Machine learning in Astroparticle Physics.

A.Kryukov*
SINP MSU, Moscow; IGU, Irkutsk

*kryukov@theory.sinp.msu.ru

Outline

- ML
  - A few statistics
  - What is ML and what is difference between ML
    and traditional programming
  - ML vs. DL. Mathematical basis of DL
- ANN
  - Classification of ML (with/wothput adviser)
  - ANN (CNN, GAN, …)
- ML in Particle Astrophysics
  - Gamma Astronomy
  - Nuetrino
- ML in other fields of physics
- Artificial intelligence vs Human brain
- Conclusion

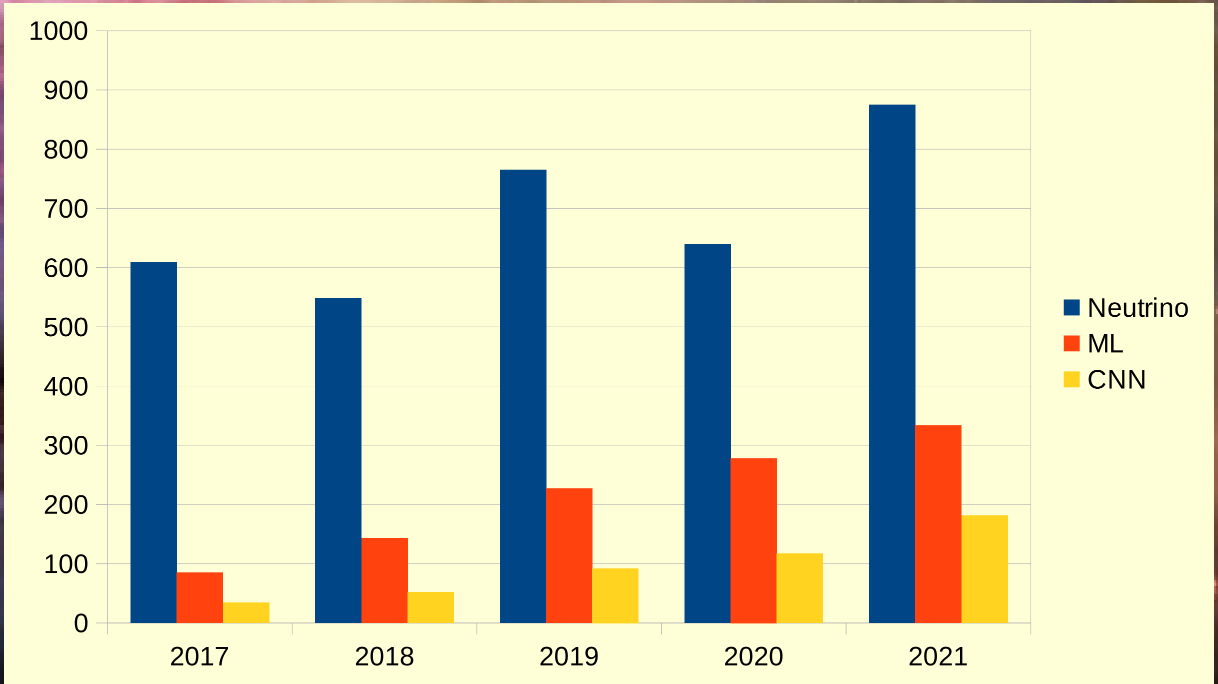
Machine Learning

- A few statistics (2022)

	Google	ArXiv, astro-ph	ArXiv, hep-ph
Machine Learning	1 690 000 000	423	138
Deep Learning	1 500 000 000	206	54
Convolutional Neural Network	45 100 000	186	29
Neutrino	12 300 000	1 079	1 480
ALL		20 508	10 141

