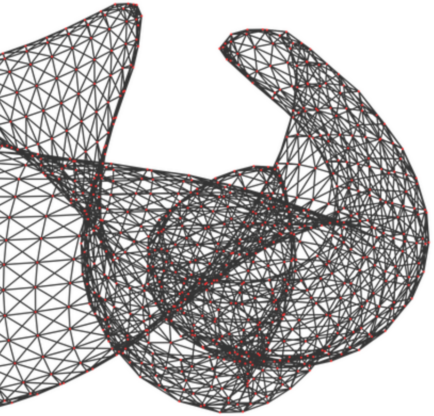
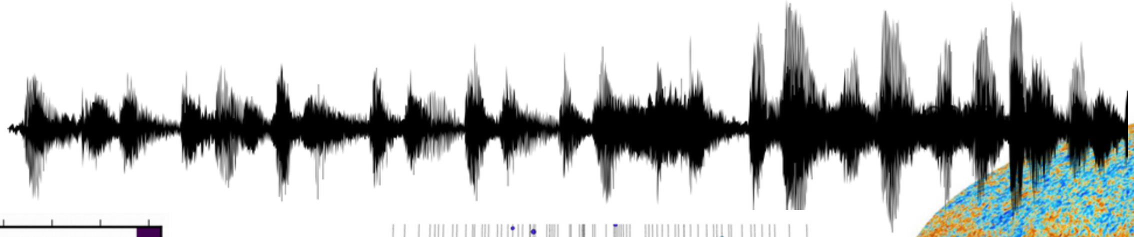
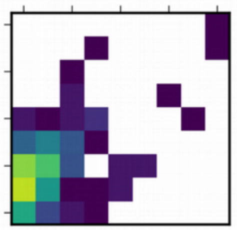
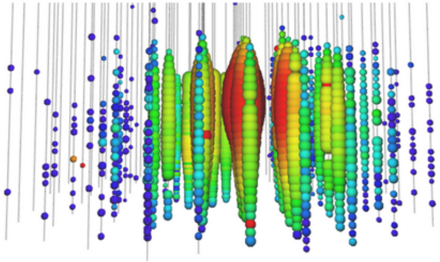
Deep learning for astroparticle physics



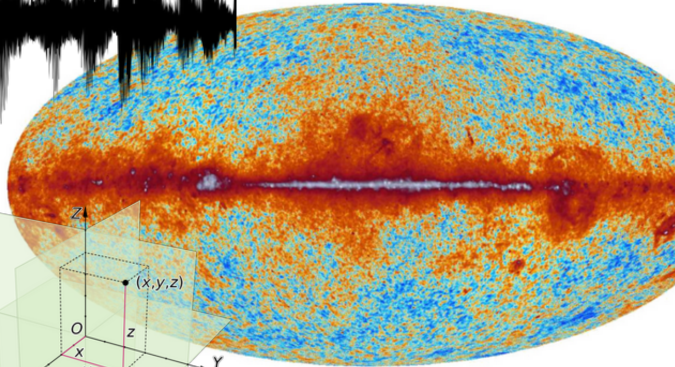




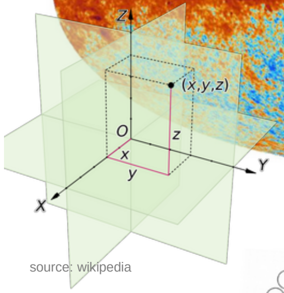
10.1016/j.astropartphys.2018.10.003



https://arxiv.org/abs/1309.7003



Astronomy and Astrophysics 641, p. 1 (2018)



source: wikipedia

Jonas Glombitza, *Martin Erdmann*

February 2022, www.deeplearningphysics.org/

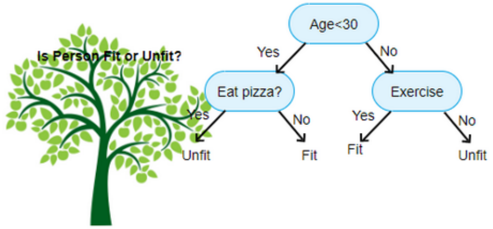
Workshop on Machine Learning for Cosmic-Ray Air Showers, Delaware

SPONSORED BY THE
Federal Ministry of Education and Research

Machine Learning and Deep Learning

Machine Learning

- applications across many physics domains, e.g., for (background rejection, multi-class classifications)
- BDTs, random forest, shallow NNs



https://www.aitimejournal.com/@akshay.chavan/a-comprehensive-guide-to-decision-tree-learning

Deep Learning

- field driven by computer science (BigTechs)
- major improvements in:
  - speech recognition, NLP
  - pattern recognition, CV
- (usually) requires huge amounts of data

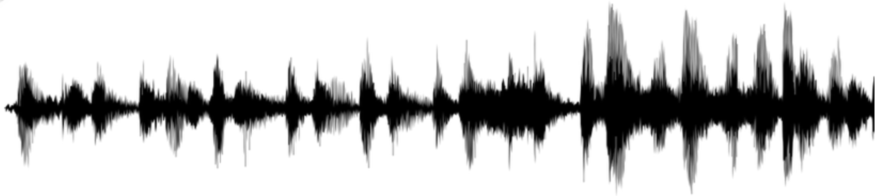
KÜNSTLICHE INTELLIGENZ

Schlau in zwei Stunden

VON ALEXANDER ARMBRUSTER - AKTUALISIERT AM 27.09.2017 -

www.faz.net

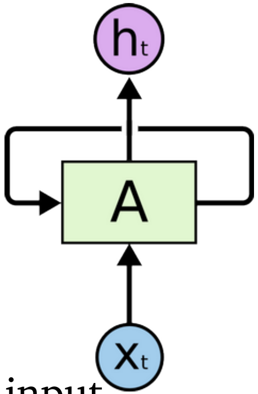
Deep Learning: RNNs & CNNs



Recurrent Networks (RNNs)

- analyze sequential data (translation)
- recurrent definition of transformation

$$h^{(t)} = A(h^{(t-1)}, x^{(t)})$$



https://colah.github.io/

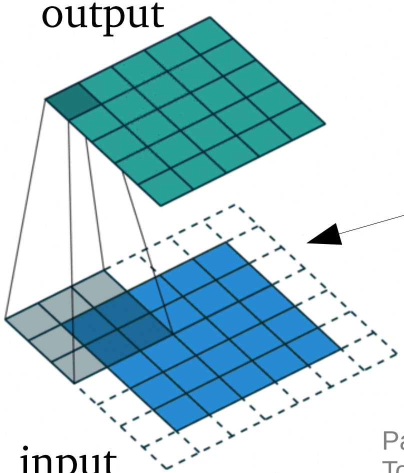
Advanced concept: LSTM

features memory

- long-range correlations

Convolutional Networks (CNNs)

- **filter** exploits image-like data
  - features translational invariance
  - prior on local correlations
  - fast on GPUs
  - (# param. independent of input size)

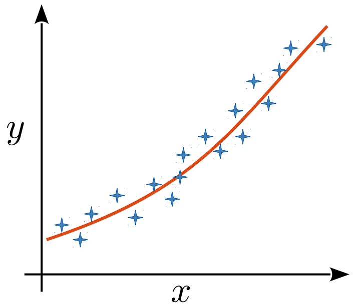


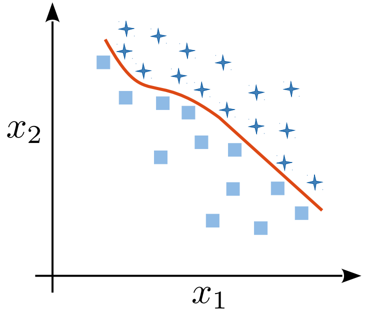
w_1	w_2	w_3
w_4	w_5	w_6
w_7	w_8	w_9

adaptive parameters of **filter**

Paul-Louis Pröve, Towards Data Science

Applications in Physics Research





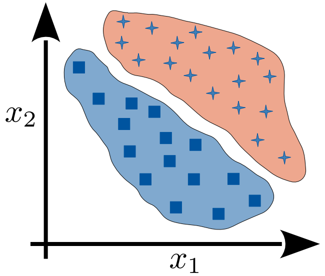
I. Supervised learning

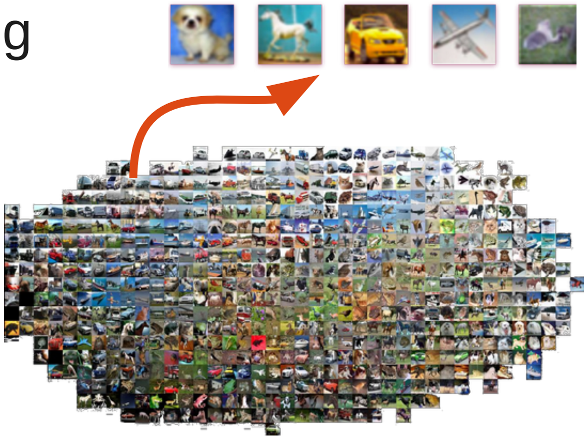
Object reconstruction

- classification
- inference of variables
- de-noising
- reconstruction of complex structures (e.g., tracks)

II. Unsupervised learning

- embedding / clustering
- generative models
- domain adaption





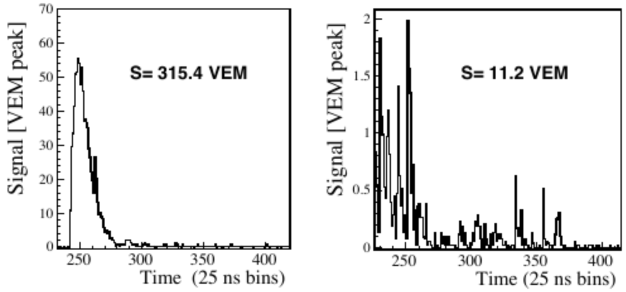
III. Reinforcement learning

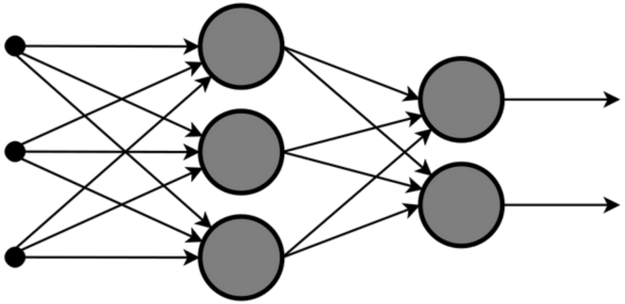
- first progress: sorting, tracking

© nature

From Classic Machine Learning to Deep Learning

- Air shower signals measured by surface detectors
  - disentangle muonic and em part at station level





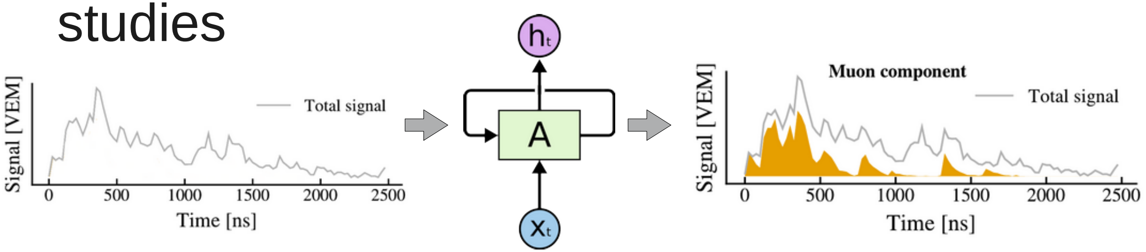
Traditional ML approach

- Extract fraction of muons measured by single station
- Feed physicist observables into a neural network

A. Gulillen et al., 10.1016/j.astropartphys.2019.03.001

Deep learning version

- Use RNN to extract time-dependent signals induced by muons
- Promising results for mass composition studies



Pierre Auger Collaboration, JINST 16 P07016 (2021)

Denoising of Signal Traces (1D)

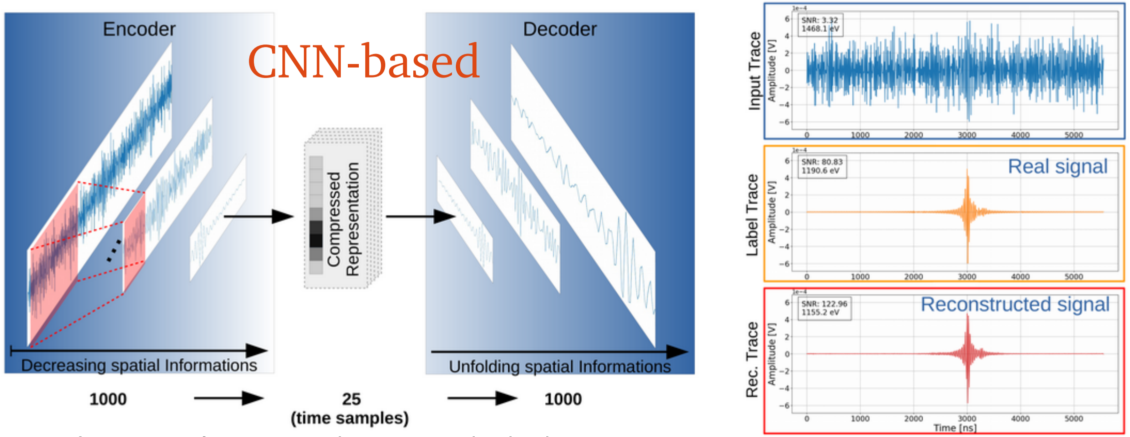
Supervised training of denoising autoencoders

