New and Cool LHC Stuff

How to GAN

Tilman Plehn

Universität Heidelberg

PHENO 5/2020

Learning from art

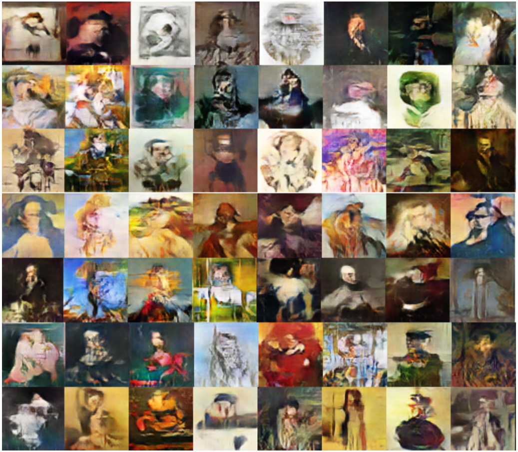
GANGogh [Bonafilia, Jones Danyluk (2017)]

- old news: NNs turning pictures into art of a certain epoch
  but can they create new pieces of art?
- train on 80,000 pictures [organized by style and genre]
- map noise vector to images
- generate flowers

Learning from art

GANGogh [Bonafilia, Jones Danyluk (2017)]

- old news: NNs turning pictures into art of a certain epoch
  but can they create new pieces of art?
- train on 80,000 pictures [organized by style and genre]
- map noise vector to images
- generate portraits



Learning from art

GANGogh [Bonafilia, Jones Danyluk (2017)]

- old news: NNs turning pictures into art of a certain epoch
  but can they create new pieces of art?
- train on 80,000 pictures [organized by style and genre]
- map noise vector to images

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

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

⇒ all about marketing and sales

GAN algorithm

GANning events

- training true events $\{x_T\}$ following $p_T(x)$
  output generated events $\{r\} \to \{x_G\}$ following $p_G(x)$
- discriminator constructing $D(x)$ [$D(x) = 1, 0$ truth/generated]

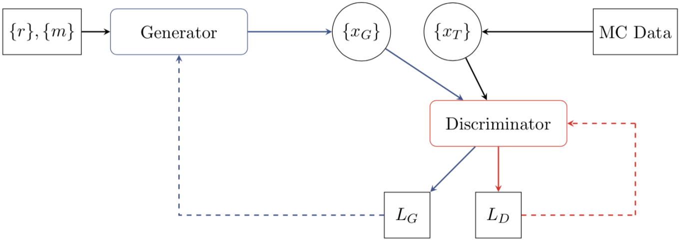
$$L_D = \big\langle -\log D(x) \big\rangle_{x \sim P_T} + \big\langle -\log(1 - D(x)) \big\rangle_{x \sim P_G} \to -2 \log 0.5$$

- generator producing good events [D needed]

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

- stabilization: gradient penalty or WassersteinGAN

$\Rightarrow$ statistically independent copy of training events



GANs at LHC

Phase space networks

- MC integration [Bendavit (2017)]
- NNVegas [Klimek (2018), not really generative network]

Existing GAN studies

- Jet Images [de Oliveira (2017), Carazza (2019)]
- Detector simulations [Paganini (2017), Musella (2018), Erdmann (2018), Ghosh (2018), Buhmann (2020)]
- Event generation [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

- generative invertable networks without generation or inversion
- neural importance sampling [Bothmann (2020)]
- i-flow in SHERPA [Gao (2020)]

1– Jet generation

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

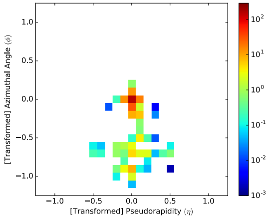
– start with calorimeter images or jet images [η vs ϕ]
  sparsity the technical challenge [cf top tagging comparison]

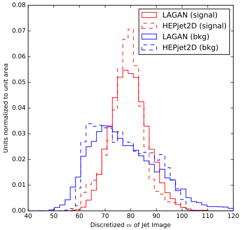
1- reproduce valid jet images from training data

