

DeepEfficiency

arXiv:1809.06101

Mikael Mieskolainen

IRN Terascale

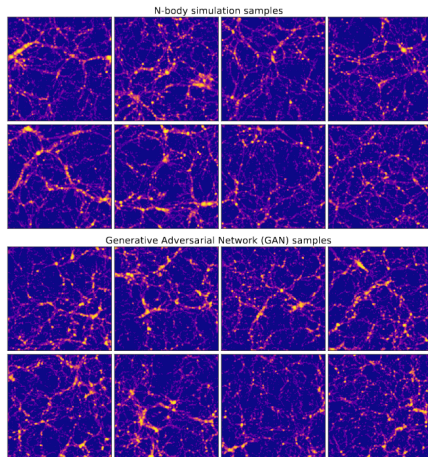
21/05/2019, Annecy, France



UNIVERSITY OF HELSINKI

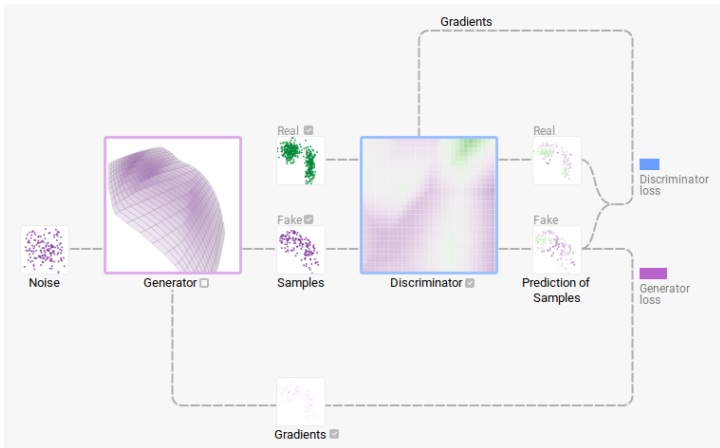
Neural networks – the meta-math-tool for everything?

Example: "Fast Cosmic Web Simulations with GAN networks" (internally a Nash equilibrium type minimax optimization problem)



[Rodriguez et al., ETH Zürich, [arXiv:1801.09070](https://arxiv.org/abs/1801.09070), orders of magnitude speed improvement in 2D slices]

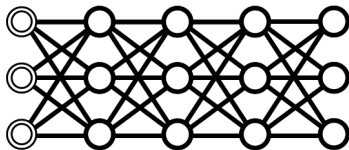
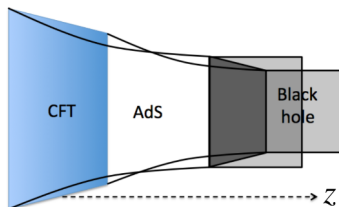
Learn the tool: play with GAN in your browser!
<https://poloclub.github.io/ganlab>



[Kahng et al., Georgia Tech/Google brain, [arXiv:1809.01587](https://arxiv.org/abs/1809.01587)]

Example: Neural networks meet holography in "AdS/CFT as a deep Boltzmann machine" (Boltzmann net is Ising model like net)

data \simeq boundary scalar field theory generating functional, hidden network weights \simeq bulk spacetime metric (geometry), the deepest layer \simeq the black-hole horizon



Now this work

[MM, DeepEfficiency, [arXiv:1809.06101](#)]

A new way implementing detector fiducial efficiency corrections in higher phase space dimensions, towards optimal measurements

Talk 15 min, so accurate details you can find in the [Tensorflow](#) framework based code at

[<github.com/mieskolainen>](https://github.com/mieskolainen), MIT open source

Try it out, make it better

Basic HEP Measurement

Cross section (differential)

$$\sigma = \frac{N - \langle B \rangle}{\mathcal{E} \times A \times L}$$

- Events: N , number of measured events after trigger & selection cuts
- Background: $\langle B \rangle$, expected number of background events
- Integrated luminosity: L , van der Meer scan calibration + online tech.
- Acceptance (Geometric): $A \rightarrow 1 \Leftrightarrow$ fiducial measurement,
 $A \rightarrow 0$, highly non-fiducial (extrapolating)
- Efficiency: $\mathcal{E} \equiv$ number of selected / number of generated within
fiducial phase space $\in [0, 1]$

In this work, we reformulate the detector efficiency term \mathcal{E} evaluation as a problem of learning a probabilistic function

$$\mathcal{E} : \mathbb{R}^D \rightarrow [0, 1]$$

Fully differential final state event kinematics in, probability out.

Before DeepEfficiency

Histograms used in *Bills of Mortality* (1662)

by John Graunt

Natural and Political
OBSERVATIONS

Mentioned in a following INDEX,
and made upon the

Bills of Mortality.