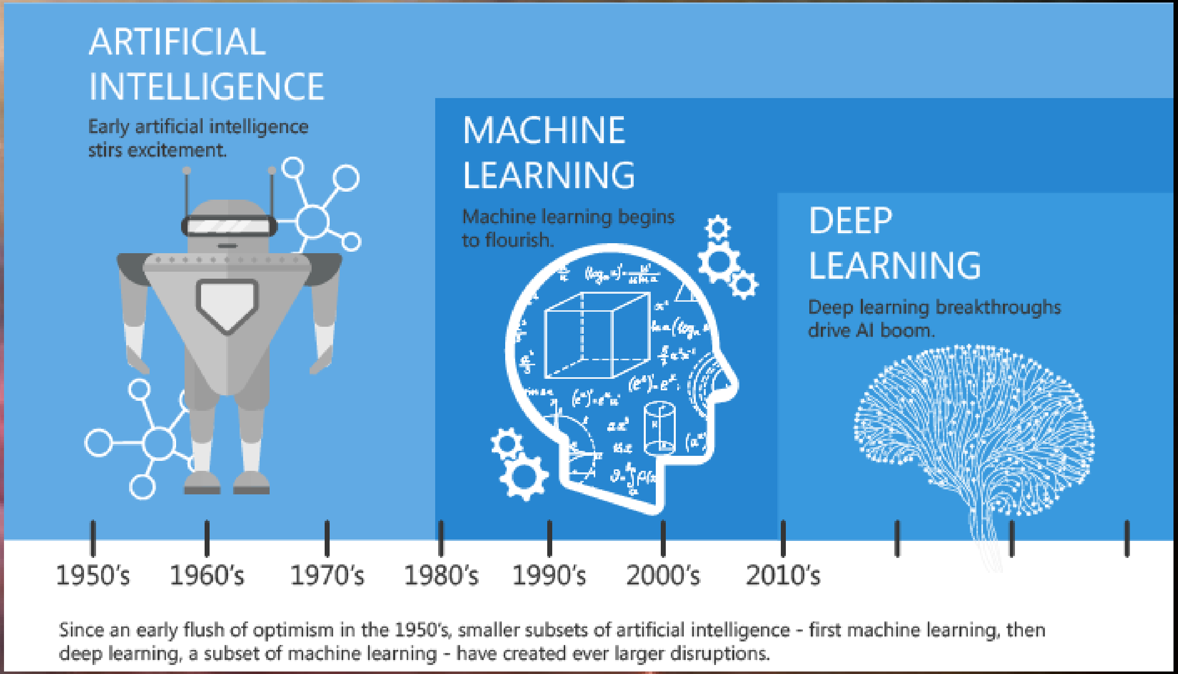
Machine Learning

- Number of publication in ArXiv, astro-ph



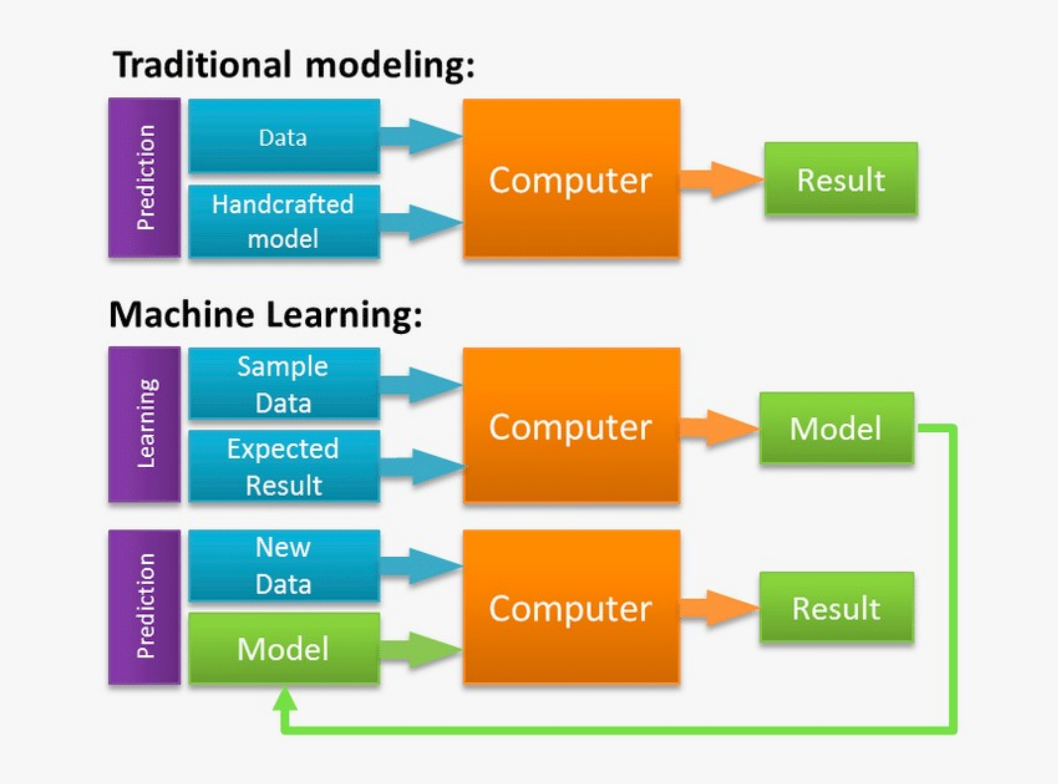
Machine Learning

- Machine learning is a branch of artificial intelligence (AI) and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.
- Deep learning is a subset of machine learning, which is essentially a neural network with three or more layers.



What is difference between ML and traditional programming

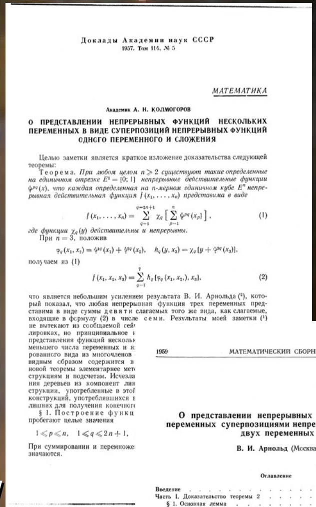
- The difference between normal programming and machine learning is that programming aims to answer a problem using a predefined set of rules or logic. In contrast, machine learning seeks to construct a model or logic for the problem by analyzing its input data and answers.

