

# How to GAN for LHC

Tilman Plehn

Universität Heidelberg

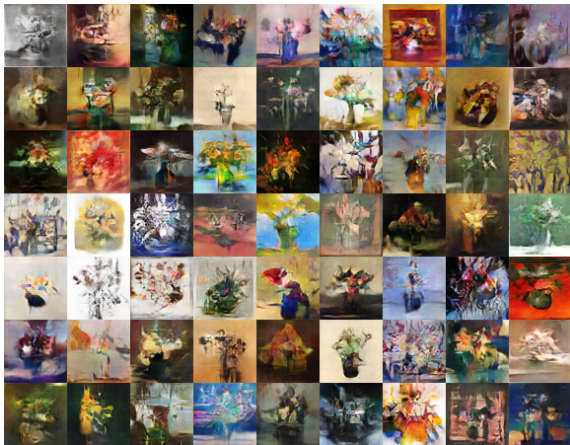
Taipei, 10/2020



# Learning from art

## GANGogh [Bonafilia, Jones Danyluk (2017)]

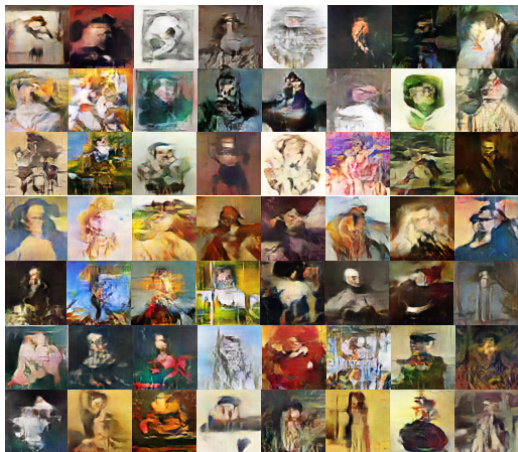
- can networks generate **something new?**
- map noise vector to images
- train on 80,000 pictures [organized by style and genre]
- generate flowers



# 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

Basics

1- Jets

2- GANpification

3- Events

4- Subtraction

5- Unfolding

6- Inverting

## 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]

- 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**

Basics

1- Jets

2- GANplication

3- Events

4- Subtraction

5- Unfolding

6- Inverting



## GAN basics (for LHC theory plumbers)

## Basics

- 1- Jets
- 2- GANplification
- 3- Events
- 4- Subtraction
- 5- Unfolding
- 6- Inverting

## Simulations crucial for LHC physics [review: Butter &amp; TP]

- 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 et al (2014)]

- training true events  $\{x_T\}$  following  $P_T$   
output generated events  $\{r\} \rightarrow \{x_G\}$  following  $P_G$
- **discriminator** constructing  $D(x)$  [ $D(x) = 1, 0$  true/generated]
 
$$L_D = \langle -\log D(x) \rangle_{x \sim P_T} + \langle -\log(1 - D(x)) \rangle_{x \sim P_G} \rightarrow -2 \log 0.5$$
- **generator** producing true-looking events [ $D$  needed]
 
$$L_G = \langle -\log D(x) \rangle_{x \sim P_G}$$

⇒ **statistically independent copy of training events**



# 1– Jet generation

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

- start with calorimeter or jet images [ $\eta$  vs  $\phi$ ]
- sparsity the technical challenge [cf top tagging comparison]

1- reproduce valid jet images from training data

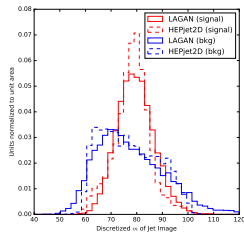
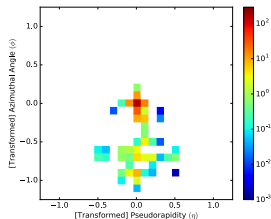
2- organize them by QCD vs  $W$ -decay jets

- high-level observables  $m, \tau_{21}$  as check

⇒ GANs generating jets

## GAN questions

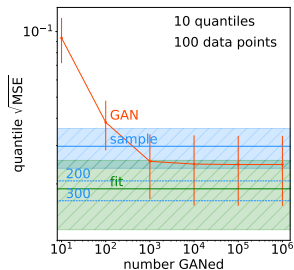
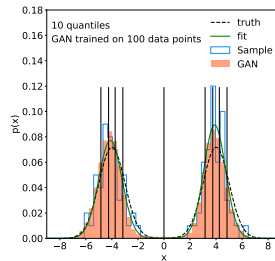
- use cases?
- control?
- uncertainty? [Bayesian networks?]
- gain?
- achievable precision?