B Y

Capt. JOHN GRAUNT,
Fellow of the *Royal Society.*

With reference to the *Government, Religion, Trade, Growth, Air, Diseases,* and the
several Changes of the said CITY.

So I expect histogram bin-by-bin ‘counting efficiency corrections’ of measurements have been used for centuries.

This is obvious and quite solid¹, but only if your observable is *truly* low-dimensional. However, in HEP, observables (differential distributions) are constructed as a function of multidimensional final state kinematics.

¹Your problem may extend to include heavy unfolding/deconvolution type of corrections of detector resolution/bias, algorithmically since 1970's.

Main Problem and Solution

The usual way of doing efficiency corrections is to rely on the MC event generator + detector simulation (GEANT). We use GEANT too, but minimize (even get rid of) the event generator dependence by **using fully differential final state kinematics**.

MC theory dependence: **Bias² can (will) propagate** from the event generator (dynamics) if no fully differential treatment, whenever the process \simeq scattering amplitudes + rest cannot be well simulated. For example: transverse momentum low- p_t parametrizations, system high- p_t ISR shower (MC resummation) uncertainties, **QCD** color coherence, non-perturbative spin polarization \sim decay angle (θ, φ) dependence e.g. in QCD spectroscopy...

² Usually handled by dumping into systematics by varying the event generators (even if they use in essence the same model...).

What people have done to handle this problem?

- Multidimensional histograms, bin over differential kinematics (e.g. system mass, rapidity, transverse momentum), approx $D = 3$ practical maximum
- Histograms + Spherical Harmonic Expansions³ on top of them – for the angular dependence
- **These do not scale to higher dimensions**, exponential statistics requirement for histogram hypercells – think about exponential number of binary combinations = 2^D as a function of D , then change binary (2) to some large integer (# bins) approximating real line, much worse

³You find this expansion algorithm in my [GRANIITTI engine](https://github.com/mieskolainen/graniitti),
<github.com/mieskolainen/graniitti>

DeepEfficiency

Universal Approximator

So, we assume that a wide and deep enough fully connected feed-forward network will work as a universal function approximator and that modern neural network optimization tools can optimize⁴ the network, based on (stochastic) gradient search, with the underlying technique being *Reverse Mode Automatic Differentiation* (backpropagation as a special case)

Original RMAD: [Seppo Linnainmaa, MSc thesis, Helsinki, 1970](#)

For the universal approximator theorems:

[Wikipedia, Universal approximation theorem](#)

See also the classic [Stone-Weierstrass theorem](#)

⁴Difficult non-convex optimization problem

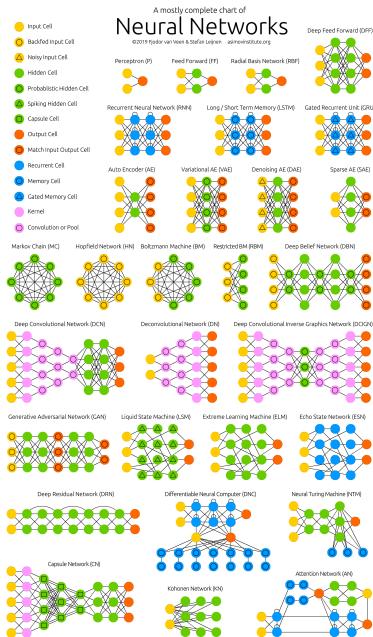


Figure: Types of neural nets, [Fjodor van Veen], we use the top right ↑

Network for Probabilities

To train for probabilities, we need to train (minimize) using cross entropy cost ⁵

$$L = -\frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} [R_i \ln \mathcal{E}_i + (1 - R_i) \ln(1 - \mathcal{E}_i)],$$

where R_i is simulated selection (0=lost,1=selected) for the i -th event in the sample \mathcal{S} and \mathcal{E}_i the network output. We found hyperbolic tangent to work fine as the activation function with 5 hidden layers and ~ 100 neurons per layer. Also, we use sigmoid $f(x) = 1/(1 + e^{-x})$ as the final layer function to get output values in $[0, 1]$.

⁵Basically, this is an extension of [Logistic Regression](#), both well known in machine learning.

How to use

Arbitrary efficiency corrected distributions (1D, 2D, profile...) of observables \mathcal{O} can be constructed with⁶.

$$\frac{dN}{d\mathcal{O}} = h_{\mathcal{O}}(\{\vec{p}\}) \odot [\mathcal{E}(\{\vec{p}\})]^{-1},$$

where $h_{\mathcal{O}}$ is a probability distribution estimator operator, such as a (bin width $\Delta\mathcal{O}$ normalized) histogram. The point-wise operator \odot is defined as an integral (sum) over the event sample, event weight $\mathcal{E}(\{\vec{p}\})$ from the trained network.

```
In ROOT ~ for ( i over data event sample ) {  
    weight_i = network(kinematics_i);  
    histObsA->Fill( observableA(kinematics_i), 1.0/weight_i );  
    histObsB->Fill( observableB(kinematics_i), 1.0/weight_i ); ...  
}
```

⁶Inverse weighting known as [Horvitz-Thompson estimator](#) in statistics. [Efron's bootstrap re-sampling](#) may be used for statistical uncertainties.

