



Take home message

This algorithm is able to discover New Physics without prior knowledge of the signal!

Data



Problems of the traditional evaluator

- Insensitive to clustering in testing data
- Resonances, shape, etc. are clusters and very important for BSM physics detection

Goals for of a new evaluator

Measure of clustering

High score Large density difference

Low score Small density difference

"Scoring according to the clustering around each testing point on top of the training data distribution in a feature space"

Novelty Evaluation

Traditional novelty measure

$$\Delta_{\text{trad}} = \frac{d_{\text{train}} - \langle d'_{\text{train}} \rangle}{\langle d'^2_{\text{train}} \rangle^{1/2}}$$

- d_{train} mean distance of a testing data point to the k nearest neighbours
- $\langle d'_{\text{train}} \rangle$ average of the mean distances of k nearest neighbours
- $\langle d'^2_{\text{train}} \rangle^{1/2}$ standard deviation of $\langle d'_{\text{train}} \rangle$

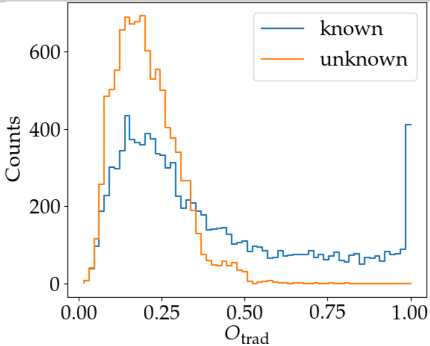
New novelty measure

$$\Delta_{\text{new}} = \frac{d_{\text{test}}^{-m} - d_{\text{train}}^{-m}}{d_{\text{train}}^{-m/2}} \propto \left. \frac{S}{\sqrt{B}} \right|_{\text{local bin}}$$

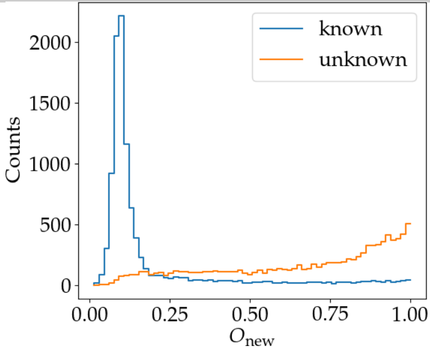
- d_{test} mean distance of a testing data point to its k nearest neighbours
- m dimension of the feature space

Compares local densities of the testing point in the training and testing datasets

Traditional performance



New performance



$\mathcal{O}_{\text{new}}$ distinguishes significantly better between signal and background data than $\mathcal{O}_{\text{trad}}$

Look elsewhere effect

Fluctuations of the background (especially in the tail)

- Appear as clusters (of potential NP) in the new evaluator

Look elsewhere effect

Fluctuations of the background (especially in the tail)

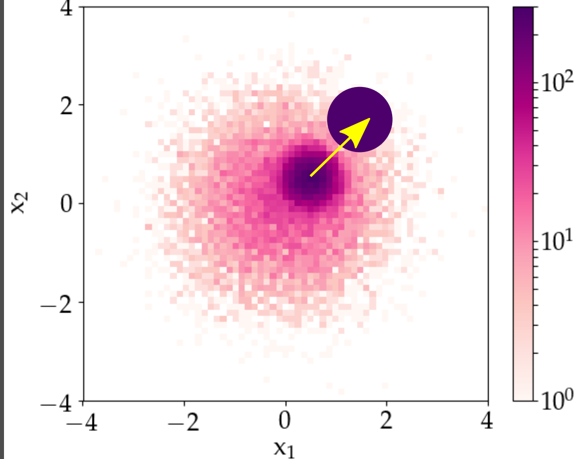
- Appear as clusters (of potential NP) in the new evaluator
- Are close to the background manifold in the old evaluator

Proposed solution: Combined evaluator

$$\mathcal{O}_{\text{comb}} = \sqrt{\mathcal{O}_{\text{trad}}\mathcal{O}_{\text{new}}} \; .$$

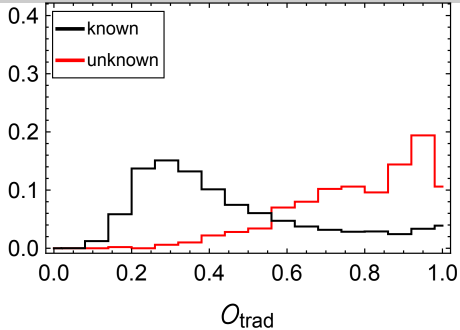
Demonstration with a fluctuation like signal

