How to GAN for LHC

Tilman Plehn

Universität Heidelberg

TRIUMF 8/2020

Learning from art

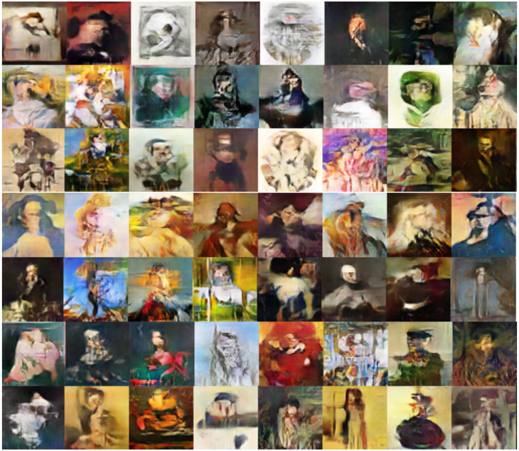
GANGogh [Bonafilia, Jones Danyluk (2017)]

- can networks generate something new?
- map noise vector to images
- train on 80,000 pictures [organized by style and genre]
- generate flowers

Learning from art

GANGogh [Bonafilia, Jones Danyluk (2017)]

- can networks generate something new?
- map noise vector to images
- train on 80,000 pictures [organized by style and genre]
- generate portraits



Learning from art

GANGogh [Bonafilia, Jones Danyluk (2017)]

– can networks generate something new?
– map noise vector to images
– train on 80,000 pictures [organized by style and genre]

Edmond de Belamy [Caselles-Dupre, Fautrel, Vernier]

– trained on 15,000 portraits
– sold for $ 432.500

⇒ all about marketing

GAN basics (for LHC theory plumbers)

Simulations crucial for LHC physics [review: Butter & TP in Kazu Terao's book; ask Wojciech]

- goal: data-to-data with fundamental physics input
- Monte Carlo challenges

  higher-order precision in bulk
  coverage of tails
  inversion to access fundamental QCD
- neural network benefits

  training on MC and/or real events
  lightning speed, once trained
  best available interpolation

GAN basics (for LHC theory plumbers)

Simulations crucial for LHC physics [review: Butter & TP in Kazu Terao's book; ask Wojciech]

– goal: data-to-data with fundamental physics input

– Monte Carlo challenges

higher-order precision in bulk
coverage of tails
inversion to access fundamental QCD

– neural network benefits

training on MC and/or real events
lightning speed, once trained
best available interpolation

GANning data [Goodfellow etal (2014)]

– training true events $\{x_T\}$ following P_T
output generated events $\{r\} \to \{x_G\}$ following P_G

– discriminator constructing $D(x)$ [$D(x) = 1, 0$ true/generated]

$$L_D = \big\langle -\log D(x) \big\rangle_{x \sim P_T} + \big\langle -\log(1 - D(x)) \big\rangle_{x \sim P_G} \to -2\log 0.5$$

– generator producing true-looking events [D needed]

$$L_G = \big\langle -\log D(x) \big\rangle_{x \sim P_G}$$

⇒ statistically independent copy of training events

1– Jet generation

GANGogh for jet images [de Oliveira, Paganini, Nachman]

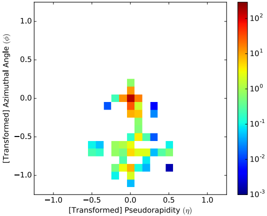
– start with calorimeter or jet images [η vs ϕ]
  sparsity the technical challenge [cf top tagging comparison]

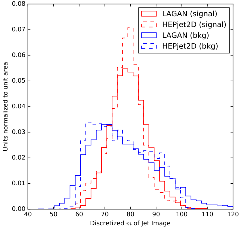
1- reproduce valid jet images from training data

2- organize them by QCD vs W-decay jets

– high-level observables m, τ_{21} as check

$\Rightarrow$ GANs generating jets





GAN questions

– use cases?

– control?

– uncertainty? [Bayesian networks?]