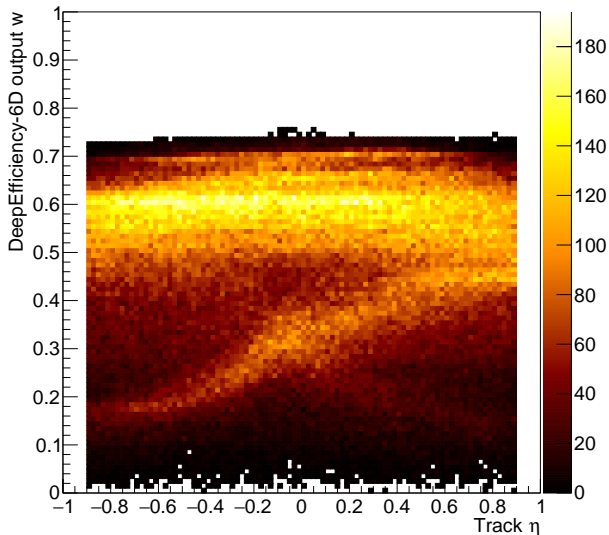


Figure: Non-linear detector visible in 6D-network event weights \mathcal{E} ($= w$) versus pion track pseudorapidity η . Setup described in the paper.

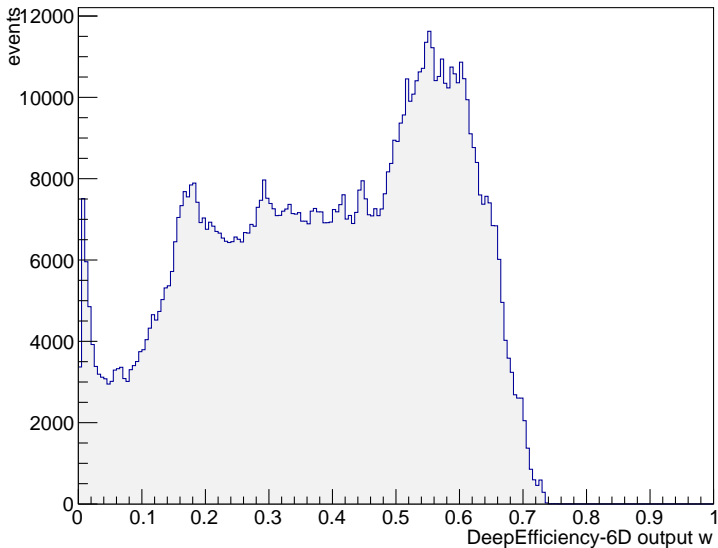


Figure: 6D-Network output weight \mathcal{E} ($= w$) statistics example.

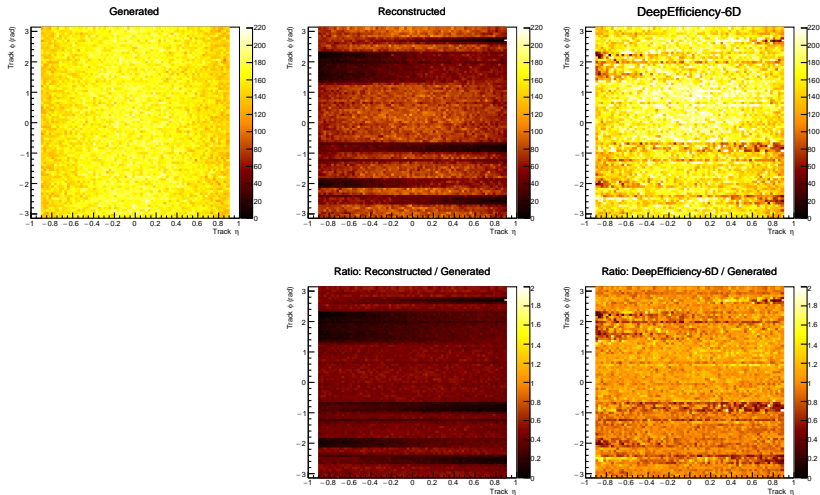


Figure: [GENERATED, RECONSTRUCTED, INVERTED]: 6D-Network inversion in terms of single pion track (η, ϕ). Stripes propagate from the inactive regimes of the inner most silicon detector layer. Train sample size < 10 million events. Quite solid inversion of sharp discontinuities.