- feature compressed space in between encoder and decoder

➔ encodes only relevant information in compressed space

Future application: brining ML close to the sensor

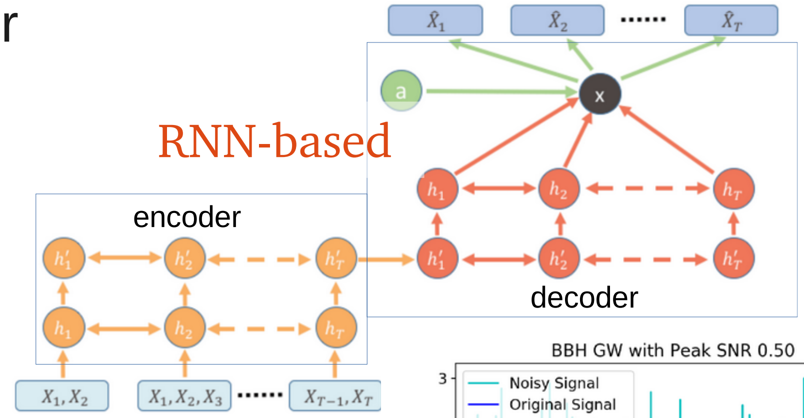
Denoising of cosmic ray radio signals

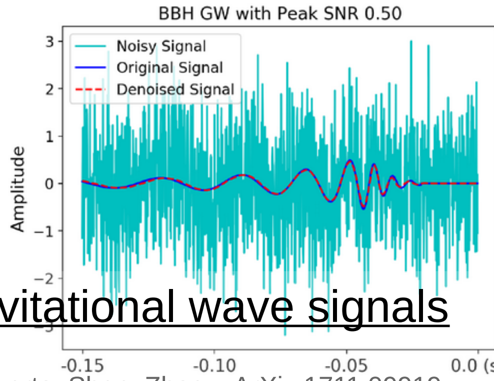


M. Erdmann et al. - 10.1088/1748-0221/14/04/P04005

A. Rehman et al., PoS ICRC2021 417

P. Bezyazeekov et al., ArXiv/2101.02943 & D. Shipilov et al., EPJ (2019) 02003



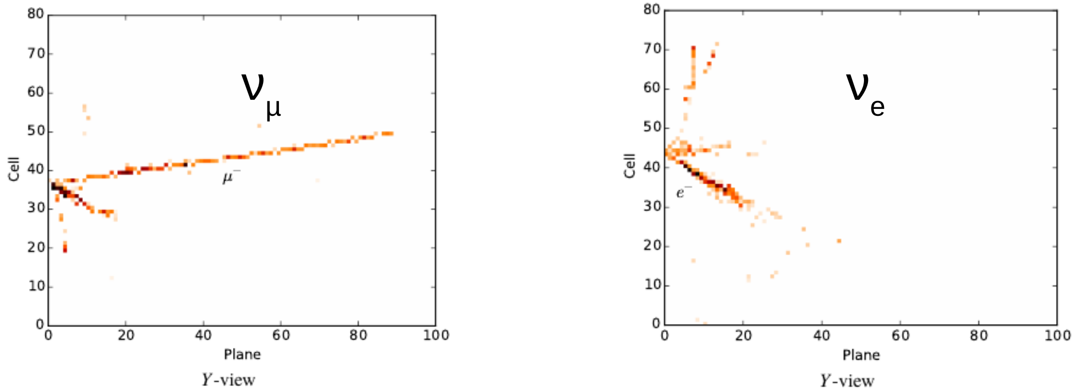


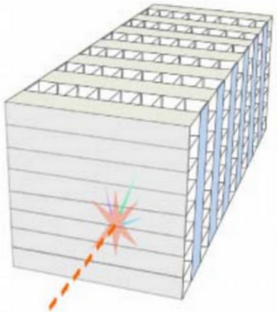
Denoising of gravitational wave signals

George, Huerta, Shen, Zhao – ArXiv 1711.09919

Electron Neutrino Identification at NovA

- First CNN application in physics at detector level
  - classification of neutrino type using image-like data
  - input: 2D projection of 3D detector
- Based on inception modules
  - apply various convolutions at different scales, in parallel



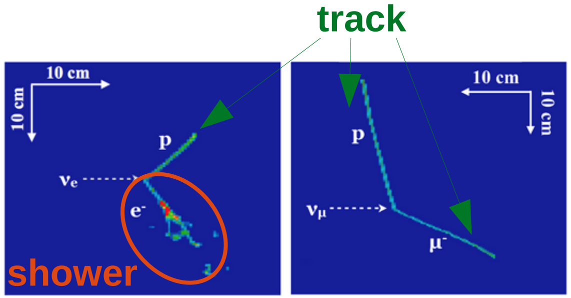


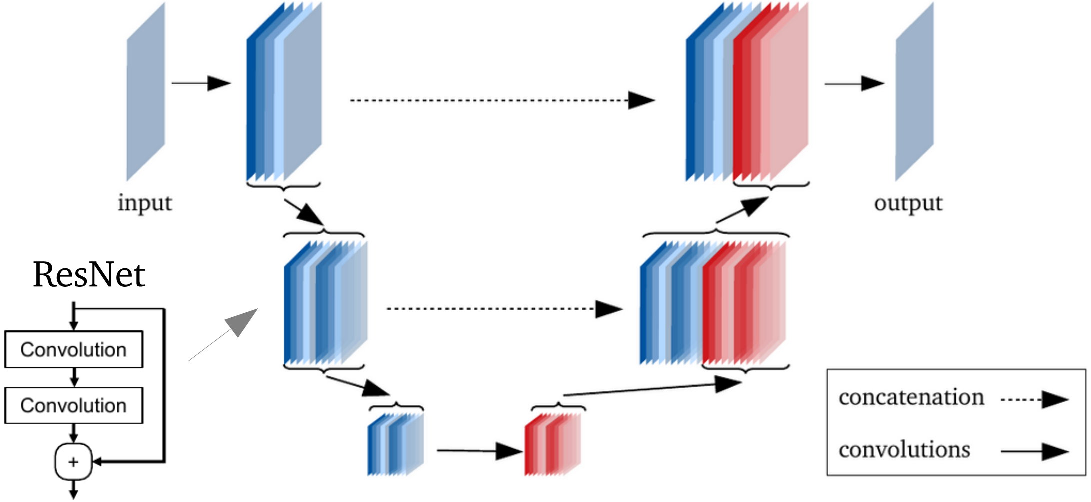
- At same purity, efficiency increased significantly (ν_e CC)
  35% (physicist algorithms) → 49% (CNN)

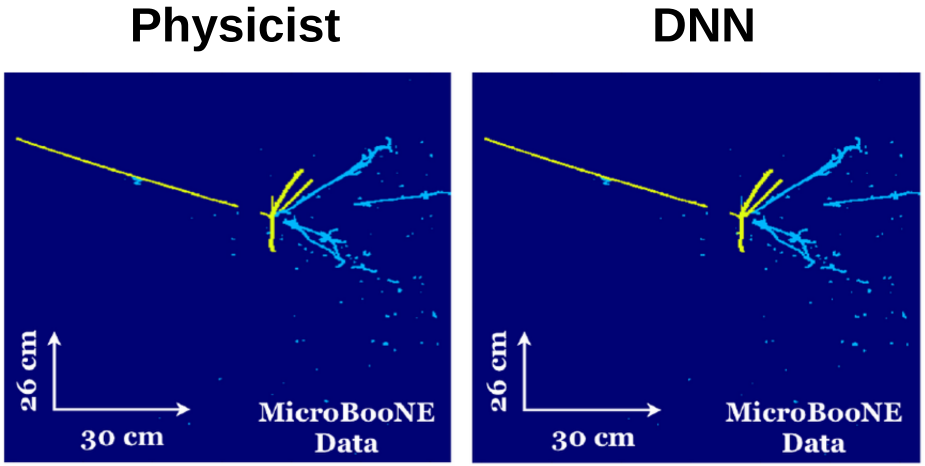
A. Aurisano et al., JINST 11 (2016) P09001

Segmentation - MircroBooNE

- Liquid Argon TPC for neutrino detection
- Segmentation (pixel-wise class prediction) into tracks and electromagnetic-showers
  - Architecture: combination of ResNet and U-Net
- Incorrectly classified pixel fraction per image ~ few percent



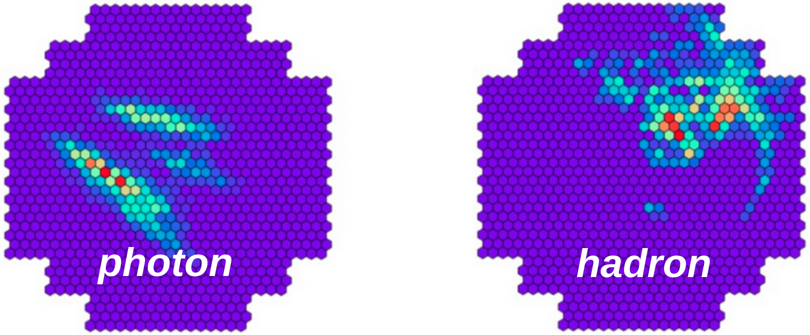


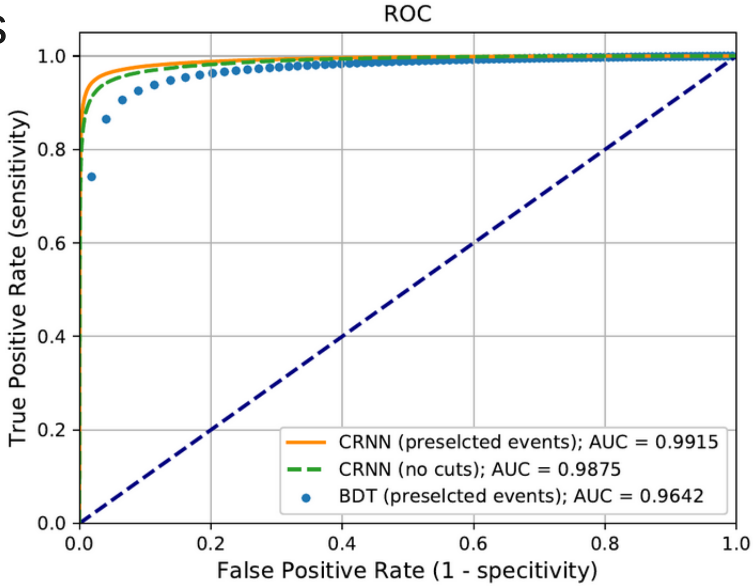


Adams et al. ArXiv: 1808.07269

Classification: H.E.S.S.

- Gamma ray telescopes in Namibia
  - background rejection (hadrons / photons)
- Hybrid approach, combining (CNNs & RNNs)
  - CNN output fed into LSTM, reflects sequential order of telescope measurements
- Network outperforms BDT



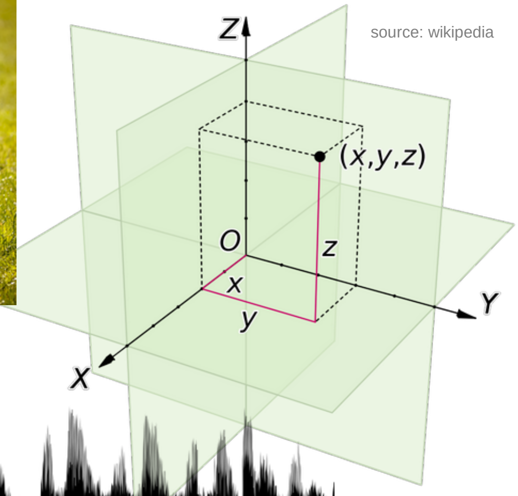


Further Developments: Deep Learning for IACT - CTlearn: https://github.com/ctlearn-project/ctlearn

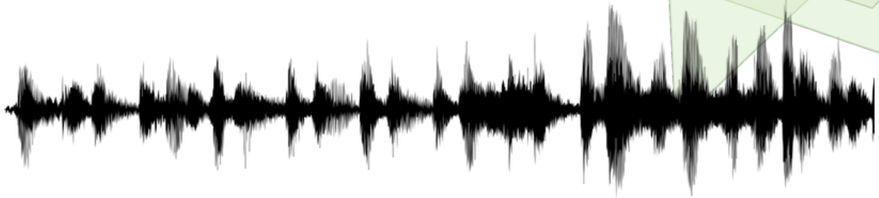
Shilon et al. - 10.1016/j.astropartphys.2018.10.003