2- organize them by QCD vs W-decay jets

– high-level observables m, τ_{21} reproduced

$\Rightarrow$ GANs can generate jets





1– Jet generation

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

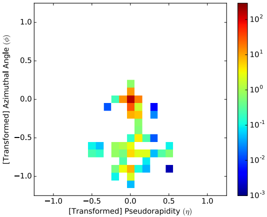
– start with calorimeter images or jet images [η vs ϕ]
  sparsity the technical challenge [cf top tagging comparison]

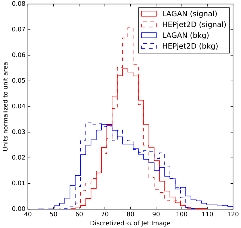
1- reproduce valid jet images from training data

2- organize them by QCD vs W-decay jets

– high-level observables m, τ_{21} reproduced

$\Rightarrow$ GANs can generate jets





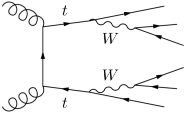
Open questions to all GANs

– use cases?

– uncertainty? [Bayesian networks?]

– achievable precision?

2– How to GAN LHC Events

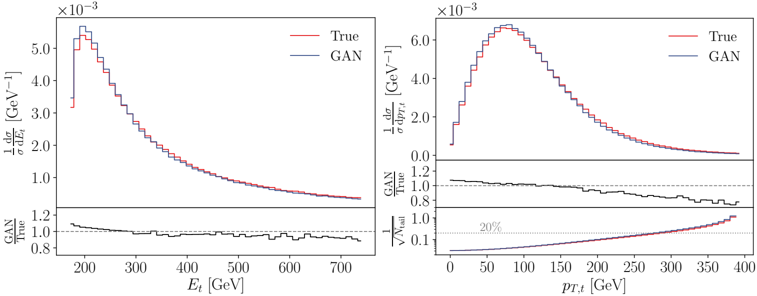


Idea: replace ME for hard process [Otten, Hashemi, Di Sipio...]

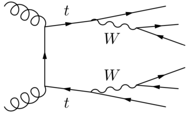
- medium-complex final state $t\bar{t} \to 6$ jets [Butter, TP, Winterhalder]

  $t/\bar{t}$ and $W^{\pm}$ on-shell with BW $6 \times 4 = 18$ dof

  on-shell external states $\to 12$ dof [constants hard to learn]
- flat observables flat [phase space coverage okay]
- constructed observables with tails [statistical error indicated]

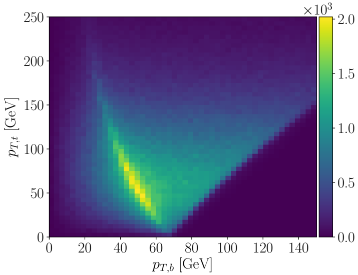


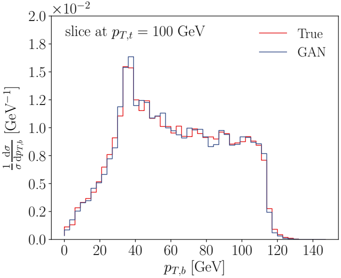
2– How to GAN LHC Events



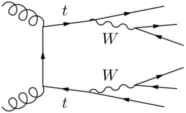
Idea: replace ME for hard process [Otten, Hashemi, Di Sipio...]

- medium-complex final state $t\bar{t} \to 6$ jets [Butter, TP, Winterhalder]
  $t/\bar{t}$ and $W^{\pm}$ on-shell with BW $6 \times 4 = 18$ dof
  on-shell external states $\to 12$ dof [constants hard to learn]
- flat observables flat [phase space coverage okay]
- constructed observables with tails [statistical error indicated]
- 2D correlations





2– How to GAN LHC Events



Idea: replace ME for hard process [Otten, Hashemi, Di Sipio...]