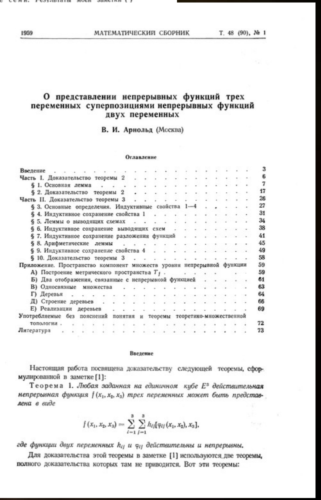


Mathematical basis of DL

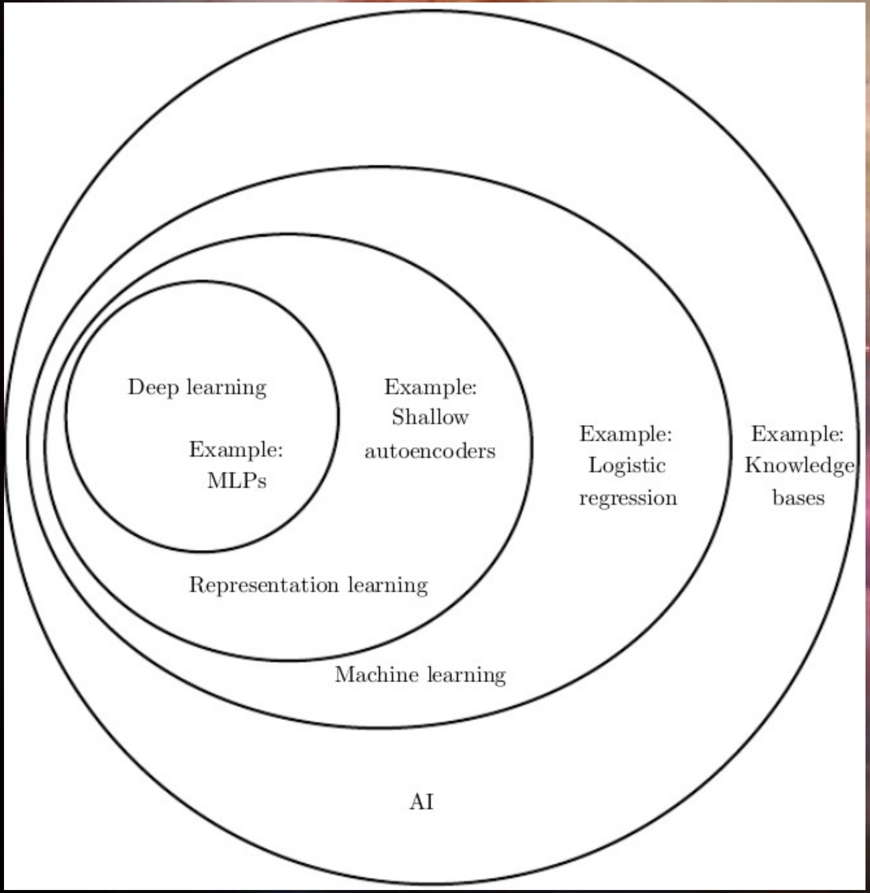
- Hilbert's 13th problem (1900): is it possible to represent a function of several variables as a superposition of functions of fewer variables.
  - A. N. Kolmogorov, “On the representation of continuous functions of several variables as superpositions of continuous functions of one variable and addition” // Dokl. - 1957. - T. 114, issue. 5. - S. 953-956 (http://www.mathnet.ru/rus/dan22050)
  - V. I. Arnold, “On the representation of continuous functions of three variables by superpositions of continuous functions of two variables” // Mat. Sb., 48(90):1 (1959), 3–74 (http://mi.mathnet.ru/msb4884)
- ANN is an universal aprroximator

$$F(x_1, x_2, \ldots, x_n) = \sum_{j=1}^{2n+1} g_j\Big(\sum_{i=1}^{n} h_{ij}(x_i)\Big)$$