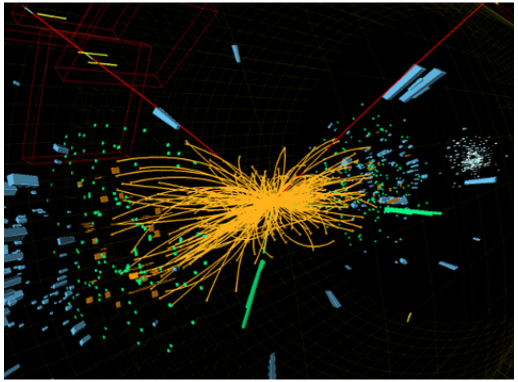
CNNs and Physics Datasets

- CNNs powerful in homogeneous and discrete (pixelized) euclidean space

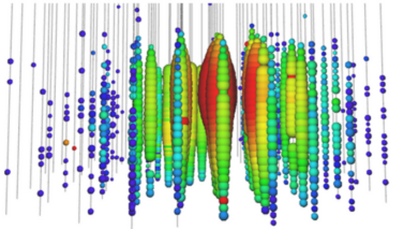


source: wikipedia



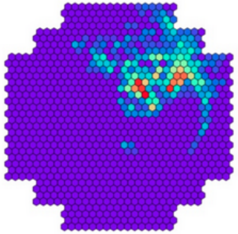


https://cds.cern.ch/record/2711418

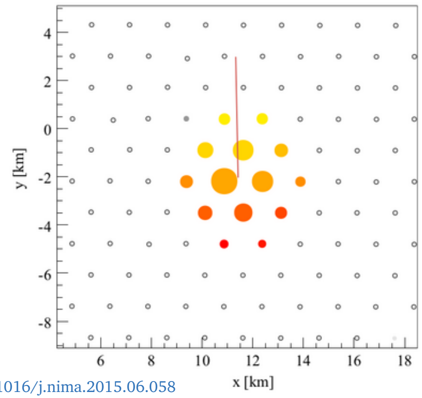


https://arxiv.org/abs/1309.7003

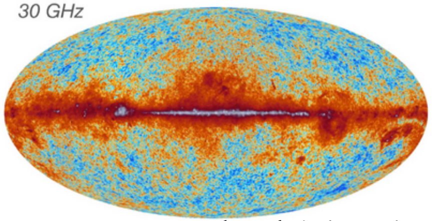
- physics data often feature different geometries



10.1016/j.astropartphys.2018.10.003



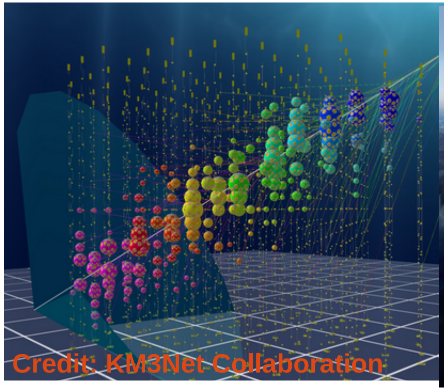
10.1016/j.nima.2015.06.058

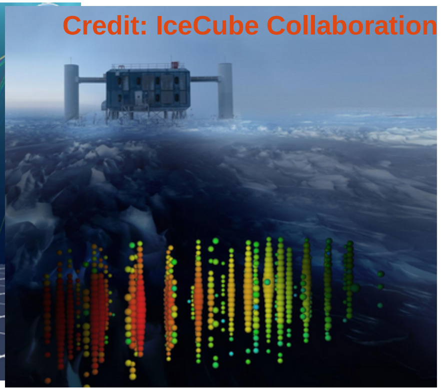


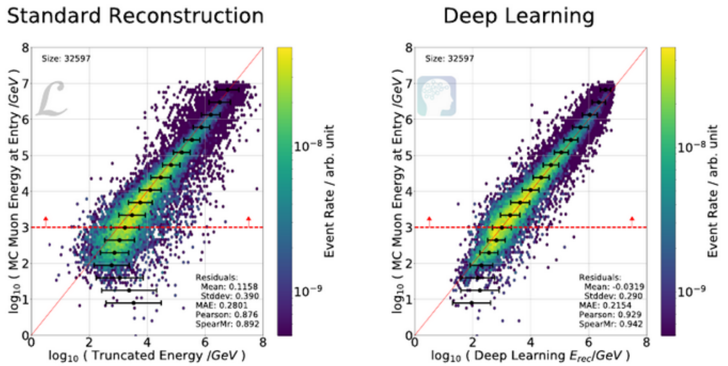
Astronomy and Astrophysics 641, p. 1 (2018)

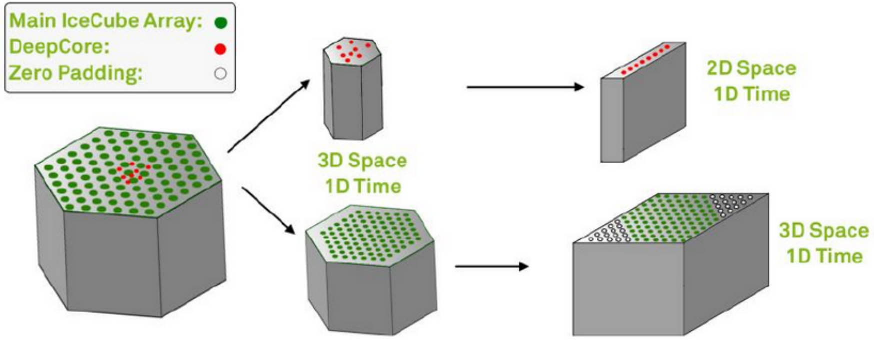
Beyond 2D Inputs

- Neutrino observatories measurements (4D)
  - position (x,y,z), time, PMTs (channels)
  - most frameworks, limited to 3D-Conv
  - projections are performed (integrating over time, or PMT)
- Reconstruction of events
  - compatible/better than comparative approaches





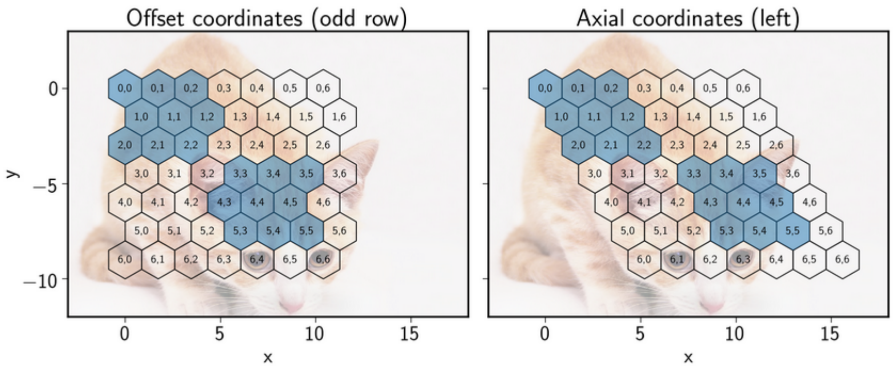


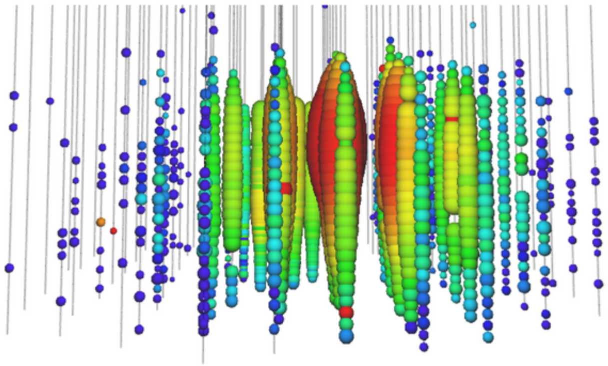


[1] M. Hünnefeld, ICRC17 – 10.22323/1.301.1057
[2] A. Aiello et al., JINST 15 (2020) P10005
[3] R. Abbasi et al., JINST 16 (2021) P07041

Hexagonal Grids

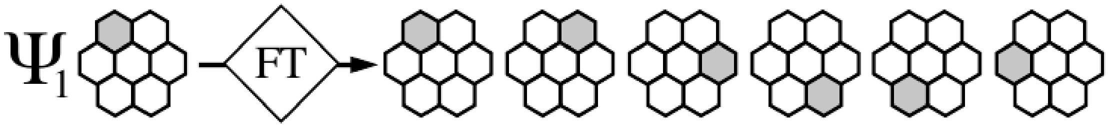
- Most astroparticle detectors feature hexagonal grids
  - change of coordinate system for indexing

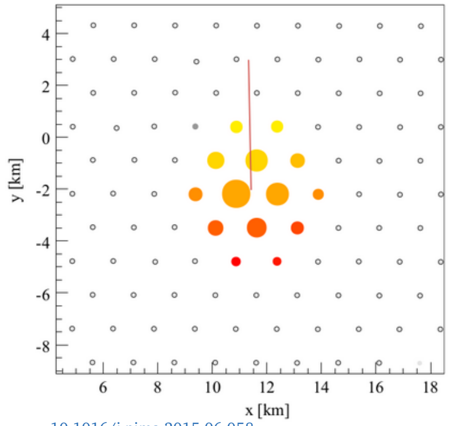




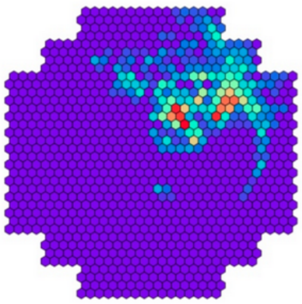
https://arxiv.org/abs/1309.7003

- Exploit rotational symmetry
  - extent convolution using orientation channels





10.1016/j.nima.2015.06.058

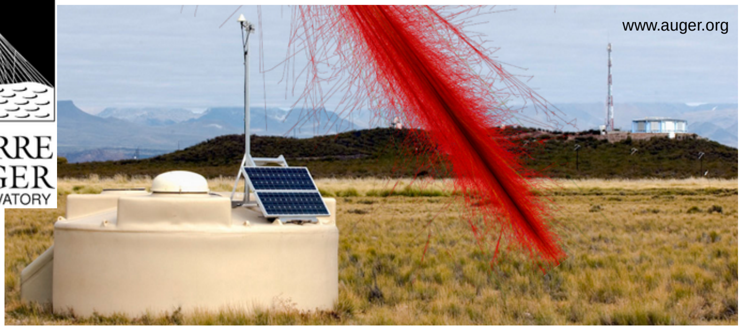


10.1016/j.astropartphys.2018.10.003

E. Hoogeboom, J. Peters, T. Cohen, M. Welling: ArXiv/1803.02108

Air-Shower Reconstruction

The Pierre Auger Collaboration, JINST 16 P07019 (2021)



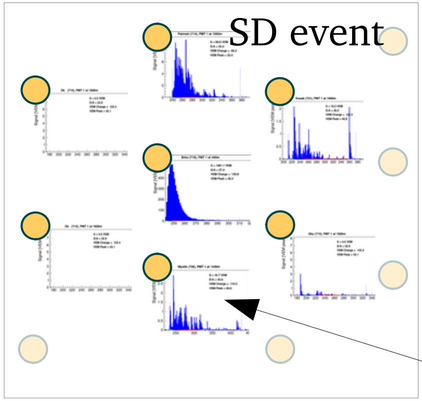
Pierre Auger Observatory

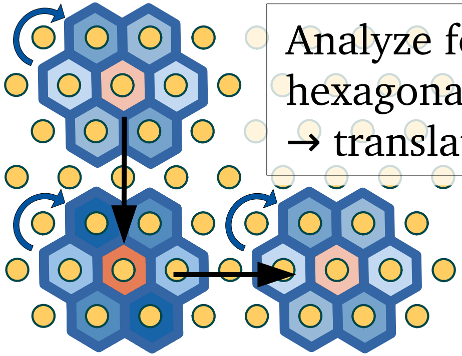
Fluorescence Detector (15% duty cycle)

- direct and precise observation of shower maximum *X*max

Surface Detector (~100% duty cycle)

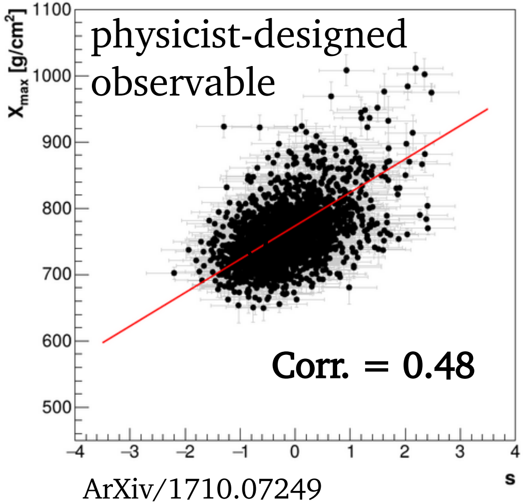
- reconstruction of shower maximum using deep learning
- verification using hybrid measurements