## 2- GANplification

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

- true function known  
compare GAN vs sampling vs fit
- $\chi^2$ -goodness in quantiles



Basics

1- Jets

2- GANplification

3- Events

4- Subtraction

5- Unfolding

6- Inverting

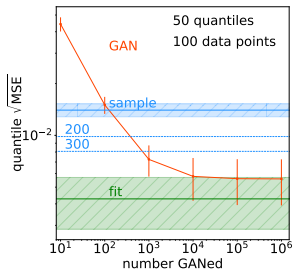
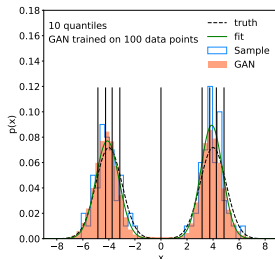




## 2- GANplification

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

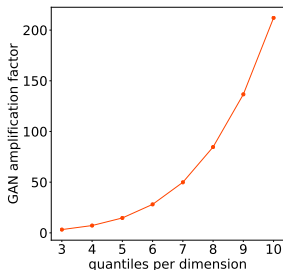
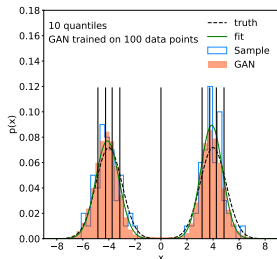
- true function known  
compare GAN vs sampling vs fit
- $\chi^2$ -goodness in quantiles
- fit like 500-1000 sampled points  
GAN like 500 sampled points [amplification factor 5]  
improvement up to 10,000 GANned events



## 2- GANplification

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

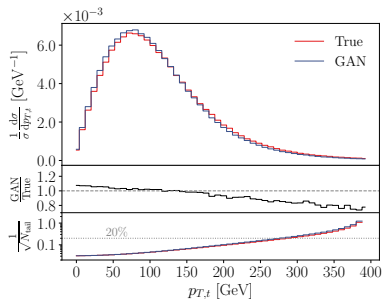
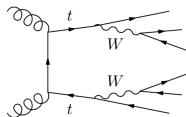
- true function known  
compare GAN vs sampling vs fit
  - $\chi^2$ -goodness in quantiles
  - fit like 500-1000 sampled points  
GAN like 500 sampled points [amplification factor 5]  
improvement up to 10,000 GANned events
  - 5-dimensional Gaussian shell  
sparsely populated  
amplification vs quantiles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



## 3– How to GAN LHC Events

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

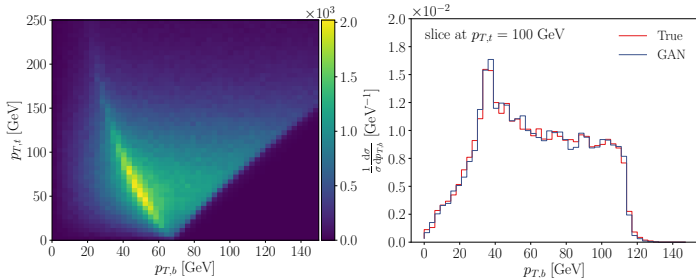
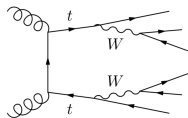
- realistic final state  $t\bar{t} \rightarrow 6$  jets [Butter, TP, Winterhalder]
- on-shell external states  $\rightarrow$  12D phase space
- top observables with tails [statistical/systematic error indicated]



## 3– How to GAN LHC Events

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

- realistic final state  $t\bar{t} \rightarrow 6$  jets [Butter, TP, Winterhalder]
- on-shell external states  $\rightarrow$  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...]