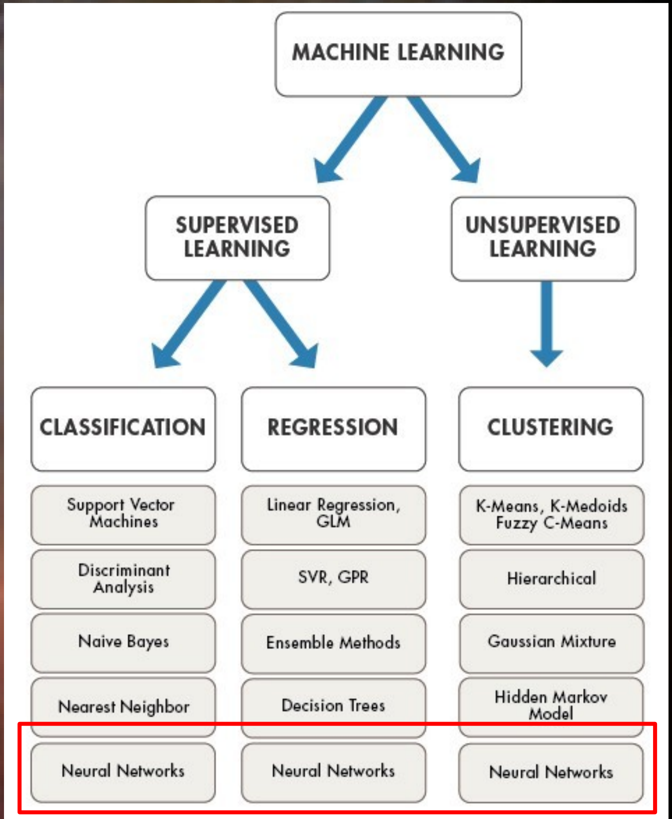




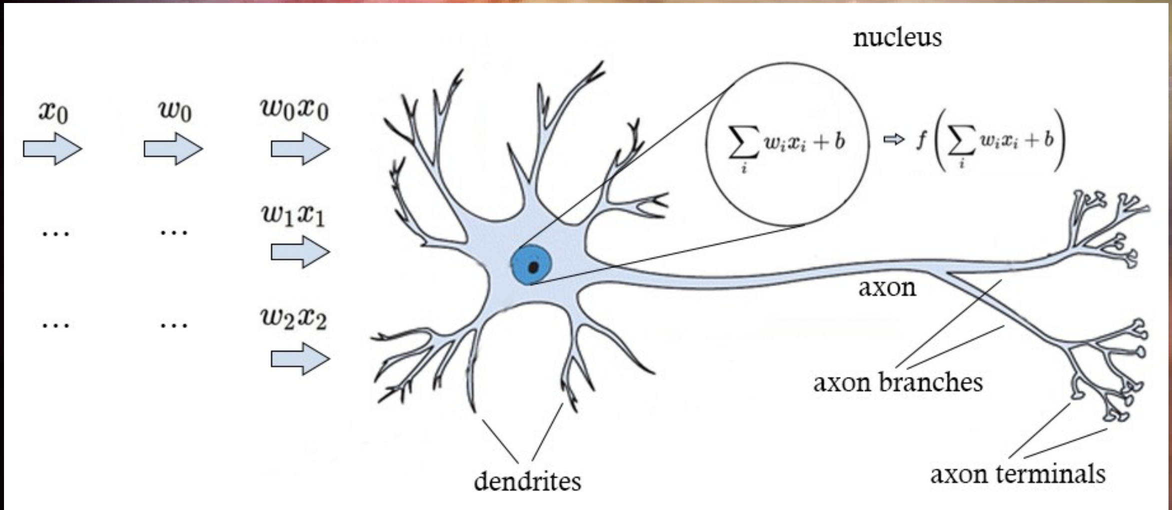
Classification of ML methods



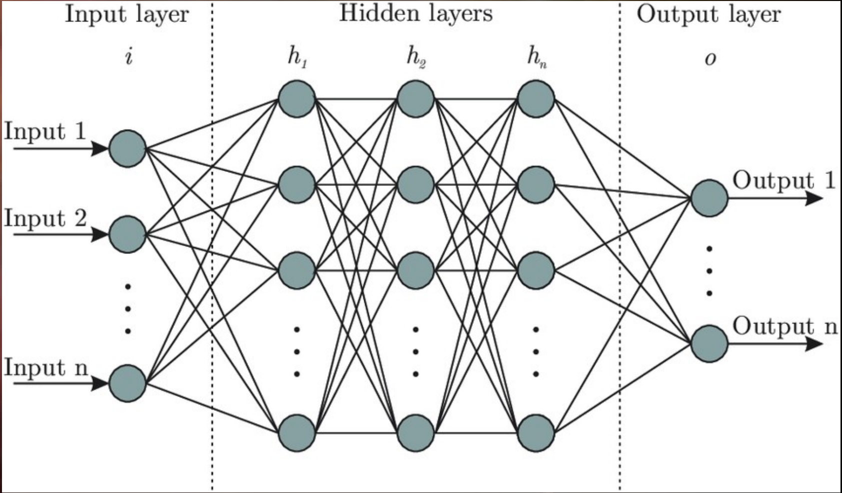
- ML algorithms
  - Classification
  - Regression
  - Clustering



Artificial Neural Network

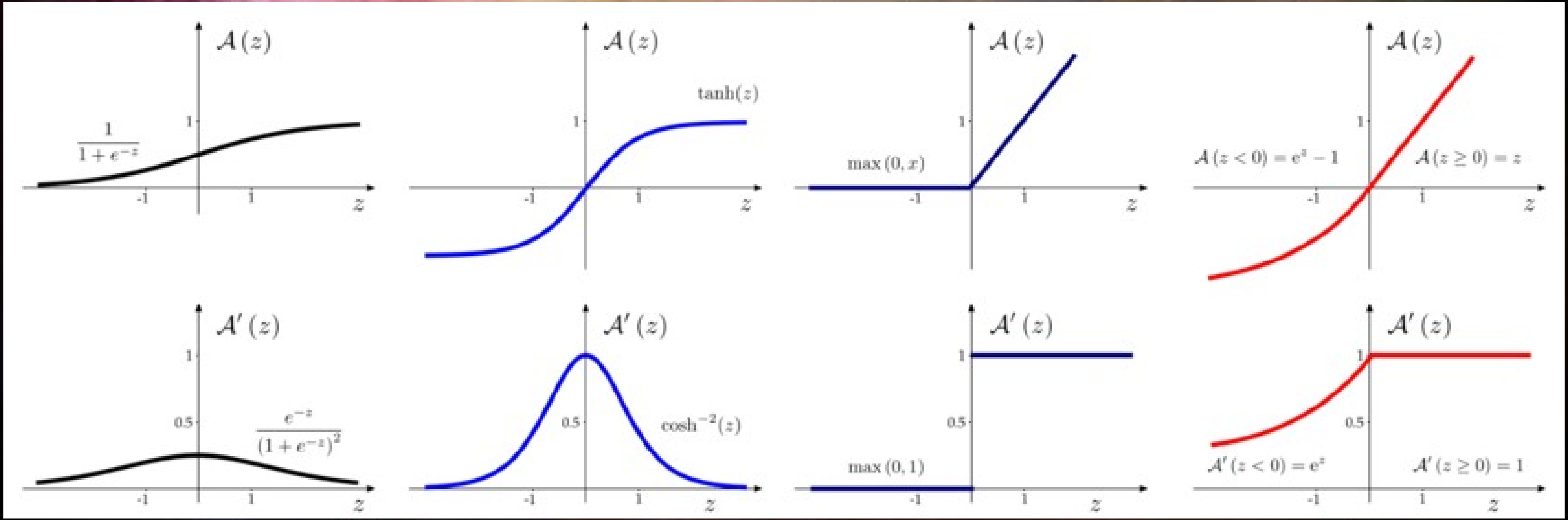


- The structure of the artificial neuron was inspired by the natural neuron in the brain.