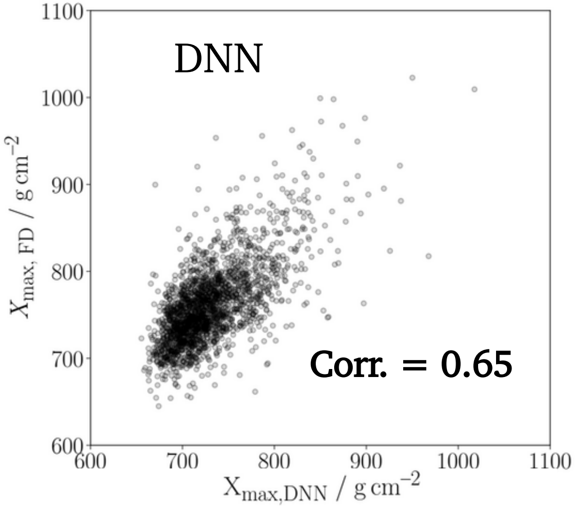


Analyze footprint with hexagonal convolution → translation + rotation

analyze traces with RNNs

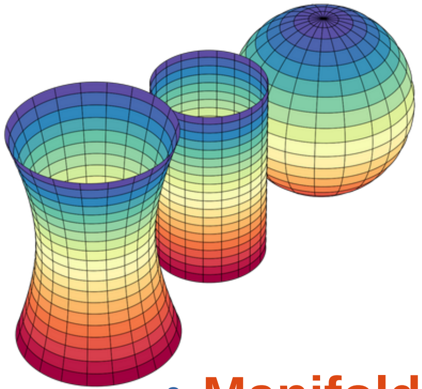


ArXiv/1710.07249



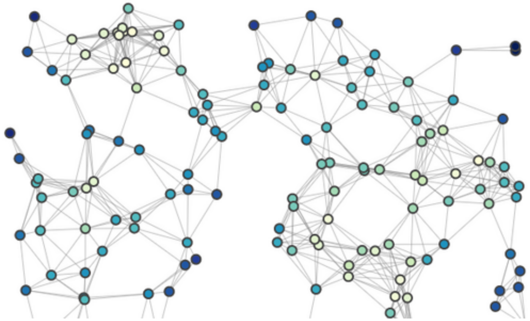
Generalization to Non-Euclidean Domains

- Defining convolutions, challenging on non-euclidean domains
  - Deformation of filters, changing neighbor relations
  - Non-isometric connections on graphs



- **Manifolds**

source: wikipedia



- **Graphs**

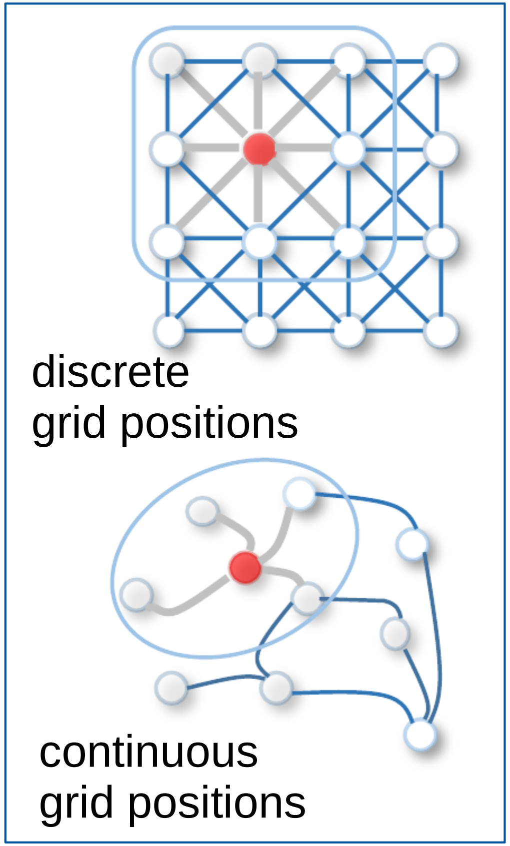
source: Cody Marie Wild, Towards Data Science

Image-like data
- collection of pixels (vector)
- coherent (rarely sparse)
- discrete, regular (symmetric)
- feature euclidean space

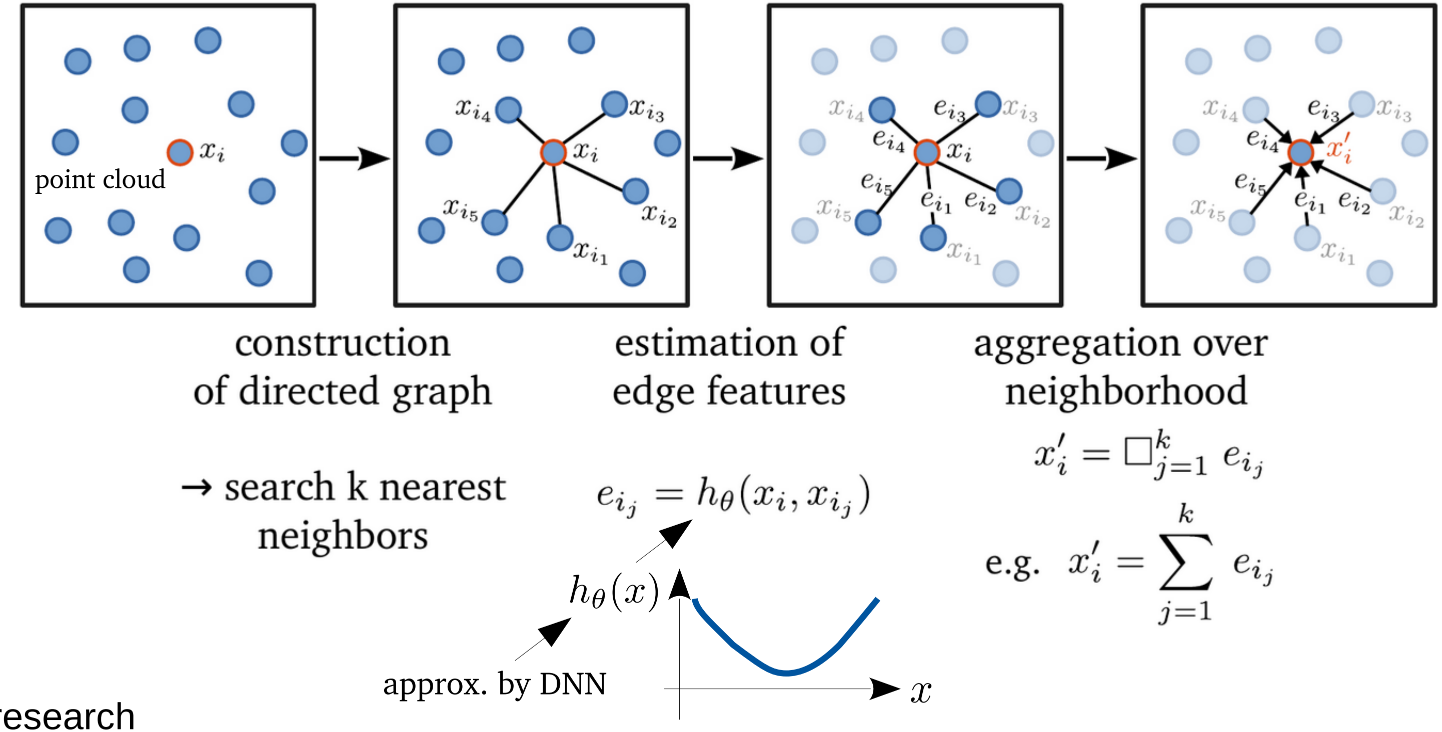
How can we generalize convolutions?

Graph Network: Dynamic Edge Convolution

Y.Wang et al,
https://arxiv.org/abs/1801.07829



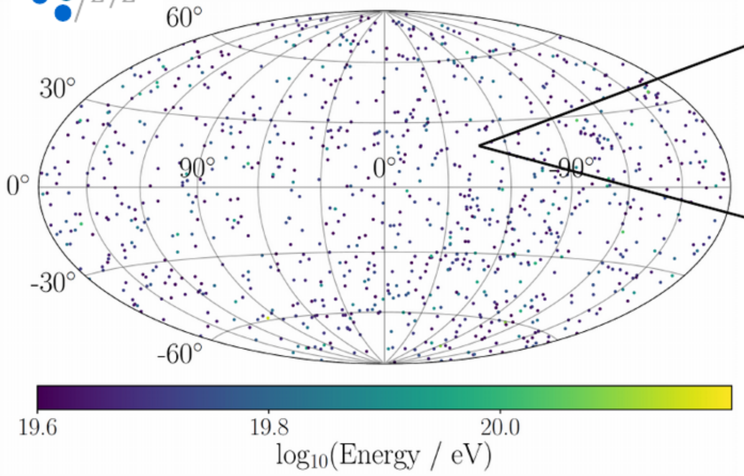
- define continuous filter (using kNN)
- flexible for many settings: irregular structures, point clouds
- dynamically ‘adapt’ fundamental graph structure each layer



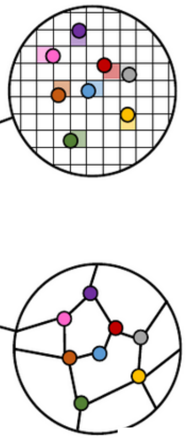
Search for UHECR Origins

Slide credit: Niklas Langner

Deflection depends on energy charge and GMF



Continuously distributed on **sphere**



sparse, spherical not suited for CNN

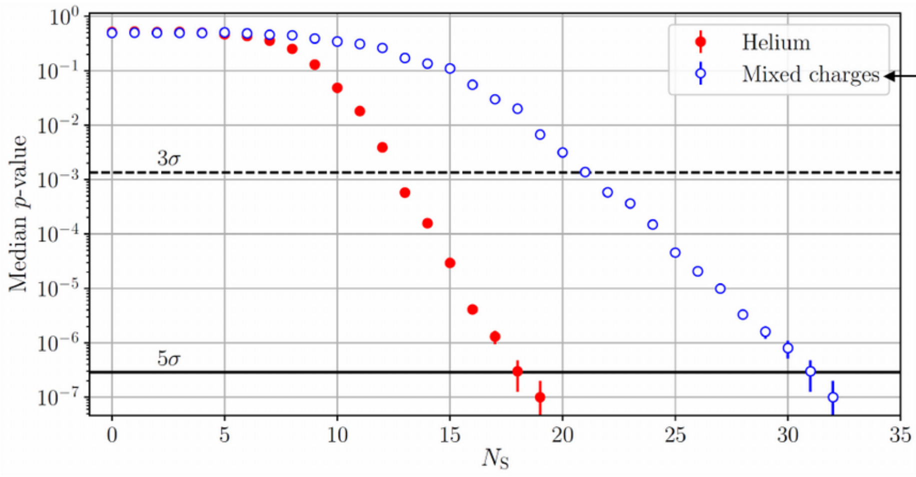
use Dynamic Graph Network

Situation:

One measured sky (spherical)

Learn to classify between

- isotropic sky / signal
- use dynamic edge convolutions



15% H
45% He
40% C, N, O

Bister et al., 10.1016/j.astropartphys.2020.102527

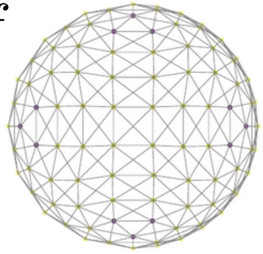
Convolutions on Spherical Domains

- (Graph) convolution in spectral domain
  smooth, localized filter → Chebychev expansion

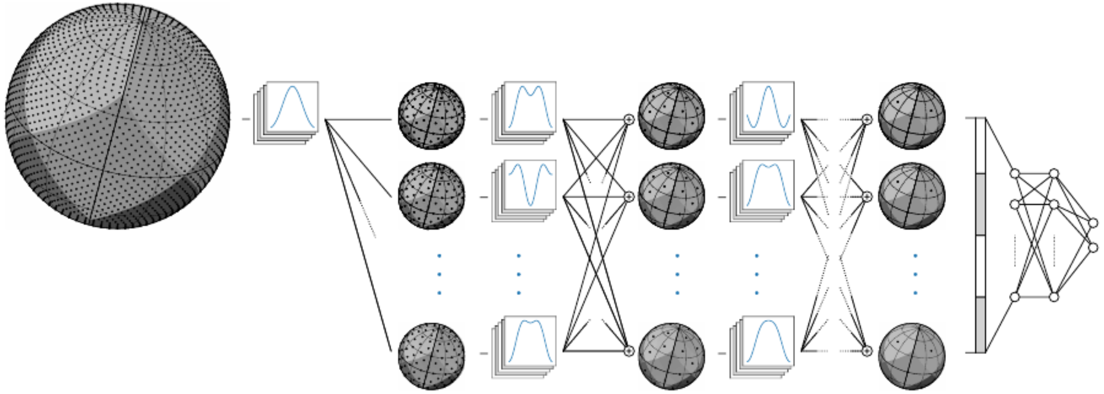
  Example: DeepSphere, for spherical data
  - HEALPix pixelization defines graph structure
- based on fixed pixels (useful for sensor configurations)

$$f * w = \Phi \hat{W} \Phi^T f$$

filter adaptive in spectral (Fourier) domain



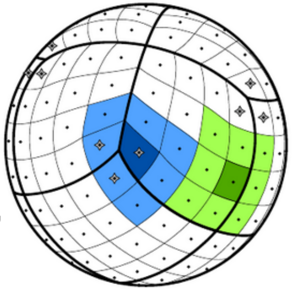
constructed graph



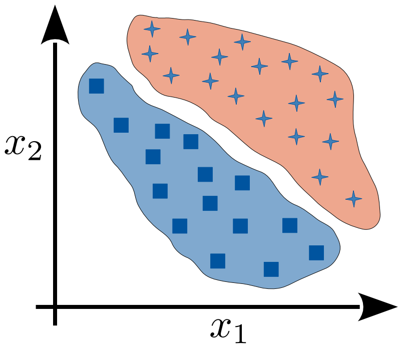
N. Perraudin el al., 10.1016/j.ascom.2019.03.004