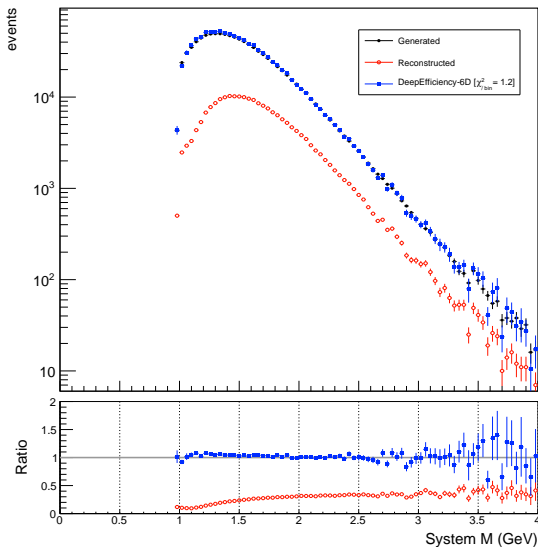


Figure: 6D-Network inversion of low-mass K^+K^- exclusive QCD diffraction events with $\chi^2 / \text{bin} = 1.2$. Train sample size < 10 million events. Fiducial domain: kaon $p_t > 0.15$ GeV, $|\eta| < 0.9$.

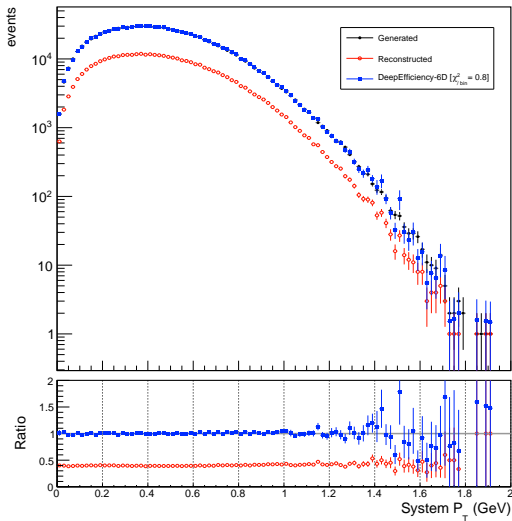


Figure: 6D-Network inversion of low-mass $\pi^+\pi^-$ exclusive QCD diffraction events with $\chi^2 / \text{bin} = 0.8$. Train sample size < 10 million events. Seemingly easy case, differentially flat observable within selected fiducial domain: pion $p_t > 0.15$ GeV, $|\eta| < 0.9$.

POSITIVE FEATURES

- ♠ Maximally MC **event generator independent efficiency corrections due to high dimensional inversion** \Rightarrow can use even pure phase space generator (to illustrate the point, in practise not smart/possible always)
- ♠ Can take highly complicated experimental correlations into account, such as complex trigger etc.
- ♠ Arbitrary observables can be efficiency corrected

NEGATIVE FEATURES

- Practical performance is limited, of course, by the MC sample sizes, network structure etc.
- Pure acceptance **holes** inside the fiducial domain need to be taken into account with an additional algorithm, cannot divide null by 0

UNIVERSAL

- ◆ Bias-Variance tradeoff theorem; no free lunch (in general)

Bias: over/undershooting the efficiency correction differentially

Variance: anomalous fluctuations in the efficiency correction weights give fluctuations in observables

After DeepEfficiency

Unsolved Neural Theory

(Riemannian geometry, statistical mechanics, information theory, topology, new math needed?)

- There is no really deep (useful) theory of generic deep neural networks – thus, the optimality claims here are formal, of course – relying on the universal approximator lemma & ‘neural engineering’
- Depth vs width (series vs parallel), type of non-linearity, architecture (feed-forward, convolution, recursive loops ...), gradient descent methods
- How to use Vapnik-Chervonenkis dimension: ‘the theoretical capacity of a statistical learning algorithm’ in practise?

Unsolved Algorithms for Theory and Experiment

- **Detector bias/resolution unfolding** in higher dimensions $\mathbb{R}^D \rightarrow \mathbb{R}^D$ (quite difficult), related to super-resolution etc.
- **Efficient high dimensional Monte Carlo integration** and event generation, beyond VEGAS, multichanneling and hand-engineered Jacobians. For work towards solving this, see: Bendavid, [arXiv:1707.00028](#); better in Müller et al, [arXiv:1808.03856](#)
- **Neural Scattering Amplitudes** (think about GAN-like generator networks). Use for probing theory landscapes beyond Lagrangian descriptions and perturbation theory – more powerful re-visit of the 1960's S -matrix era?

Conclusions

A new efficiency inversion algorithm for **precision fiducial measurements** using fully differential (multidimensional) final state kinematics

⇒ can be used also with likelihood analyses such as the matrix element method, (fast) simulations, replacing other efficiency parametrizations

Full code at [<github.com/mieskolainen>](https://github.com/mieskolainen)