Novel data in the tail with small S/B of $1/20$

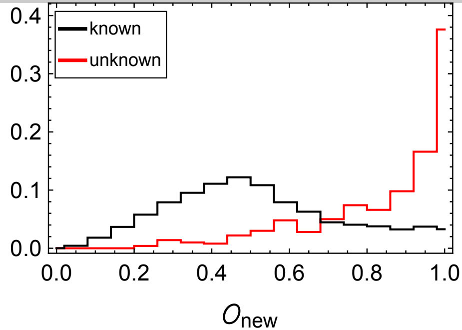


Comparison of all three evaluators

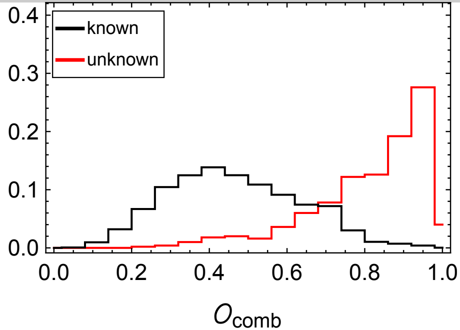
Traditional evaluator



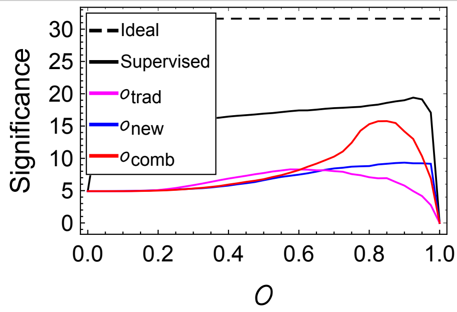
New evaluator



Combined evaluator



Significances



$\mathcal{O}_{\text{comb}}$ outperforms both $\mathcal{O}_{\text{trad}}$ and $\mathcal{O}_{\text{new}}$.

Proof of concept: Application to HEP Monte Carlo data

Exotic Higgs decays at 240 GeV e^+e^- collider

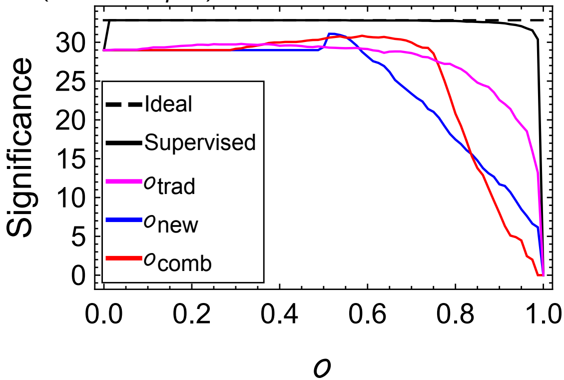
SM processes

- $e^+e^- \to hZ \to Z^*_{\text{inv}} Z_{\bar{b}b} l^+l^-$
  $\sigma = 0.00686\,\text{fb}$
- $e^+e^- \to hZ \to Z^*_{\bar{b}b} Z_{\text{inv}} l^+l^-$
  $\sigma = 0.00259\,\text{fb}$

$h \to \widetilde{\chi}_1\widetilde{\chi}_2 \to \widetilde{\chi}_1\widetilde{\chi}_1 a$ in the NMSSM

$m_{N_1} = m_{N_2}/9 = m_a/4 = 10$ GeV

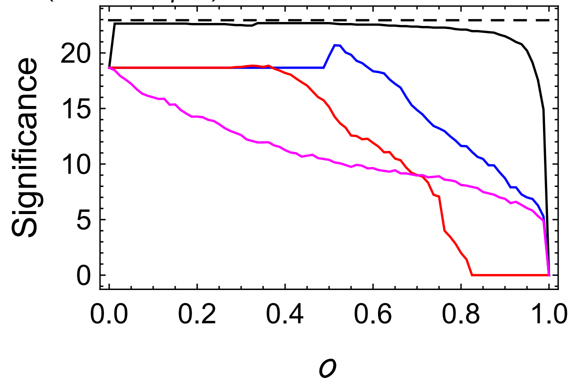
BR$(h \to \bar{b}bE_T^{\text{miss}}) = 1\,\%$ $\sigma = 0.108\,\text{fb}$



$h \to Za$ in the 2HDM and the NMSSM

$m_a = 25$ GeV

BR$(h \to \bar{b}bE_T^{\text{miss}}) = 1\,\%$ $\sigma = 0.053\,\text{fb}$



$\mathcal{O}_{\text{new}}$ has the best performance due to the large S/B

Di-top (leptonic) production at the 14 TeV LHC $\bar{b}bl^+l^-E_T^{\text{miss}}$

SM processes (for simplification σ scaled by 1/2000)