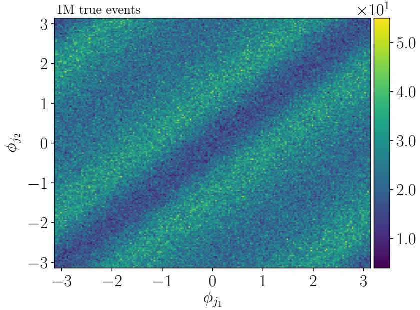
– medium-complex final state $t\bar{t} \to 6$ jets [Butter, TP, Winterhalder]
  $t/\bar{t}$ and $W^{\pm}$ on-shell with BW $6 \times 4 = 18$ dof
  on-shell external states $\to$ 12 dof [constants hard to learn]

– flat observables flat [phase space coverage okay]

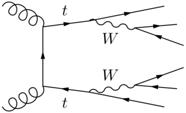
– constructed observables with tails [statistical error indicated]

– 2D correlations

– improved resolution [1M training events]

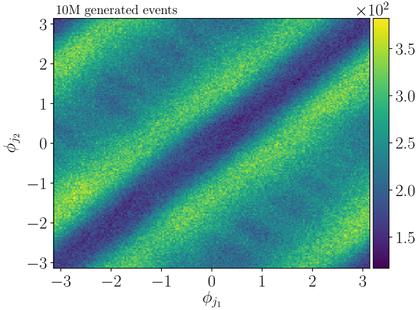


2– How to GAN LHC Events

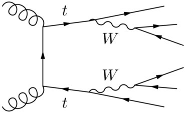


Idea: replace ME for hard process [Otten, Hashemi, Di Sipio...]

- medium-complex final state $t\bar{t} \to 6$ jets [Butter, TP, Winterhalder]
  $t/\bar{t}$ and $W^\pm$ on-shell with BW $6 \times 4 = 18$ dof
  on-shell external states $\to 12$ dof [constants hard to learn]
- flat observables flat [phase space coverage okay]
- constructed observables with tails [statistical error indicated]
- 2D correlations
- improved resolution [10M generated events]



2– How to GAN LHC Events



Idea: replace ME for hard process [Otten, Hashemi, Di Sipio...]

– medium-complex final state $t\bar{t} \to 6$ jets [Butter, TP, Winterhalder]
$t/\bar{t}$ and $W^\pm$ on-shell with BW $6 \times 4 = 18$ dof
on-shell external states $\to 12$ dof [constants hard to learn]

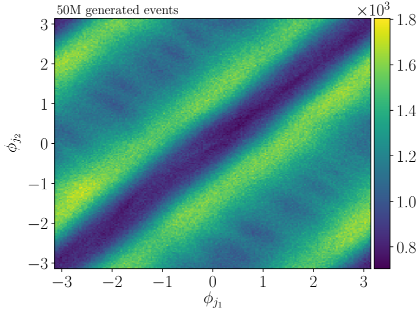
– flat observables flat [phase space coverage okay]

– constructed observables with tails [statistical error indicated]

– 2D correlations

– improved resolution [50M generated events]

– GAN generation working



Intermediate resonances

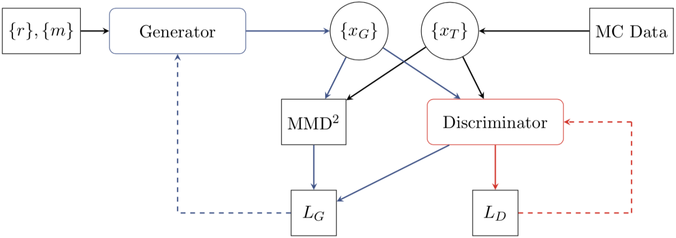
Narrow phase space structures

- MC: phase space mapping [BW $\to$ flat, multi-channel]
- generally 1D features

  cuts and phase space boundaries
  invariant masses [top, W]
- batch-wise comparison of distributions, MMD loss with kernel k

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

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



Intermediate resonances

Narrow phase space structures

– MC: phase space mapping [BW $\to$ flat, multi-channel]

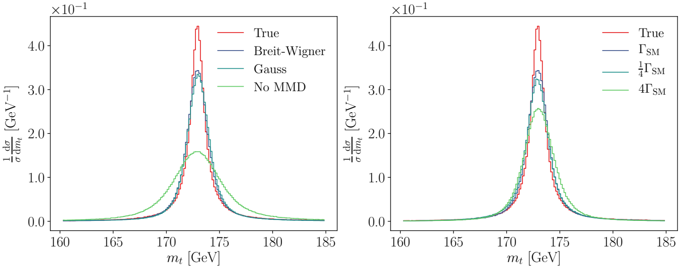
– generally 1D features

cuts and phase space boundaries
invariant masses [top, W]