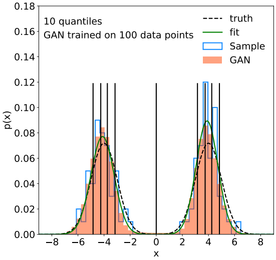
– gain?

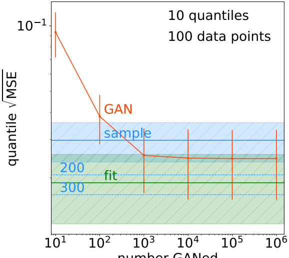
– achievable precision?

2– GANplification

Gain beyond training data [Butter, Diefenbacher, Kasieczka, Nachman, TP]

- true function known
  compare GAN vs sampling vs fit
- χ^2-goodness in quantiles

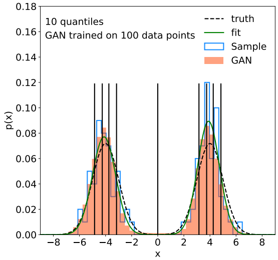


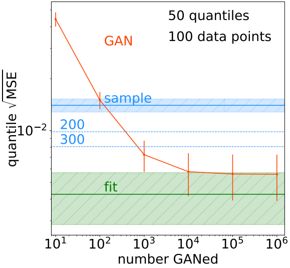


2– GANplification

Gain beyond training data [Butter, Diefenbacher, Kasieczka, Nachman, TP]

- true function known
  compare GAN vs sampling vs fit
- χ^2-goodness in quantiles
- fit like 500-1000 sampled points
  GAN like 500 sampled points [amplifictation factor 5]
  improvement up to 10,000 GANned events



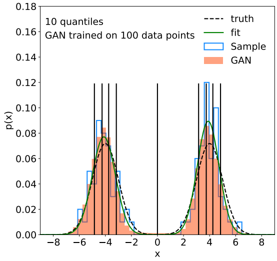


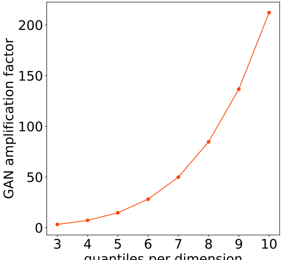
2– GANplification

Gain beyond training data [Butter, Diefenbacher, Kasieczka, Nachman, TP]

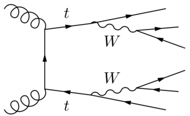
- true function known
  compare GAN vs sampling vs fit
- χ^2-goodness in quantiles
- fit like 500-1000 sampled points
  GAN like 500 sampled points [amplifictation factor 5]
  improvement up to 10,000 GANned events
- 5-dimensional Gaussian shell
  sparsely populated
  amplification vs quantiles
- fit-like additional information
- interpolation and resolution the key [NNPDF]

$\Rightarrow$ GANs enhance training data





3– How to GAN LHC Events

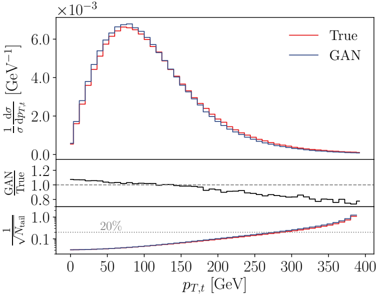


Replace ME for hard scattering [Otten, Hashemi, Di Sipio...]

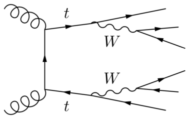
– realistic final state $t\bar{t} \to 6$ jets [Butter, TP, Winterhalder]

on-shell external states $\to$ 12D phase space

– top observables with tails [statistical/systematic error indicated]

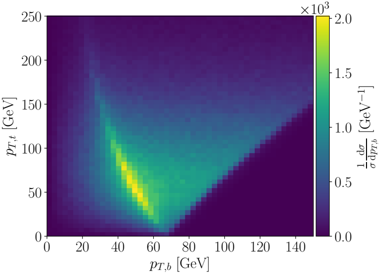


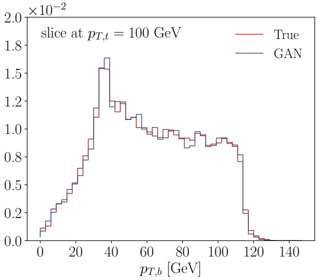
3– How to GAN LHC Events



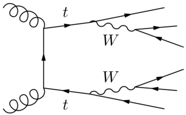
Replace ME for hard scattering [Otten, Hashemi, Di Sipio...]