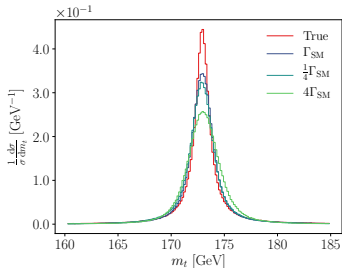
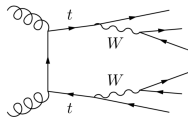
- realistic final state  $t\bar{t} \rightarrow 6$  jets [Butter, TP, Winterhalder]
- on-shell external states  $\rightarrow$  12D phase space
- top observables with tails [statistical/systematic error indicated]
- 2D correlations
- 1D-invariant masses [top, W]

batch-wise discrimination, MMD loss with kernel  $k$

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

$$L_G \rightarrow L_G + \lambda_G \text{MMD}^2$$

- 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-collinear 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-collinear subtraction, multi-jet merging

on-shell subtraction

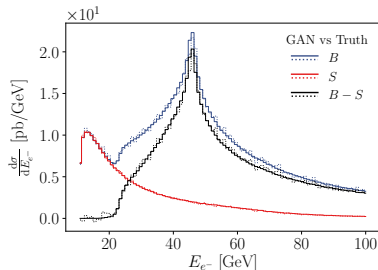
background subtraction [4-body decays]

- event-based background subtraction

$$pp \rightarrow e^+e^- \quad (\text{Base}) \quad pp \rightarrow \gamma \rightarrow e^+e^- \quad (\text{Subtracted})$$

- Z-pole events generated

⇒ Why did we ever bin?



## 5— How to GAN away detector effects

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

- detector simulation — unfolding established use case
- inversion possible, in principle [entangled convolutions]
- GAN task

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

⇒ Full phase space unfolding





## 5– How to GAN away detector effects

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

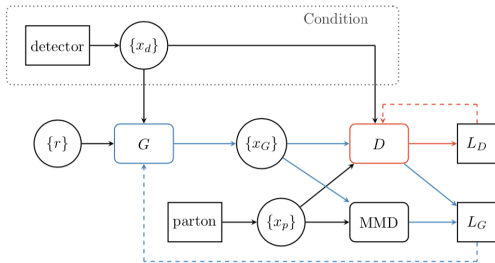
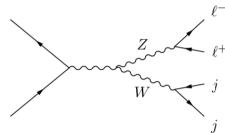
- detector simulation — unfolding established use case
- inversion possible, in principle [entangled convolutions]
- GAN task

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

⇒ Full phase space unfolding

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

- broad  $jj$  mass peak
- narrow  $\ell\ell$  mass peak
- modified  $2 \rightarrow 2$  kinematics
- (conditional) GAN like for event generation



# Fully conditional GAN

Test data modified from training data

– full inversion no point in showing...

Basics

1- Jets

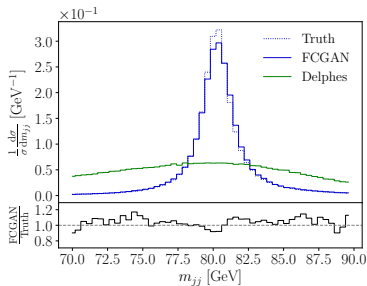
2- GANplification

3- Events

4- Subtraction

5- Unfolding

6- Inverting



## Fully conditional GAN

## Test data modified from training data

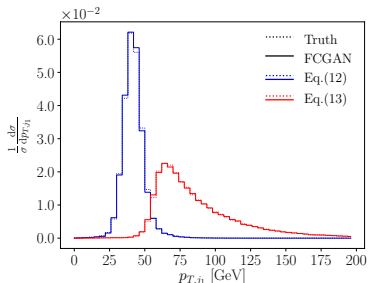
– full inversion no point in showing...

– test cuts [14%, 39% events]

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

$$p_{T,j_1} > 60 \text{ GeV} \quad (13)$$

⇒ Phase space unfolding working



## Fully conditional GAN

## Test data modified from training data

– full inversion no point in showing...

– test cuts [14%, 39% events]

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

$$p_{T,j_1} > 60 \text{ GeV} \quad (13)$$

⇒ Phase space unfolding working

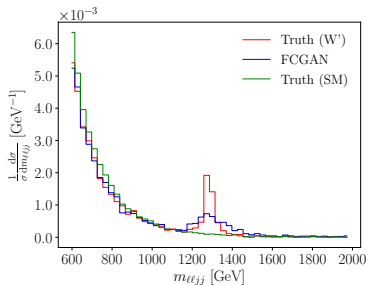
## New physics in data [model dependence]

– train: Standard Model events

test: 10% events with  $W'$  in s-channel

– nightmare: unfold  $W'$  onto Standard Model?