– batch-wise comparison of distributions, MMD loss with kernel k

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

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



$\Rightarrow$ Phase space resolution no show-stopper

3– How to GAN event subtraction

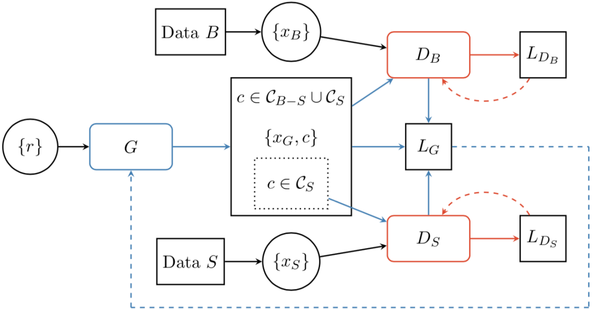
Idea: subtract event samples without bins [Butter, TP, Winterhalder]

– statistical uncertainty

$$\Delta_{B-S} = \Delta_{n_B N_B - n_S N_S} = \sqrt{\Delta^2_{n_B N_B} + \Delta^2_{n_S N_S}} = \sqrt{n_B^2 N_B + n_S^2 N_S} > \max(B, S)$$

– many applications

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



3– How to GAN event subtraction

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

– statistical uncertainty

$$\Delta_{B-S} = \Delta_{n_B N_B - n_S N_S} = \sqrt{\Delta^2_{n_B N_B} + \Delta^2_{n_S N_S}} = \sqrt{n_B^2 N_B + n_S^2 N_S} > \max(B, S)$$

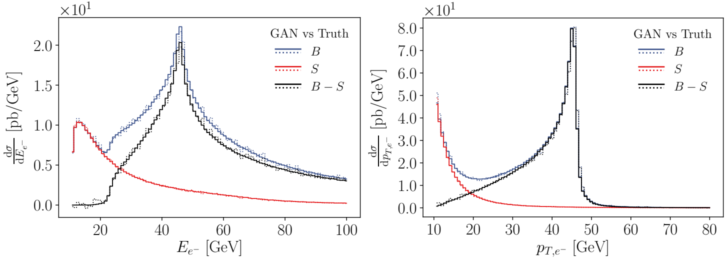
– many applications

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

– event-based background subtraction [weird notation, sorry]

$$pp \to e^+ e^- \quad \text{(B)} \qquad pp \to \gamma \to e^+ e^- \quad \text{(S)}$$

– Z-pole events generated



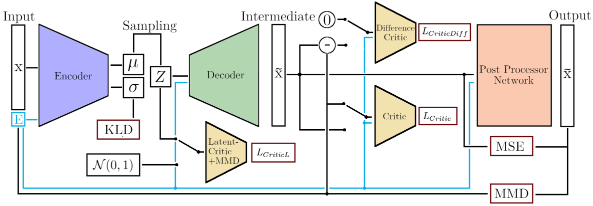
4– Detector simulation

Fast detector simulation [Paganini, Musella, Erdmann, Ghosh, Buhmann,...]

– weakest link in simulation chain
  fast simulation established problem [ATLfast, Delphes,...]

– training on GEANT4

– comparison of GAN, WGAN, new BIB-AE



4– Detector simulation

Fast detector simulation [Paganini, Musella, Erdmann, Ghosh, Buhmann,...]

– weakest link in simulation chain
  fast simulation established problem [ATLfast, Delphes,...]