Most common activation functions

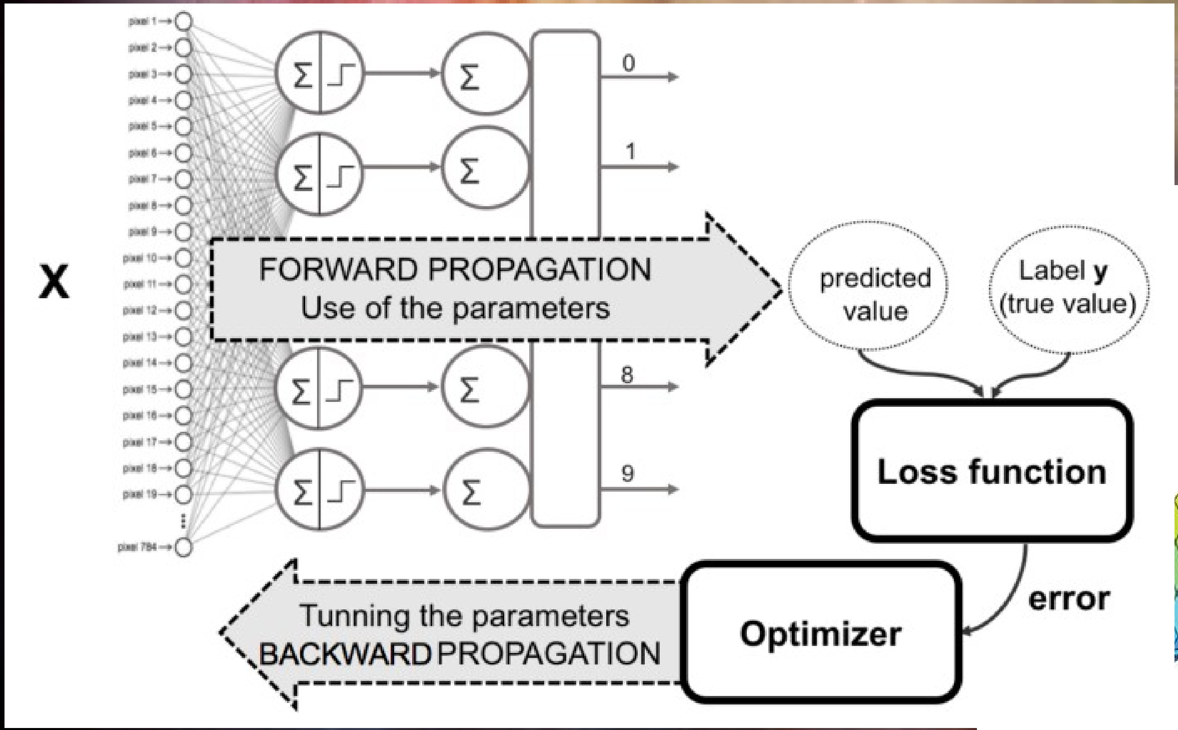
- Most popular AF

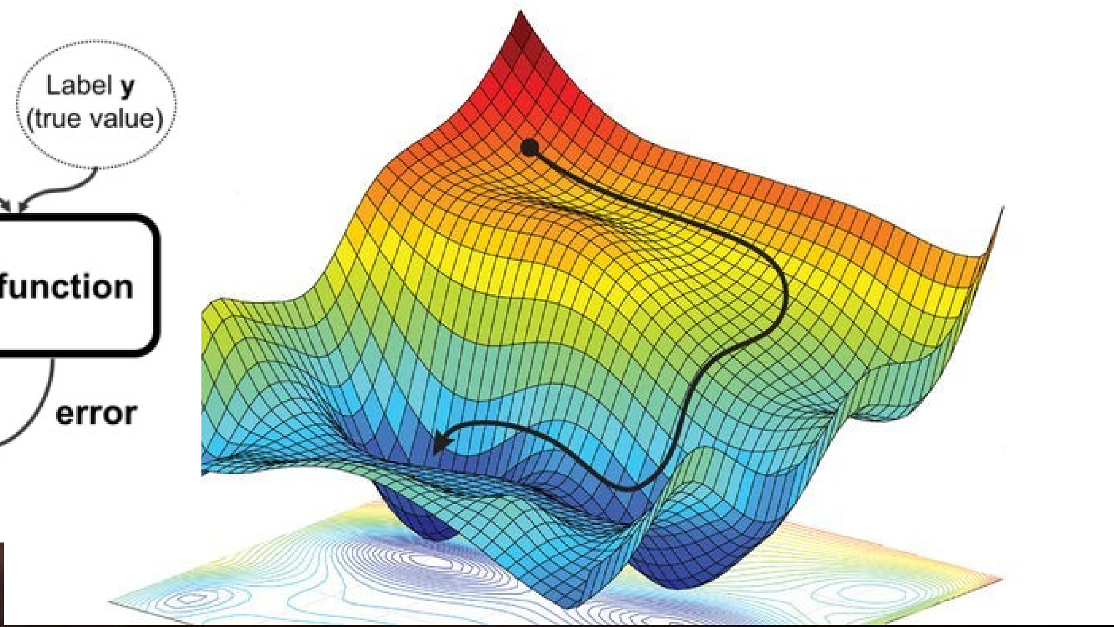


Training process

- **Loss Function.** The function used to estimate the performance of a model with a specific set of weights on examples from the training dataset.
- **Weight Initialization.** The procedure by which the initial small random values are assigned to model weights at the beginning of the training process.
- **Batch Size.** The number of examples used to estimate the error gradient before updating the model parameters.
- **Learning Rate.** The amount that each model parameter is updated per cycle of the learning algorithm.
- **Epochs.** The number of complete passes through the training dataset before the training process is terminated.