⇒ Model dependence under control



## 6– Unfolding as inverting

**Invertible networks** [Bellagente, Butter, Kasieczka, TP, Rousselot, Winterhalder, Ardizzone, Köthe]

- network as bijective transformation — normalizing flow
- Jacobian tractable — normalizing flow
- evaluation in both directions — INN [Ardizzone, Rother, Köthe]

Basics

1- Jets

2- GANplification

3- Events

4- Subtraction

5- Unfolding

6- Inverting



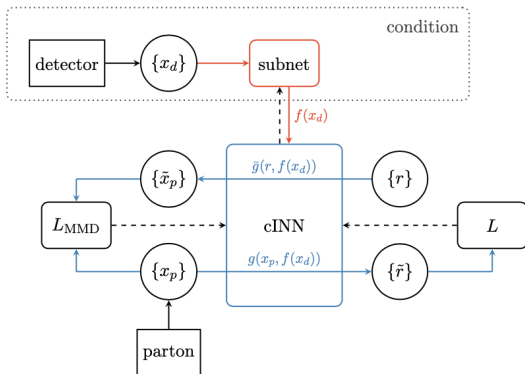
## 6– Unfolding as inverting

## Invertible networks [Bellagente, Butter, Kasieczka, TP, Rousselot, Winterhalder, Ardizzone, Köthe]

- network as bijective transformation — normalizing flow
- Jacobian tractable — normalizing flow
- evaluation in both directions — INN [Ardizzone, Rother, Köthe]

## Conditional INN

- same procedure as for GAN
- parton-level events from random numbers



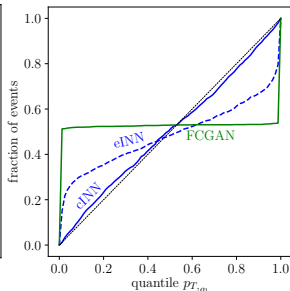
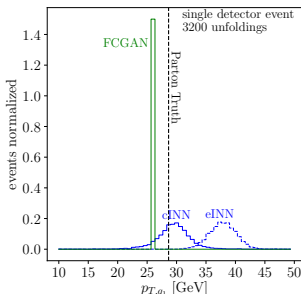
## 6– Unfolding as inverting

## Invertible networks [Bellagente, Butter, Kasieczka, TP, Rousselot, Winterhalder, Ardigzone, Köthe]

- network as bijective transformation — normalizing flow
- Jacobian tractable — normalizing flow
- evaluation in both directions — INN [Ardizzone, Rother, Köthe]

## Conditional INN

- same procedure as for GAN
- parton-level events from random numbers
- calibration for statistical unfolding



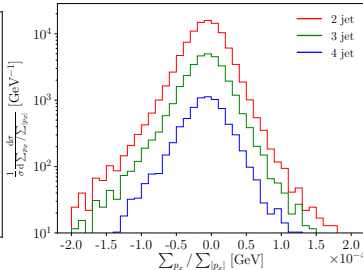
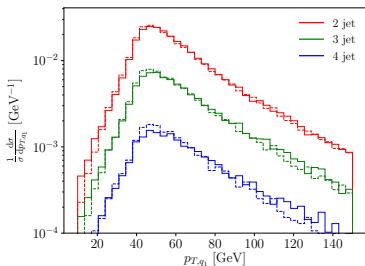
## 6– Unfolding as inverting

## Invertible networks [Bellagente, Butter, Kasieczka, TP, Rousselot, Winterhalder, Ardizzone, Köthe]

- network as bijective transformation — normalizing flow
- Jacobian tractable — normalizing flow
- evaluation in both directions — INN [Ardizzone, Rother, Köthe]

## Unfolding extra jets

- detector-level process  $pp \rightarrow ZW+\text{jets}$  [variable number of objects]
- parton-level hard process chosen  $2 \rightarrow 2$  [whatever you want]
- ME vs PS jets decided by network [including momentum conservation]



⇒ Proper statistical inversion and of QCD





# 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 [Ottens (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)]