- realistic final state $t\bar{t} \to 6$ jets [Butter, TP, Winterhalder]
  on-shell external states $\to$ 12D phase space
- top observables with tails [statistical/systematic error indicated]
- 2D correlations





3– How to GAN LHC Events



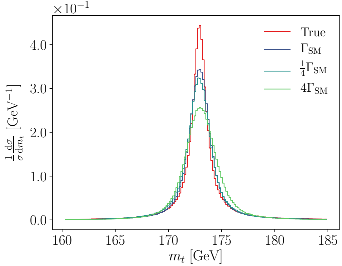
Replace ME for hard scattering [Otten, Hashemi, Di Sipio...]

- realistic final state $t\bar{t} \to 6$ jets [Butter, TP, Winterhalder]
  on-shell external states $\to$ 12D phase space
- top observables with tails [statistical/systematic error indicated]
- 2D correlations
- 1D-invariant masses [top, W]
  batch-wise discrimination, MMD loss with kernel k

$$\text{MMD}^2 = \langle k(x,x') \rangle_{x,x' \sim P_T} + \langle k(y,y') \rangle_{y,y' \sim P_G} - 2\langle k(x,y) \rangle_{x \sim P_T, y \sim P_G}$$

$$L_G \to L_G + \lambda_G \,\text{MMD}^2$$

- GANning 1.6M evts/sec on laptop



4– How to GAN event subtraction

Idea: subtract event samples without binning [Butter, TP, Winterhalder]

- statistical uncertainty

$$\Delta_{B-S} = \sqrt{\Delta_B^2 + \Delta_S^2} > \max(\Delta B, \Delta S)$$

- possible applications

  soft-collinar subtraction, multi-jet merging
  on-shell subtraction
  background subtraction [4-body decays]

4– How to GAN event subtraction

Idea: subtract event samples without binning [Butter, TP, Winterhalder]

– statistical uncertainty

$$\Delta_{B-S} = \sqrt{\Delta_B^2 + \Delta_S^2} > \max(\Delta B, \Delta S)$$

– possible applications

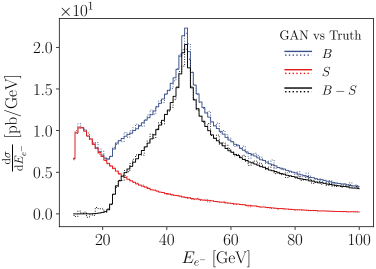
soft-collinar subtraction, multi-jet merging
on-shell subtraction
background subtraction [4-body decays]

– event-based background subtraction

$pp \to e^+ e^-$ (Base) $pp \to \gamma \to e^+ e^-$ (Subtracted)

– Z-pole events generated

$\Rightarrow$ Why did we ever bin?



5– How to GAN away detector effects

Idea: invert Monte Carlos [Datta; Bellagente, Butter, Kasiczka, TP, Winterhalder]

– detector simulation — unfolding established use case

– inversion possible, in principle [entangled convolutions]

– GAN task

$$\text{partons} \xrightarrow{\text{DELPHES}} \text{detector} \xrightarrow{\text{GAN}} \text{partons}$$

$\Rightarrow$ Full phase space unfolding

5– How to GAN away detector effects

Idea: invert Monte Carlos [Datta; Bellagente, Butter, Kasiczka, TP, Winterhalder]

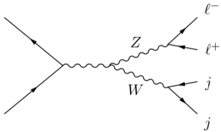
– detector simulation — unfolding established use case

– inversion possible, in principle [entangled convolutions]

– GAN task

$$\text{partons} \xrightarrow{\text{DELPHES}} \text{detector} \xrightarrow{\text{GAN}} \text{partons}$$