Training process: Gradienr decent





Loss functions

- A loss function is one of the parameters required to quantify how close a particular neural network is to the ideal weight during the training process.
- Mean Absolute Error (L1 Loss)
- Mean Squared Error (L2 Loss)
- Cross-Entropy(a.k.a Log loss)
- Relative Entropy(a.k.a Kullback–Leibler divergence)

$$MAE = \frac{1}{N}\sum_{i=1}^{N}|y_i - \hat{y}|$$

$$H(X) = \begin{cases} -\int_x p(x)\log p(x), & \text{if } X \text{ is continous} \\ \sum_x p(x)\log p(x), & \text{if } X \text{ is discrete} \end{cases}$$

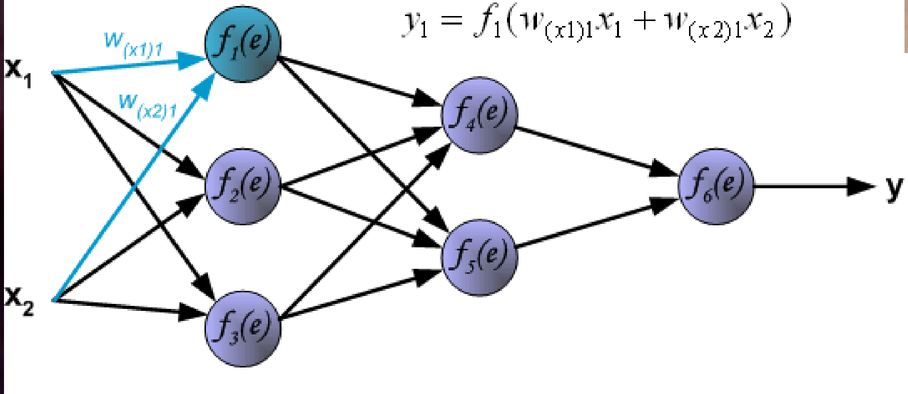
$$MSE = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y})^2$$

$$KL(P||Q) = \sum p_i(x) log(\frac{p_i(x)}{q_i(x)})$$

Problem	Output Type	Activation Function	Loss Function
Regression	Numerical	Linear	Mean Squared Error
Classification	Binary	Sigmoid	Binary Cross Entropy
Classification	Single Label, Multiple Class	Softmax	Cross Entropy
Classification	Multiple Label, Multiple Class	Sigmoid	Binary Cross Entropy

Forward and Back propagation

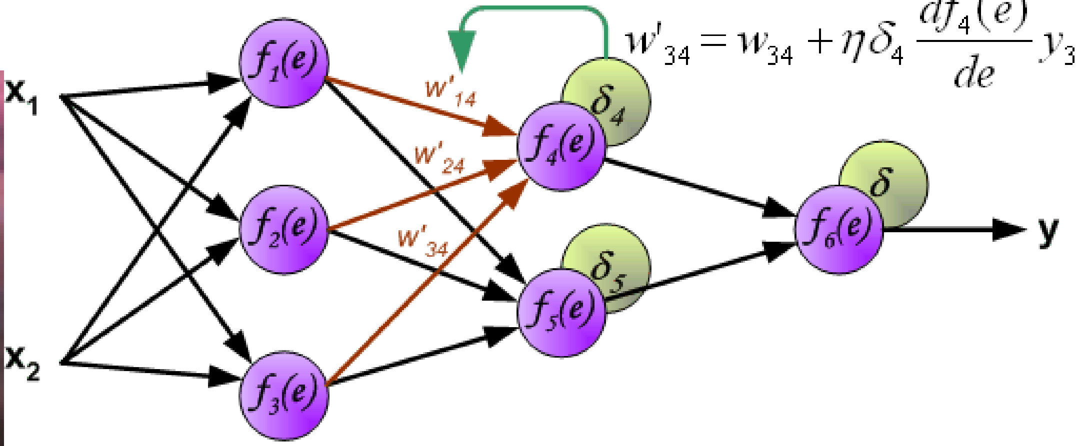
$$y_1 = f_1(w_{(x1)1}x_1 + w_{(x2)1}x_2)$$