N. Krachmalnicoff et al., A&A 628, A129 (2019)

Hybrid approach:
'Indexed Conv'
Define 'HEALPix filters'



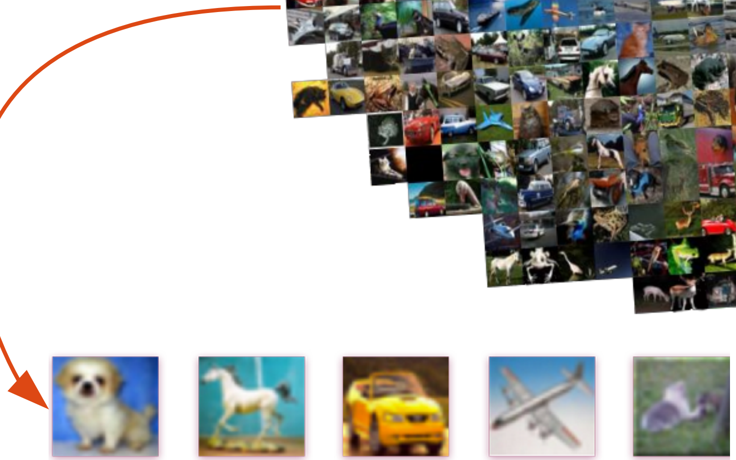
Application to search for UHECR sources:
O. Kalashev et al., 10.1088/1475-7516/2020/11/005



Unsupervised Learning

- Density estimation
- Anomaly detection
- Generative Models
- Simulation Refinement

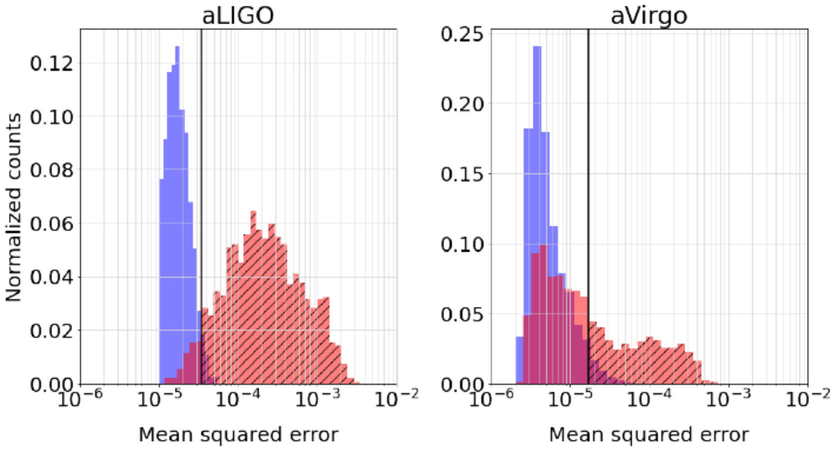
CIFAR10

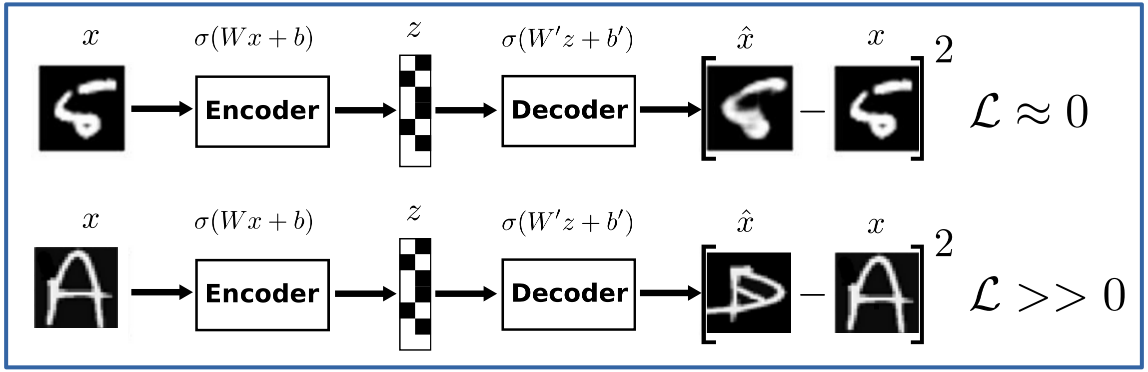


Learn to generate new samples

Anomaly Detection

- Search for data, different than used for training, using autoencoders
- indication for new physics, proposed for BSM searches at LHC
- training without limited data (no signal labels)
  - first approaches in astroparticle physics
    - detection of gravitational waves





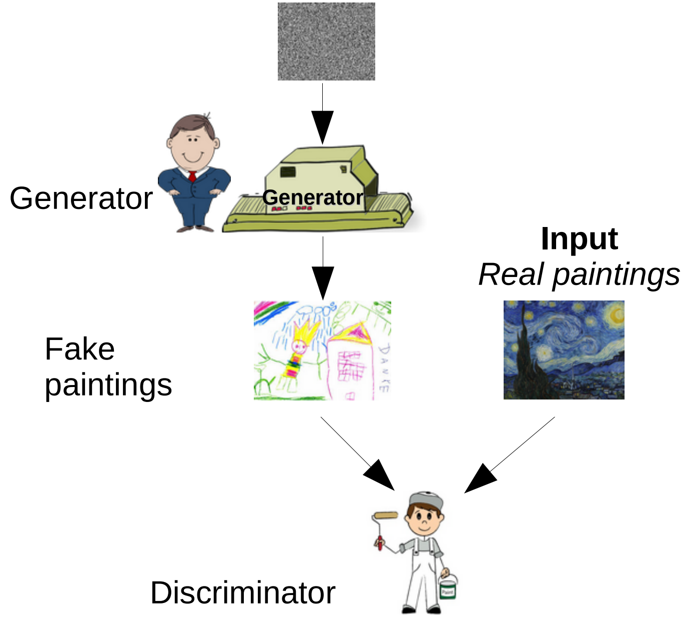
F. Morawski et al., Mach. Learn.: Sci. Technol. 2 045014

Generative Models

- Which picture is generated which is part of CELEB A ?

T. Karras et al. - https://arxiv.org/abs/1812.04948

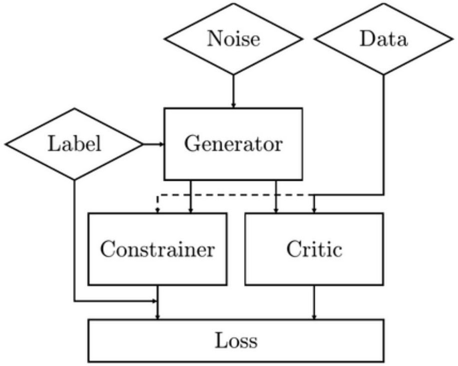
Play the game:
https://www.whichfaceisreal.com



Generative Adversarial Networks
discriminator trained to classify fake/real
generator trained to fool discriminator
→ learns to generate realistic fake samples

Application in Particle Physics

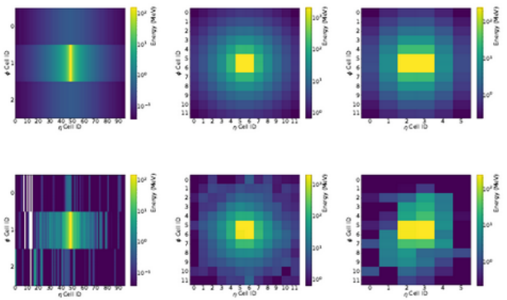
- Detector simulation are very time consuming
  - accelerate using GANs (speed-up of $10^3 - 10^5$)
- Add constraining networks to condition the generation
  - e.g., (energy, particle type, arrival direction)
- Samples must comply with physics laws
- Samples have to follow phase space density → usually no cherry-picking



Layer

Geant4

GAN

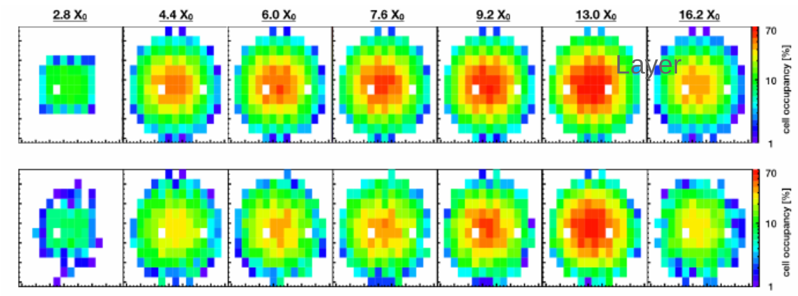


Paganini, Oliviera, Nachman - Phys. Rev. D 97, 014021 (2018)

Layer

Geant4

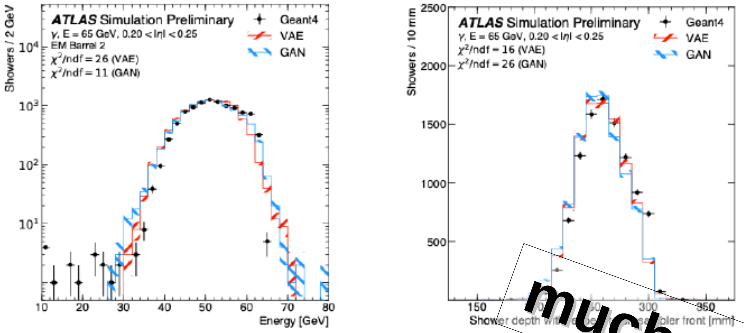
WGAN



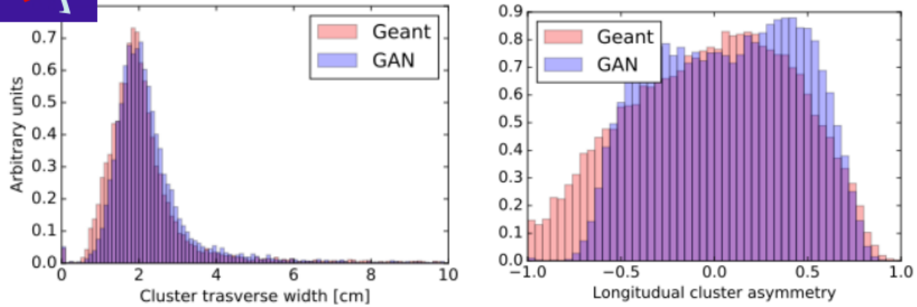
Erdmann, Glombitza, Quast - T. Comput Softw Big Sci (2019) 3: 4

Application of Generative Models at LHC

Deep generative models for fast shower simulation in ATLAS, http://cds/2680531

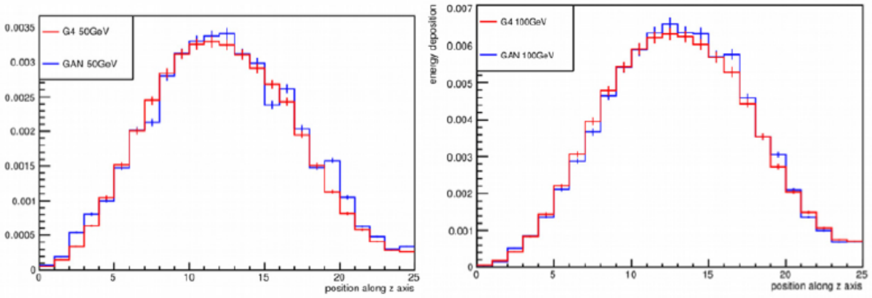


Generative Models for Fast Calorimeter Simulation - LHCb case, 1812.01319

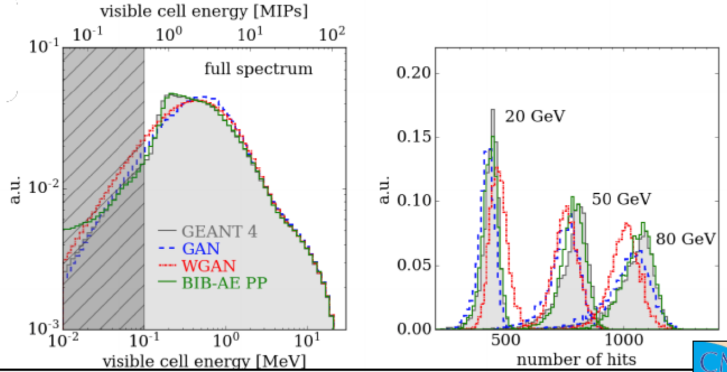


much progress!