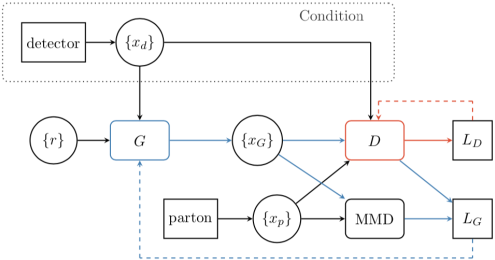
$\Rightarrow$ Full phase space unfolding



Reconstructing parton-level $pp \to ZW \to (\ell\ell)\ (jj)$

– broad jj mass peak
narrow $\ell\ell$ mass peak
modified $2 \to 2$ kinematics

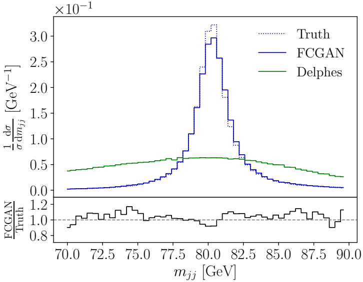
– (conditional) GAN like for event generation



Fully conditional GAN

Test data modified from training data

– full inversion no point in showing...



Fully conditional GAN

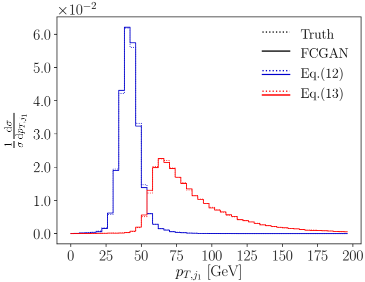
Test data modified from training data

- full inversion no point in showing...
- test cuts [14%, 39% events]

$$p_{T,j_1} = 30 \dots 50 \text{ GeV} \quad p_{T,j_2} = 30 \dots 40 \text{ GeV} \quad p_{T,\ell-} = 20 \dots 50 \text{ GeV} \tag{12}$$

$$p_{T,j_1} > 60 \text{ GeV} \tag{13}$$

$\Rightarrow$ Phase space unfolding working



Fully conditional GAN

Test data modified from training data

- full inversion no point in showing...
- test cuts [14%, 39% events]

$$p_{T,j_1} = 30 \dots 50 \text{ GeV} \quad p_{T,j_2} = 30 \dots 40 \text{ GeV} \quad p_{T,\ell-} = 20 \dots 50 \text{ GeV} \tag{12}$$

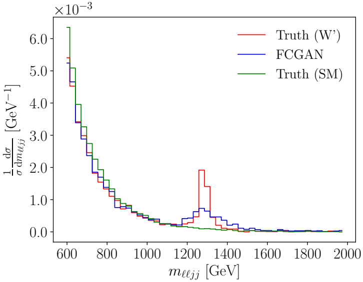
$$p_{T,j_1} > 60 \text{ GeV} \tag{13}$$

$\Rightarrow$ Phase space unfolding working

New physics in data [model dependence]

- train: Standard Model events
  test: 10% events with W' in s-channel
- nightmare: unfold W' onto Standard Model?

$\Rightarrow$ Statistical model: cINN [Bellagente etal]



No conclusion...

LHC physics really is big data

- NN best interpolation [Butter (2020)]
- training on MC and/or data
- latent space structured

GAN studies

- Jet Images [de Oliveira (2017), Carazza (2019)]
- Detector simulations [Paganini (2017), Musella (2018), Erdmann (2018), Ghosh (2018), Buhmann (2020)]
- Events [Otten(2019), Hashemi (2019), Di Sipio (2019), Butter (2019), Martinez (2019), Alanazi (2020)]
- Unfolding [Datta (2018), Bellagente (2019)]
- Templates for QCD factorization [Lin (2019)]
- EFT models [Erbin (2018)]
- Event subtraction [Butter (2019)]

Event generators

- neural importance sampling [Bothmann (2020)]
- i-flow in SHERPA [Gao (2020)]
- statistical unfolding [Bellagente (2020)]