$$w'_{14} = w_{14} + \eta\delta_4 \frac{df_4(e)}{de} y_1$$

$$w'_{24} = w_{24} + \eta\delta_4 \frac{df_4(e)}{de} y_2$$

$$w'_{34} = w_{34} + \eta\delta_4 \frac{df_4(e)}{de} y_3$$

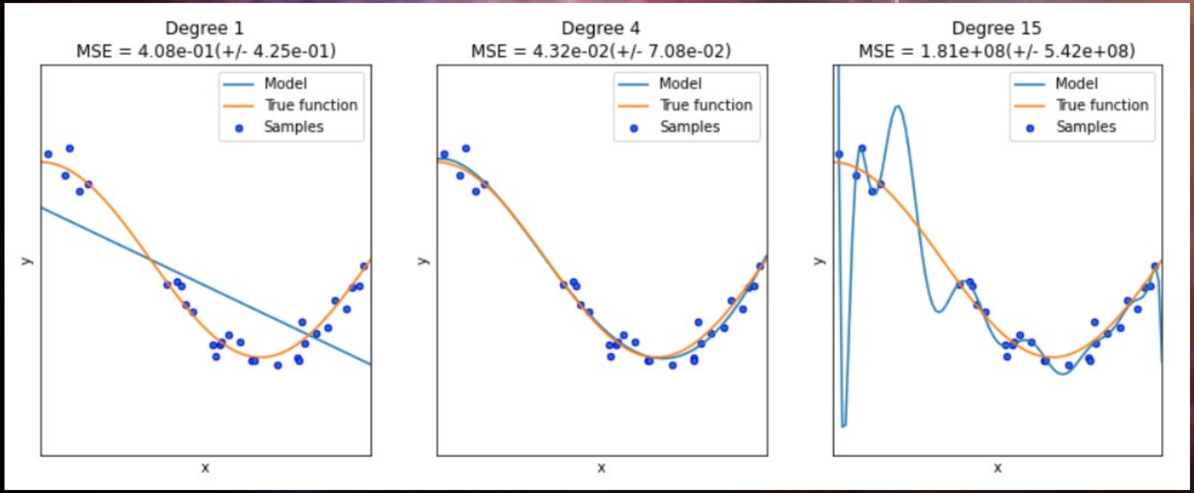


Optimazers

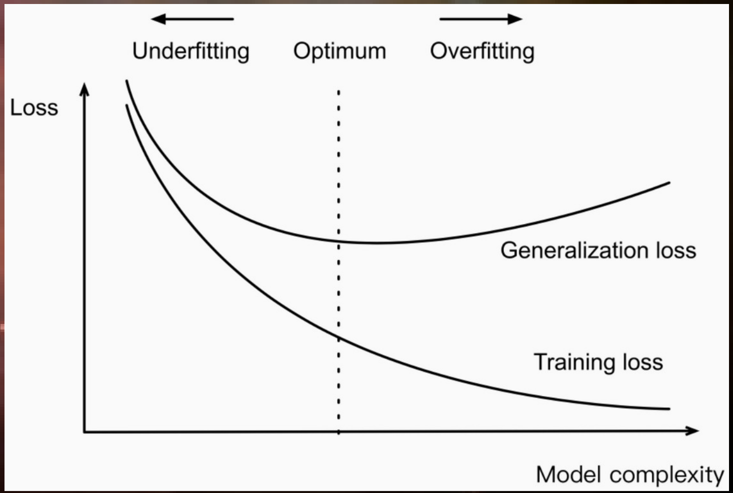
- The optimizer is another of the arguments required in the compile() method. For example Keras currently has different optimizers that can be used:
  - SGD,
  - RMSprop,
  - Adagrad,
  - Adadelta,
  - Adam,
  - Adamax,
  - Nadam.

Underfitting and overfitting

- Underfitting
  - High bias and low variance
  - The size of the training dataset used is not enough.
  - The model is too simple.
  - Training data is not cleaned and also contains noise in it.

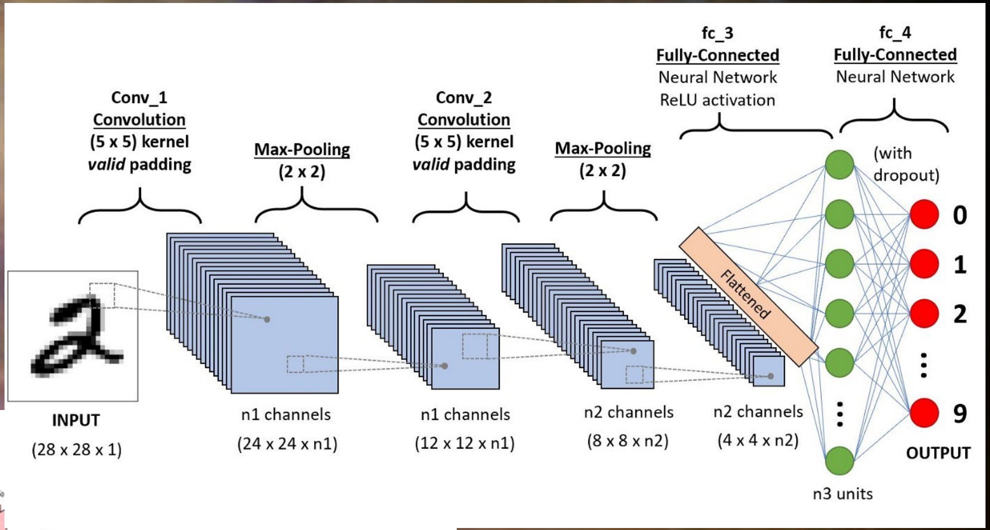


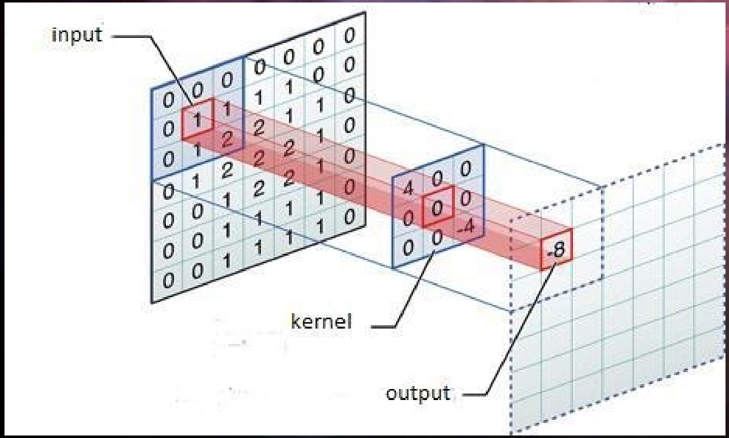
- Overfitting
  - High variance and low bias
  - The model is too complex
  - The size of the training data

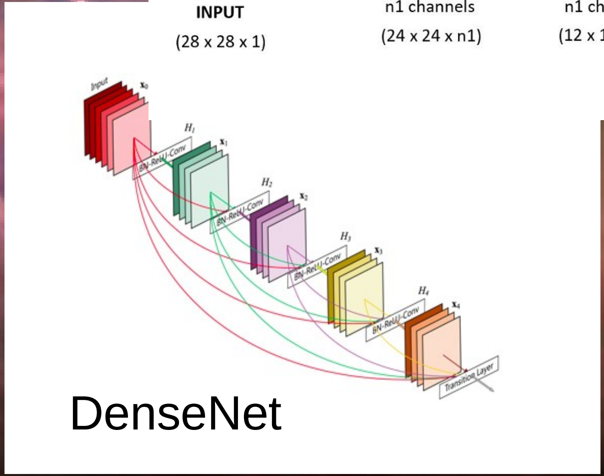


Convolutional neural network

- A convolution is a linear operation that involves the multiplication of a set of weights with the input, much like a traditional neural network. Given that the technique was designed for two-dimensional input, the multiplication is performed between an array of input data and a two-dimensional array of weights, called a filter or a kernel.