3D convolutional GAN for fast simulation, https://doi.org/10.1051/epjconf/201921402010



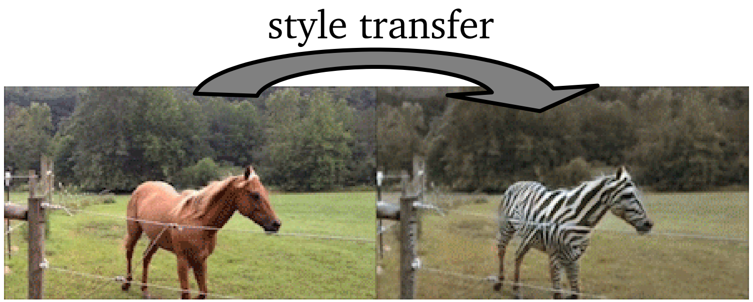
Getting High: High Fidelity Simulation of High Granularity Calorimeters with High Speed, 2005.05334



Generalization Capacities on Data

DNNs and Domain Adaption

- models are trained using physics simulations
- trained models are applied to data
  - ➔ can lead to reconstruction biases



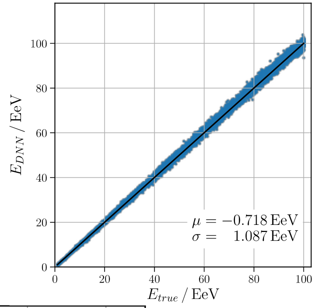
https://bair.berkeley.edu/static/blog/humans-cyclegan/

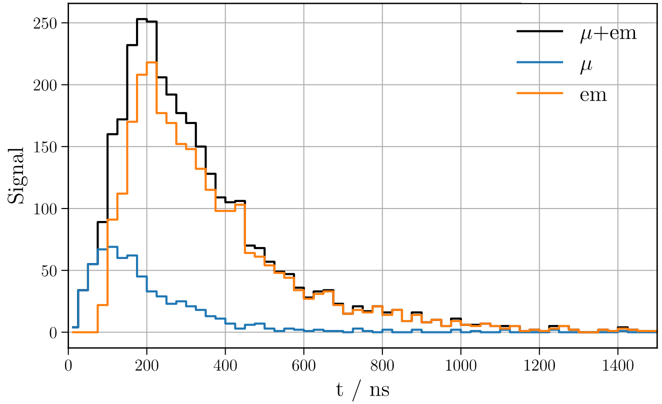
Simulation Refinement

Erdmann et al.
Comput Softw Big Sci (2018) 2: 4

- Training on **simulations** but application on **data**
  - Model can be sensitive to artifacts / mismatches existing in simulation

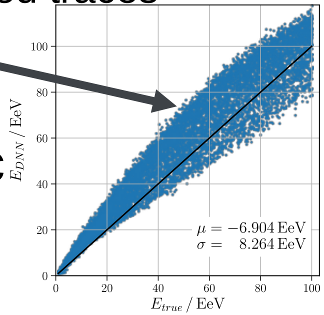
Simulation
70% electromagnetic
30% muonic

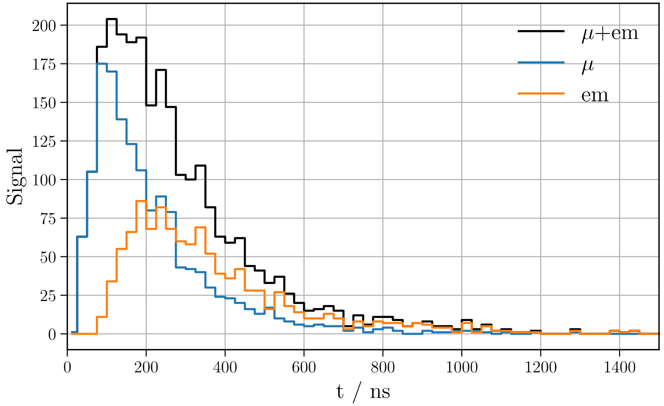




Neural network can not handle modified traces

Data
30% electromagnetic
70% muonic
+ Increased noise

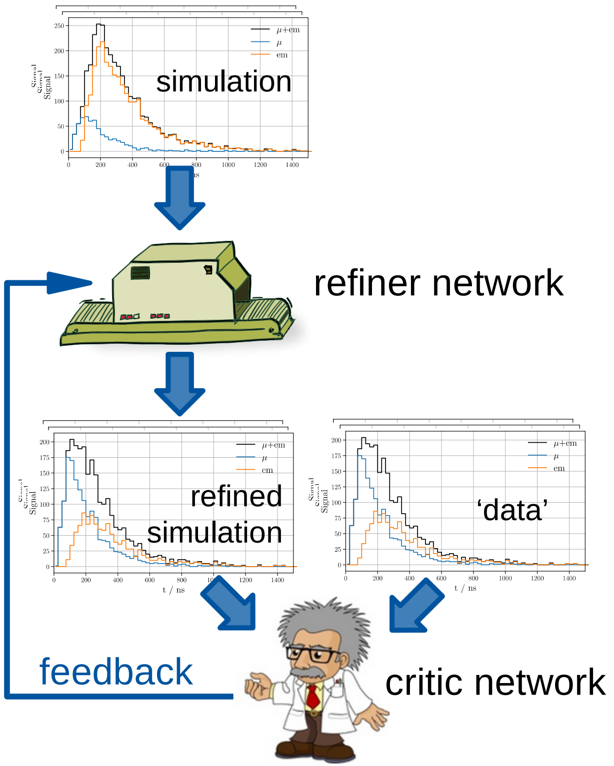




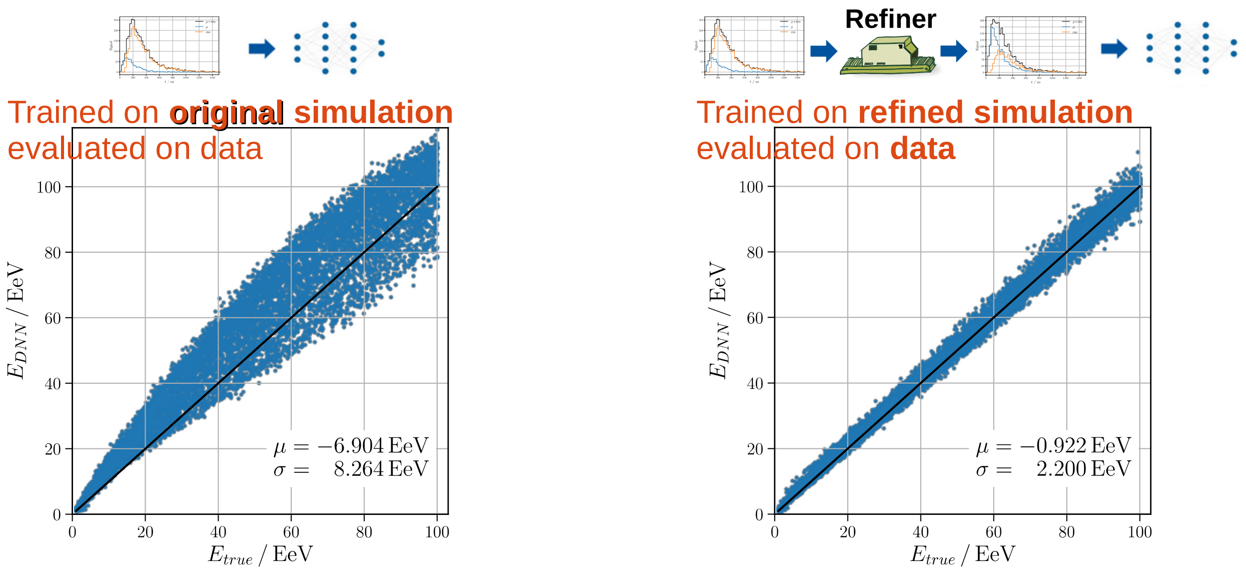
Simulation Refinement

Erdmann et al.
Comput Softw Big Sci (2018) 2: 4

mitigate data / simulation mismatches → train *refiner* to refine simulated data



- feedback given by adversarial *critic* network, rating the refined simulation quality
  - refiner uses feedback to improve performance
- improved performance when training with refined simulation

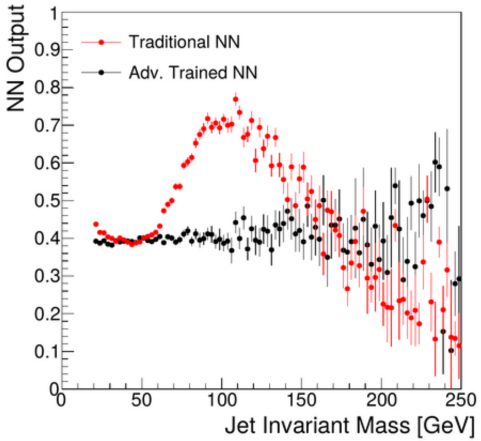


Decorrelated Jet Tagging

C. Shimmin at al., Phys. Rev. D 96, 074034 (2017)

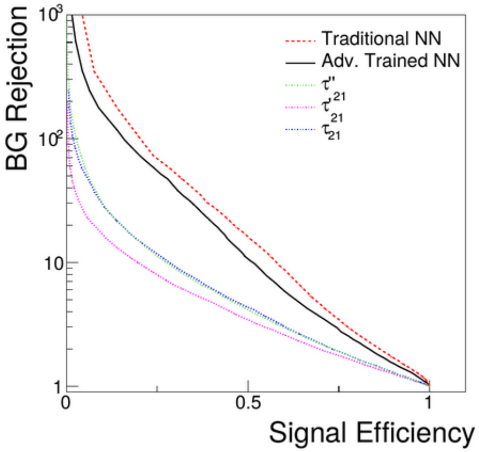
- Reduce systematic uncertainties
  - Decorrelate jet mass of background processes
  - → Reduce model dependence
- Train adversary to constrain classifier

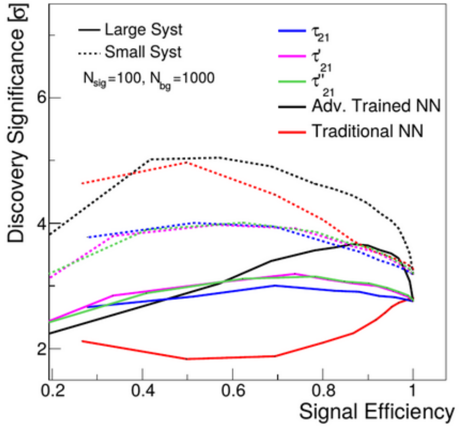
Classifier *Jet or QCD* → **Adversary** *Infer jet mass* → *10 mass bins*



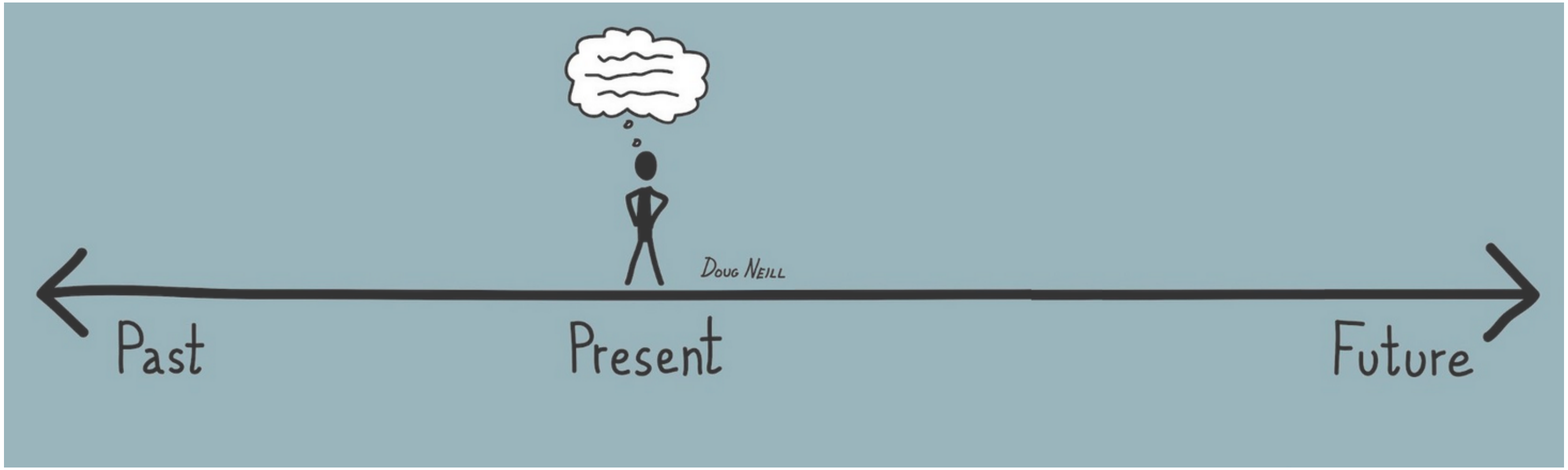
$$\mathcal{L} = \mathcal{L}_{Classification} - \mathcal{L}_{Adversarial}$$

- Classification performance
  - *Best:* traditional network
- Discovery significance
  - *Best:* adversarial network





(Past), State, and Future*



*Attempt of reviewing the current status and open a fruitful discussion image credit: Doug Neill

*Note: There is no free lunch → ”Older algorithms” are not necessarily worse

Past, Present, and Future – Deep Learning in Astroparticle Physics

I. Classic ML
Published physics analyses using high-level observables, BDTs, RFs

II. Proof of concept
- First SAL publications of applying DL at low- & high level data
- Use of standard architectures: FCNs, RNNs, CNNs mostly on simulations and toy simulations

Exploiting symmetries
Incorporating symmetries into DNNs, GCNs, transformer

III. Verified reconstruction mechanisms
First publications by Collaborations, e.g., Pierre Auger, IceCube, KM3Net, …

IV. Physics analyses with DL
- Publications by Collaborations
- Application to full data sets
- Extensive study on systematic unc.