– training on GEANT4

– comparison of GAN, WGAN, new BIB-AE

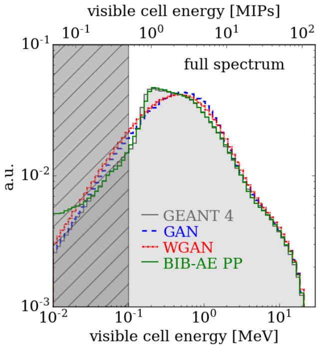
Recent ILD study using particle flow [Buhmann]

– 950k photon showers $E = 10 \dots 100$ GeV

– challenge:
  get entire spectrum
  maintain correlations

– MMD post-processing, transfer learning

$\Rightarrow$ Not easy, but working



5– How to GAN away detector effects

Open problem of publishing kinematic information [e.g. global SMEFT analyses]

- total rates losing information
  best STXS model-dependent
  unfolded distributions extremely convenient [$t\bar{t}$ results]
- challenges in unfolding
  non-invertible detector simulation
  model dependence
  flexibility/reliability [training on some event set]
- benefits from unfolding data [Omnifold]
  access to hard matrix element/first-principles QCD
  matrix element method

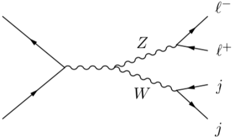
General: how to invert Markov processes [Datta; Bellagente, Butter, Kasiczka, TP, Winterhalder]

- detector simulation typical Markov process
- inversion possible, in principle [entangled convolutions]
- GAN task

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

$\Rightarrow$ Full unfolded phase space

Fully conditional GAN



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

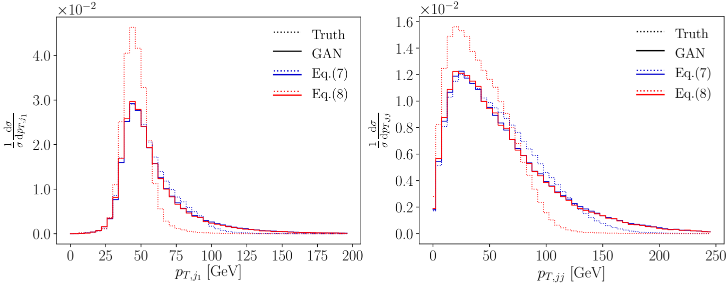
– broad jj mass peak
  narrow $\ell\ell$ mass peak
  modified $2 \to 2$ kinematics

– GAN same as event generation [with MMD]

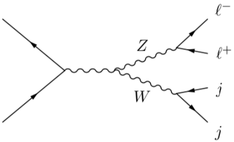
– problem: cuts in test data [88%, 38% events]

$$p_{T,j_1} = 30 \dots 100\ \text{GeV} \tag{7}$$

$$p_{T,j_1} = 30 \dots 60\ \text{GeV} \quad \text{and} \quad p_{T,j_2} = 30 \dots 50\ \text{GeV} \tag{8}$$



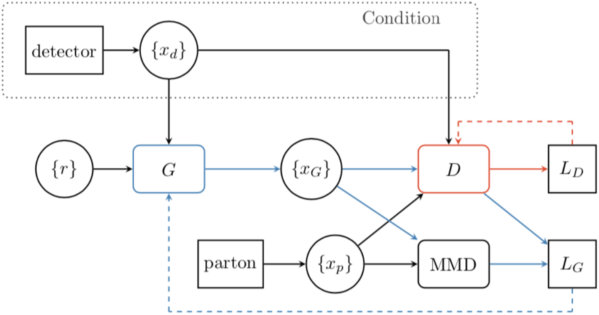
Fully conditional GAN



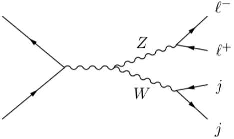
Reconstructing parton level $pp \to ZW \to (\ell\ell)\ (jj)$

- broad jj mass peak
  narrow $\ell\ell$ mass peak
  modified $2 \to 2$ kinematics
- GAN same as event generation [with MMD]

Conditional GAN



Fully conditional GAN

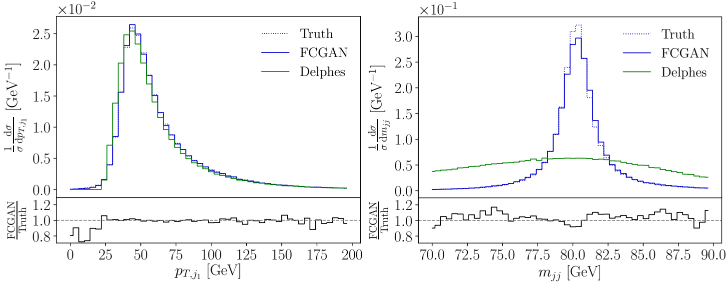


Reconstructing parton level $pp \to ZW \to (\ell\ell)\ (jj)$

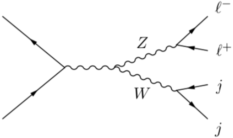
- broad jj mass peak
  narrow $\ell\ell$ mass peak
  modified $2 \to 2$ kinematics
- GAN same as event generation [with MMD]