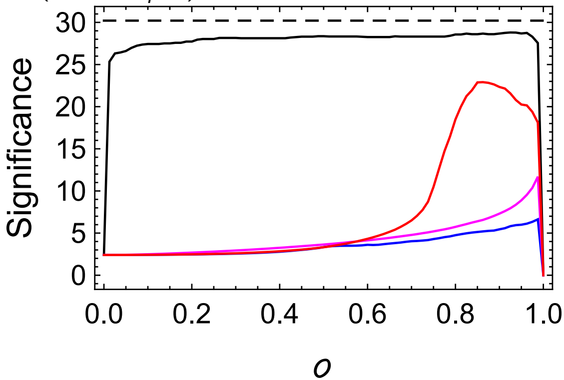
- $pp \to \bar{t}_l t_l$
  $\sigma = 11.5\,\text{fb}$
- $pp \to t_l \bar{b} W_l^\pm$
  $\sigma = 0.365\,\text{fb}$
- $pp \to Z_b Z_l$
  $\sigma = 0.0765\,\text{fb}$

$pp \to \overline{T}T \to W_l^+ W_l^- \bar{b}b$

$\overline{T}$ and T are fermionic top partners

$m_T = m_{\overline{T}} = 1.2\,\text{TeV}$

$\text{BR}(T \to W_l^+ b) = 50\,\%$ $\sigma = 0.152\,\text{fb}$



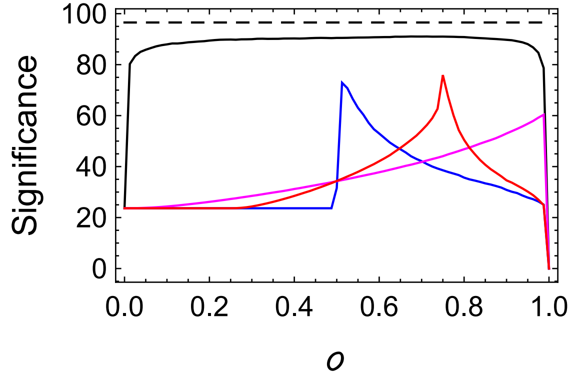
performance comparable to 2d Gaussian sample

$pp \to Z' \to \bar{t}t$

Z' is a new gauge boson

$g_{Z'} = g_Z$ $m_{Z'} = 3\,\text{TeV}$

$\text{BR}(Z' \to \bar{t}t) = 16.7\,\%$ $\sigma = 1.55\,\text{fb}$



O_{comb} and O_{comb} are comparable due to large S/B

Result

Successes of our DNN

Effective	works well at a realistic level (e.g. after hadronization)
Efficient	small sensitivity discrepancy between novelty detection and supervised learning
Economic	Comparably low computing resources necessary

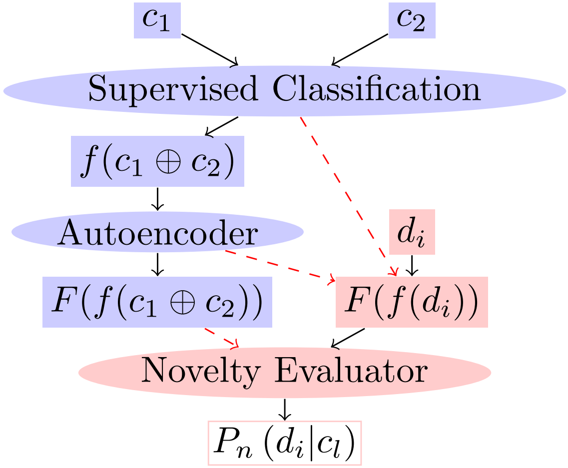
Final problem

Every uncertainty appears as novel events

- uncertainties in MC tools
- theoretical errors
- missed backgrounds
- etc.

Summary

- Rapid development of DNNs has far-reaching impact on particle physics
- A combination of supervised learning and novelty detection may lay out the framework for future data analysis
- Properly designed novelty evaluators show encouragingly high sensitivity in detecting unexpected NP
- More efforts are needed to fill the gap between proof of concept and real data analysis



References I

J. Hajer, Y.-Y. Li, T. Liu, and H. Wang. "Novelty Detection Meets Collider Physics". arXiv: 1807.10261 [hep-ph].

M. A. Pimentel, D. A. Clifton, L. Clifton, and L. Tarassenko. "A review of novelty detection". *Signal Processing* 99, pp. 215–249. DOI: 10.1016/j.sigpro.2013.12.026.

H.-P. Kriegel, P. Kröger, E. Schubert, and A. Zimek. "LoOP: Local Outlier Probabilities". *Proceedings of the 18th ACM Conference on Information and Knowledge Management*. CIKM '09. Hong Kong, China: ACM, pp. 1649–1652. ISBN: 978-1-60558-512-3. DOI: 10.1145/1645953.1646195.

R. Socher, M. Ganjoo, C. D. Manning, and A. Ng. "Zero-Shot Learning Through Cross-Modal Transfer". *Advances in Neural Information Processing Systems 26*. Ed. by C. J. C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Q. Weinberger. Curran Associates, Inc., pp. 935–943. URL: http://papers.nips.cc/paper/5027-zero-shot-learning-through-cross-modal-transfer.pdf.