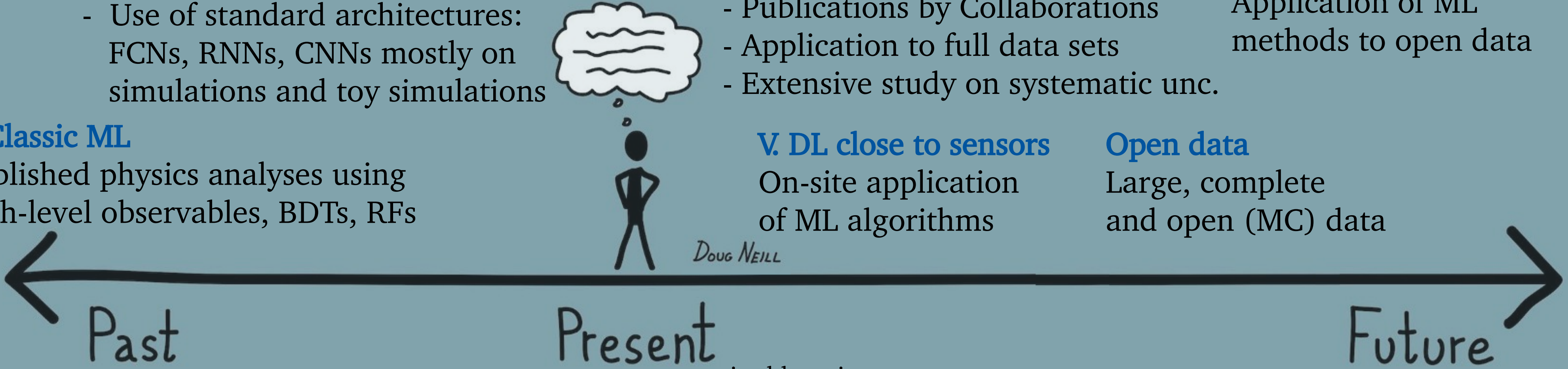
V. DL close to sensors
On-site application of ML algorithms

‘Unsupervised era’
- exploiting measured data
- refinement of simulations
- domain-robust DNNs

Interpretability
- DNN introspection & causality studies
- Distilling physics laws from DNNs

Multi-experiment DL
Application of ML methods to open data

Open data
Large, complete and open (MC) data



supervised learning

unsupervised learning

self-supervised learning

Future Challenges

Causality and visualization

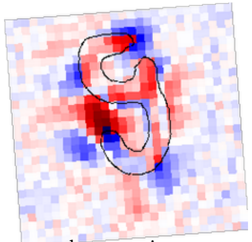
- Deep networks hold ~millions of parameters
  - Improve understanding of the emerging "tool"
- What makes a "9" a "9" for a deep neural network?
  - Find patterns important for the reconstruction
  - Explore model ← →data, what is most useful?
  - Open the black box → find (new) physics?!

Uncertainty of predictions

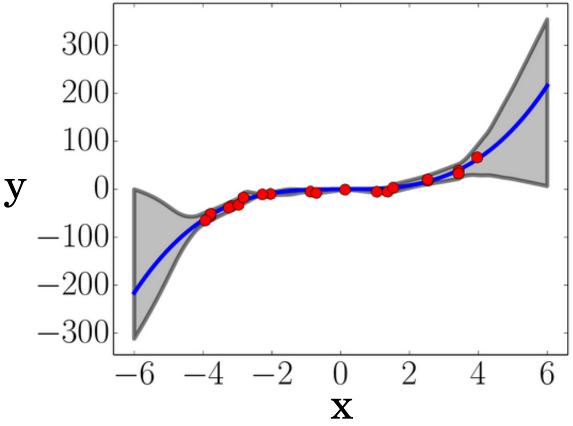
- Physics analyses require careful uncertainty estimation
- De-correlation to reduce systematic biases
- ➢ Has to be solved by physicists

https://arxiv.org/abs/1602.04938

(a) Husky classified as wolf (b) Explanation



www.heatmapping.org



Lakshminarayanan et al. - ArXiv:1612.01474

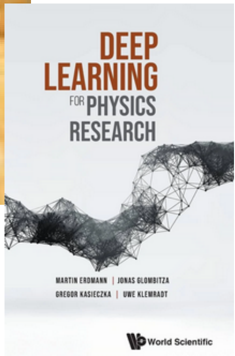
Tryout
Deep Learning
Yourself!

Find many physics examples at:

http://www.deeplearningphysics.org/

For example:

- CNNs, RNNs, GCNs
- GANs and WGANs
- Anomaly detection, Denosing AEs
- Visualization & introspection and more



References and further Readings

- [1] I. Goodfellow, Y. Bengio, A. Courville, Deep Learning, Chapter 10, MIT Press, 2016, www. deeplearningbook.org
- [2] M. Erdmann, J. Glombitza, G. Kasieczka, U. Klemradt, Deep Learning for Physics Research, World Scientific, 2021

Deep learning for astroparticle physics

- BACKUP -

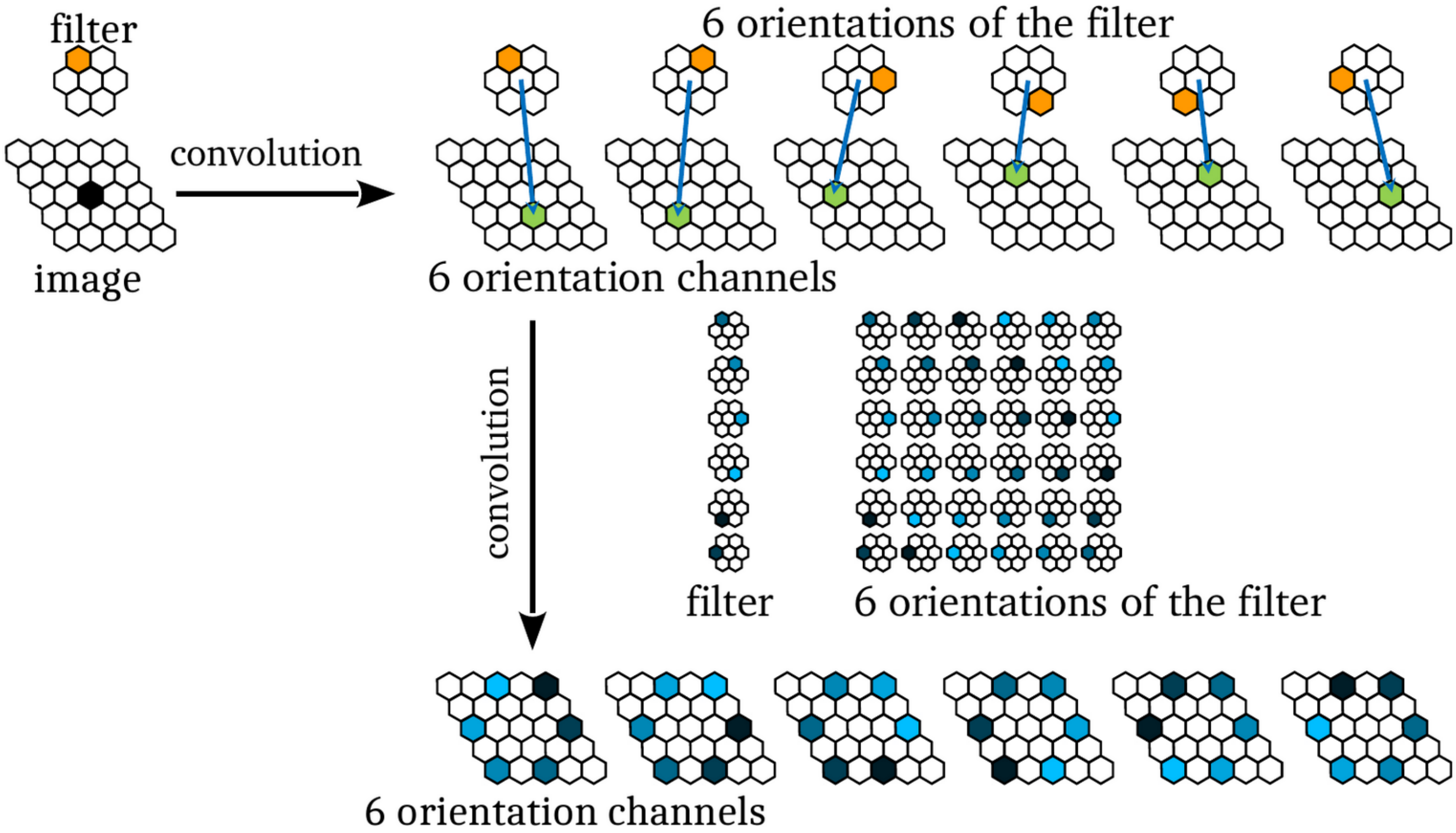
Jonas Glombitza, *Martin Erdmann*

RWTH Aachen University

SPONSORED BY THE

Federal Ministry
of Education
and Research

HexaConv



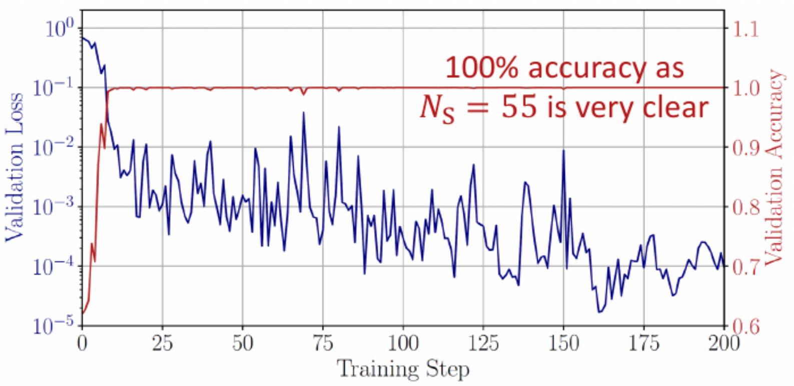
filter
image
convolution
6 orientations of the filter
6 orientation channels
convolution
filter
6 orientations of the filter
6 orientation channels

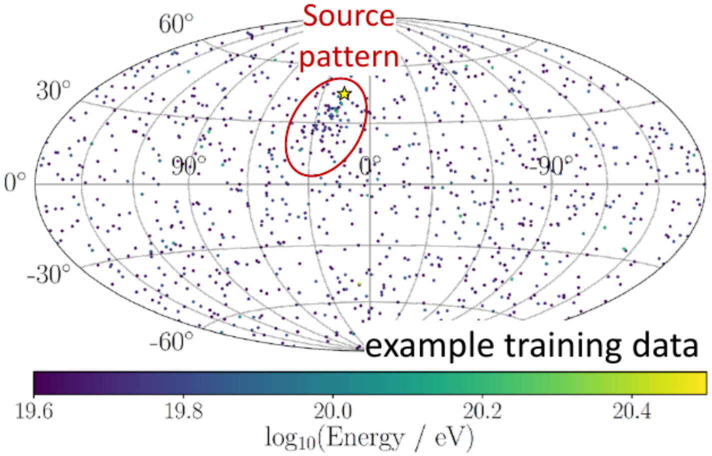
Training

- **1000 cosmic rays** with $E > 40$ **EeV**, spectrum similar to measurements of Pierre Auger Observatory
- Simulate on the fly during training → **no overfitting**
- Train on **strong multiplets** and let the network generalize

Composition	Turb. deflection T	Coherent deflection D	Source CRs
Pure Helium	50% of JF12 maximum in train values from JF12 in validation	Typical values from JF12 but larger than turbulent	55

EdgeConv dims	Loss	Optimizer	Concatenation
16/32/64	Categorical cross entropy	Adam	No





Slide credit: Niklas Langner

Smoothing in Spectral domain

- Approximate $\hat{W}_\theta$ in spectral domain $\tau(L)f = \Phi\tau(\Lambda)\Phi^T f$

$$\Phi(\hat{W}_\theta \Phi^T f) = \Phi \begin{pmatrix} \tau_\theta(\lambda_1) & & \\ & \ddots & \\ & & \tau_\theta(\lambda_n) \end{pmatrix} \Phi^T f$$

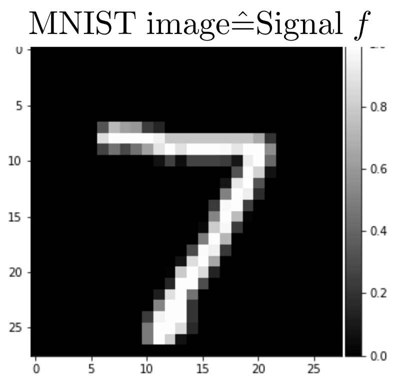
- $\hat{W}_\theta \approx \tau_\theta(\lambda) = \sum_{k=1}^{K} \theta_k f_k(\lambda)$