Conditional GAN

- full inversion fine



Fully conditional GAN



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

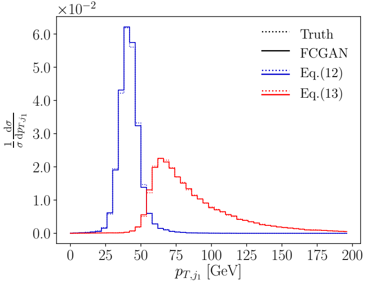
- broad jj mass peak
  narrow $\ell\ell$ mass peak
  modified $2 \to 2$ kinematics
- GAN same as event generation [with MMD]

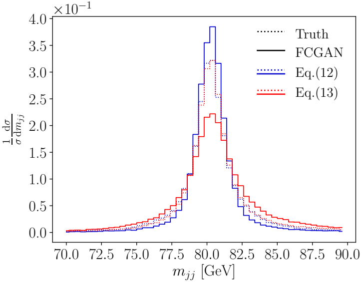
Conditional GAN

- full inversion fine
- tougher cuts challenging m_{jj} [14%, 39% events, no interpolation, MMD not conditional]

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

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





⇒ FCGAN unfolding works...

BSM Injection

Different training (MC) and actual data... [not in v1, thank you to Ben Nachman]

...or model dependence of unfolding

...or localization in latent space

– train: SM events
  test: 10% events with W' in s-channel $\Rightarrow$ Any guesses?

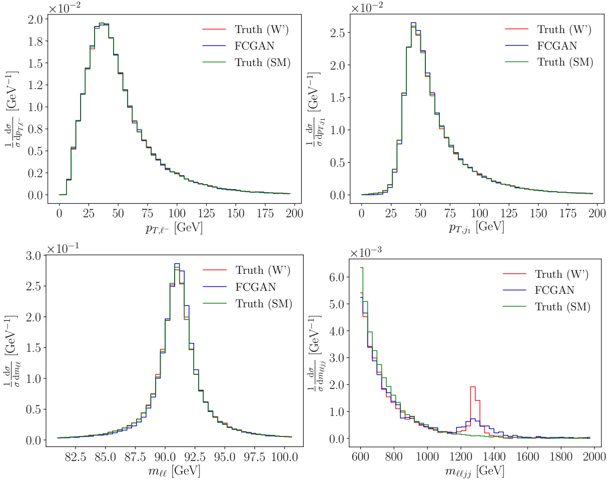
BSM Injection

Different training (MC) and actual data... [not in v1, thank you to Ben Nachman]

...or model dependence of unfolding

...or localization in latent space

– train: SM events
  test: 10% events with W' in s-channel ⇒ Any guesses?



6– Superresolution GANs (preview)

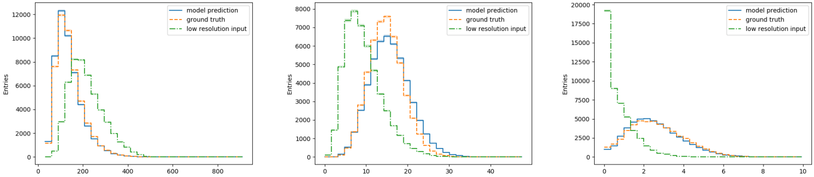
Getting inspired [Blecher, Butter, Keilbach, TP + Irvine]

– take high-resolution calorimeter images
down-sample to 1/8th 1D resolution
GAN inversion

– start from low-resolution calorimeter images
GAN high-resolution images

– works because GANs learn structure [showers are QCD]

– energy of constituents no.1,10,30



⇒ GANs are (kind of) magic

Outlook

LHC physics really is big data

jet classification largely established

advantages of generative networks [upcoming review: Butter & TP]

1: NN interpolation

2: latent space structures

3: training on MC and/or data

open questions

1: uncertainties

2: possible precision...