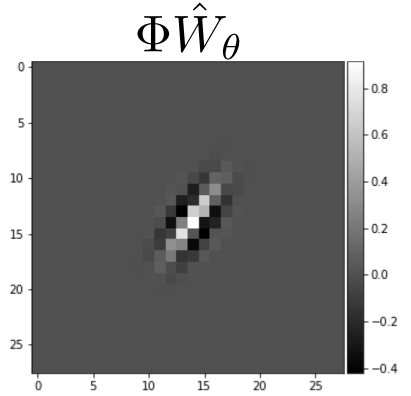
some function

adaptive parameters

- Learn only K parameters → parameter reduction ✓
- For $K << N$, $\hat{W}_\theta$ gets smooth in spectral domain ✓
  - Spectral theory: filter become local!

proposed by Bruna et al. https://arxiv.org/abs/1312.6203



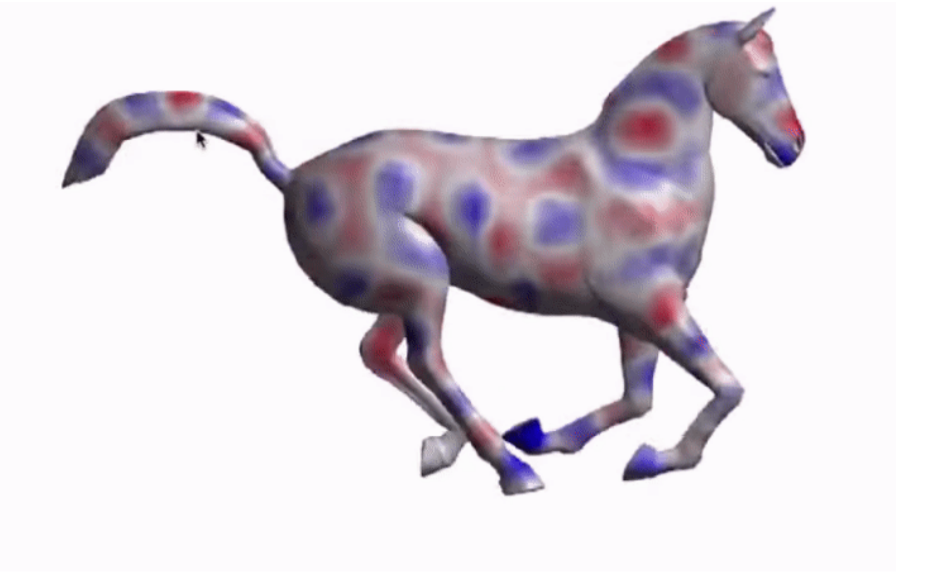


Boris Knyazev, Towards data science

Spectral Convolutions

- We can perform the convolution in the spectral domain
  - Signal $X^{(l)}$
  - Weight matrix $W^{(l)}$
- $X^{(l+1)} = \Phi(\Phi^T X^{(l)} \cdot \Phi^T W^{(l)})$
  $= \Phi \hat{W}_\theta^{(l)} \Phi^T X^{(l)}$
- $\hat{W}_\theta^{(l)} = \mathrm{diag}(\theta_1, ..., \theta_n)$

Adaptive parameters in Fourier domain



NIPS2017: M. Bronstein, J. Bruna, A. Szlam, X. Bresson, Y. LeCun

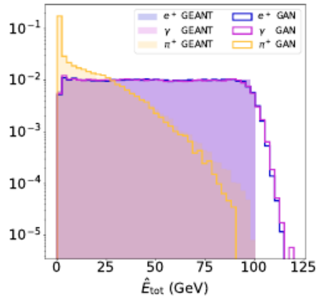
Problems:

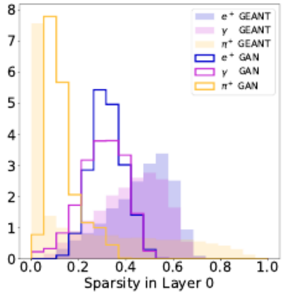
- Weights scale with number of graph nodes
  - Act global! No prior on local features!
- $\hat{W}_\theta^{(l)}$ strongly depends on L (Λ, Φ)
  - Bad generalization performance!

Generation of Calorimeter Images

- Quality of images is crosschecked using physics observables
- Challenges: Sparsity, logarithmic intensity distribution

Traditional GAN



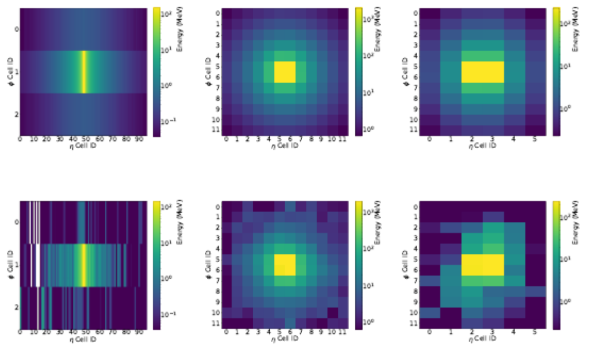


Paganini, Oliviera, Nachman - https://arxiv.org/abs/1712.10321

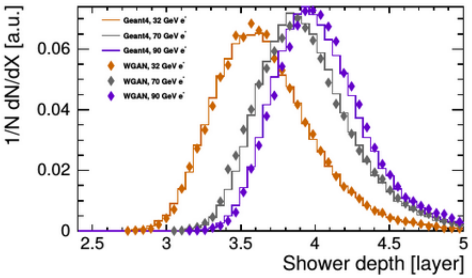
Layer

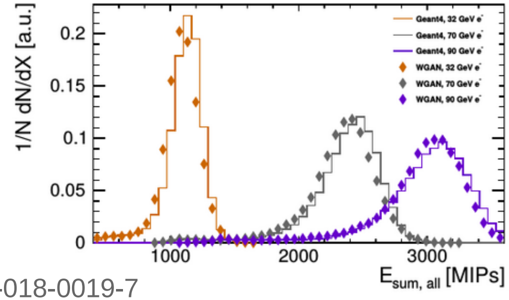
Geant4

GAN

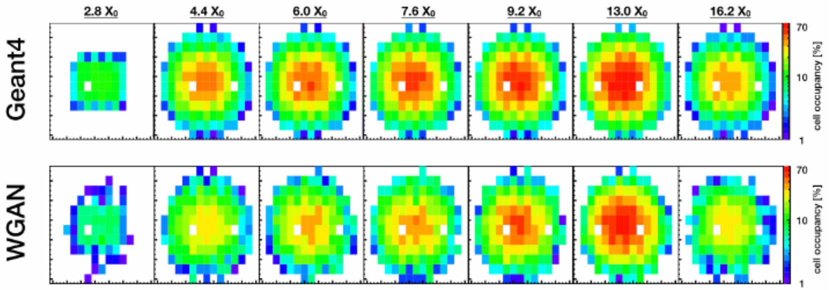


Wasserstein GAN





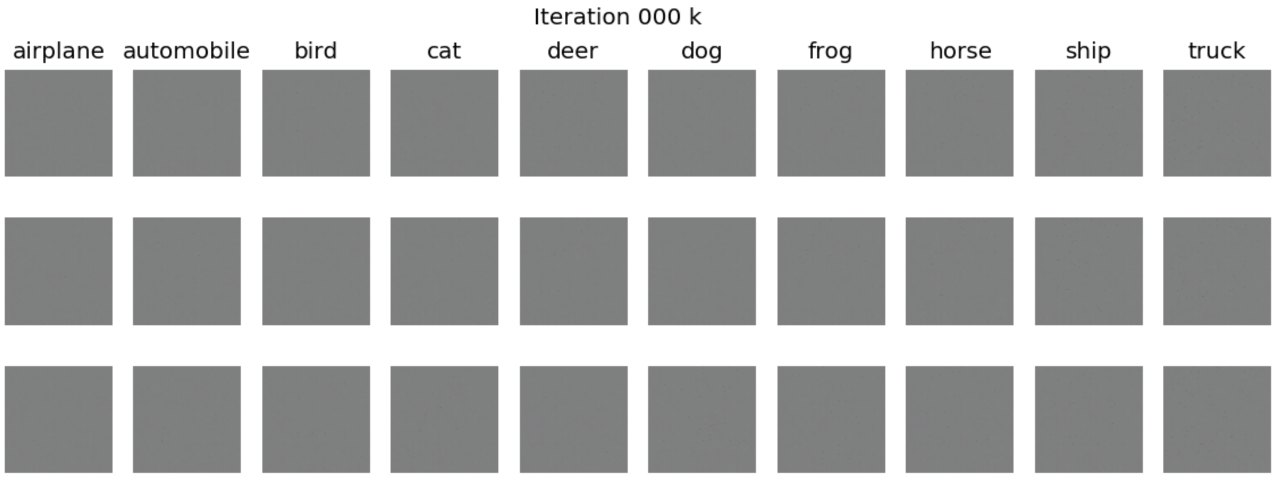
Layer



Erdmann, Glombitza, Quast - 10.1007/s41781-018-0019-7

Results

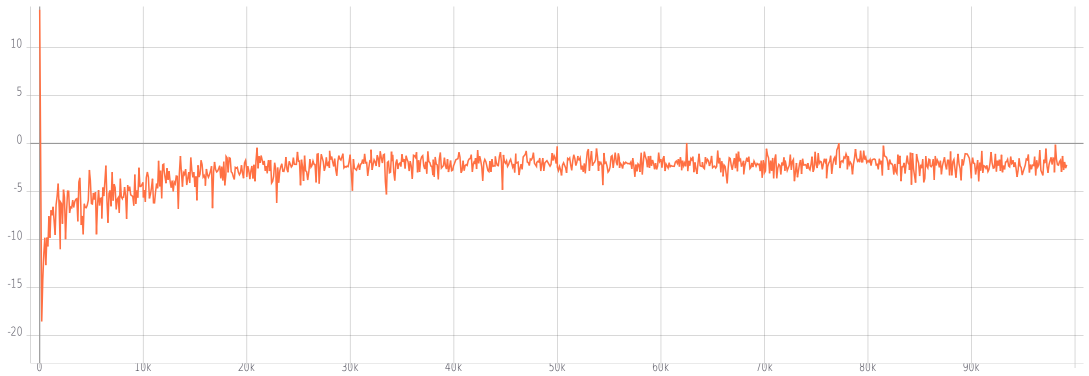
- WGAN generates images with much better quality
- Critic loss converges
- Loss correlates with images quality



Wasserstein GANs

- Allow stable training of GANs
  - Train critic to convergence
  - Precise feedback for generator
- Prevent mode collapsing
- Provide meaningful loss

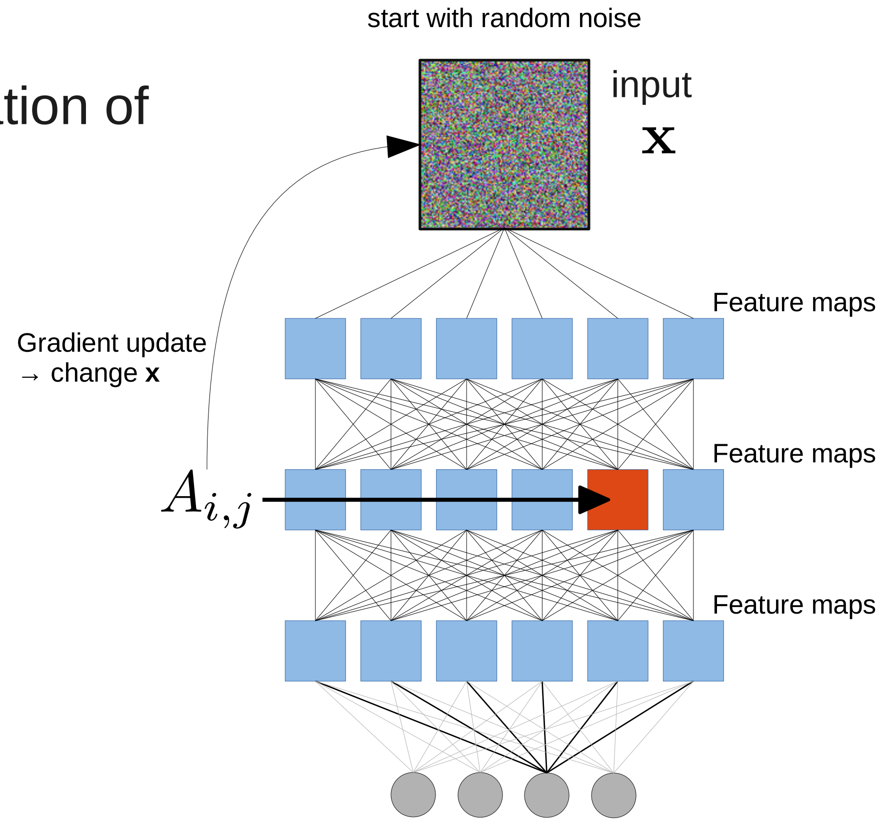
Critic loss



Activation Maximization - CNN

Idea:

- Construct pattern which maximizes the activation of a specific feature map
- Model f_θ pre-trained, weights θ fixed
- Find $\tilde{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmax}} \; h(\mathbf{x}, \theta)$
- $h(\mathbf{x}, \theta) = \sum_{i,j} A_{i,j}(\mathbf{x}, \theta) + b$
- Gradient **ascent** $\mathbf{x}' \to \mathbf{x} + \alpha \dfrac{dh(\mathbf{x}, \theta)}{d\mathbf{x}}$



https://doi.org/10.1142/12294