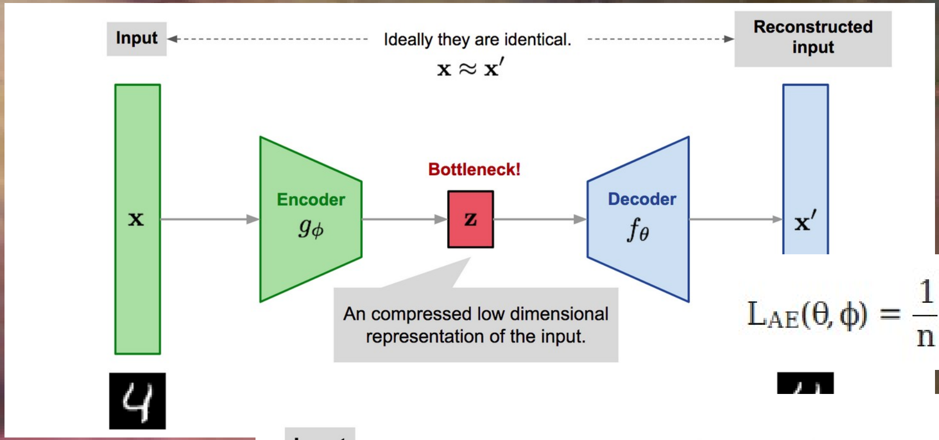






Variational Autoencoder (VAE)

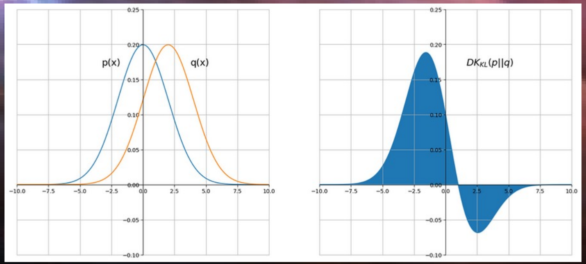
- Loss function is $L = MSE + D_{KL}$
- The Kullback-Leibler Divergence score, or KL divergence score, quantifies how much one probability distribution differs from another probability distribution.

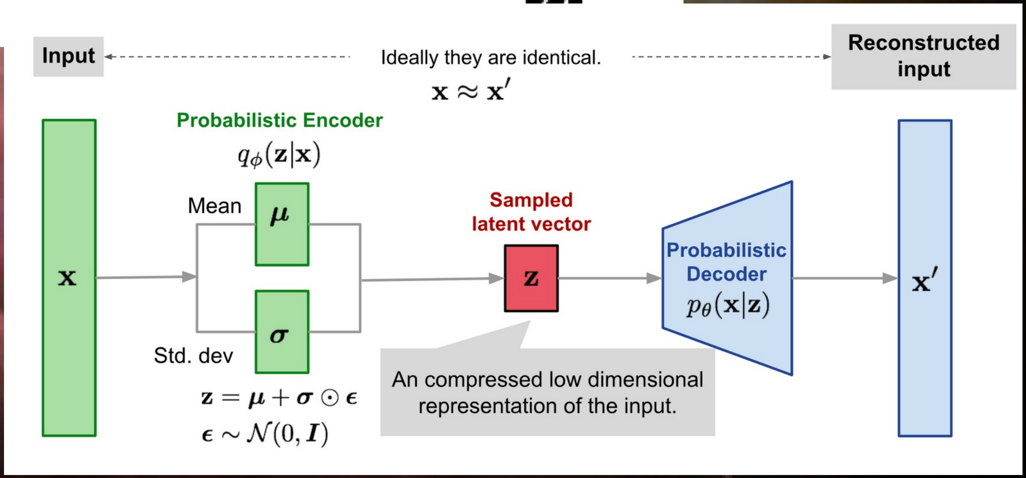


$$L_{AE}(\theta, \phi) = \frac{1}{n} \sum_{i=1}^{n} (x^{(i)} - f_\theta(g_\phi(x^{(i)})))^2$$

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

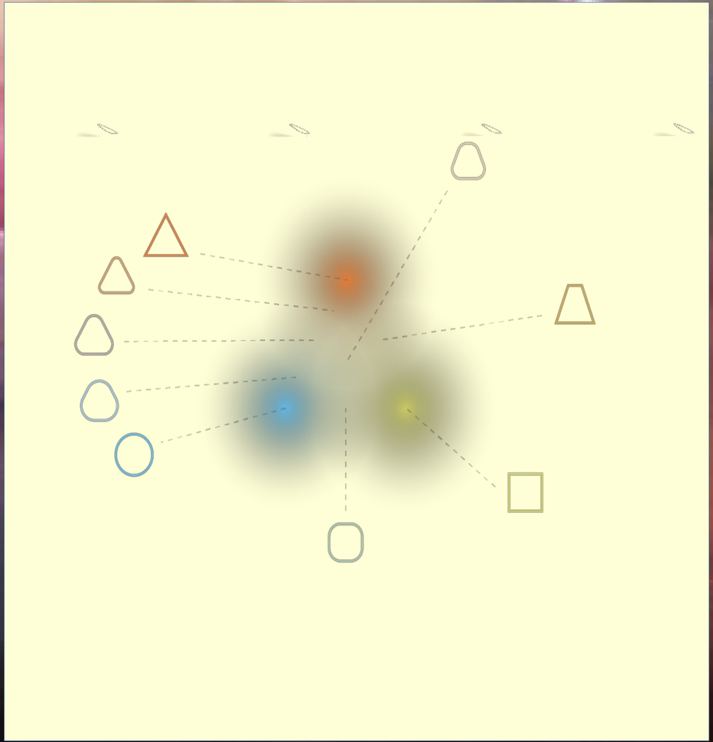
$$D_{KL}(P||Q) = \int_x P(x) \log \frac{P(x)}{Q(x)} dx$$

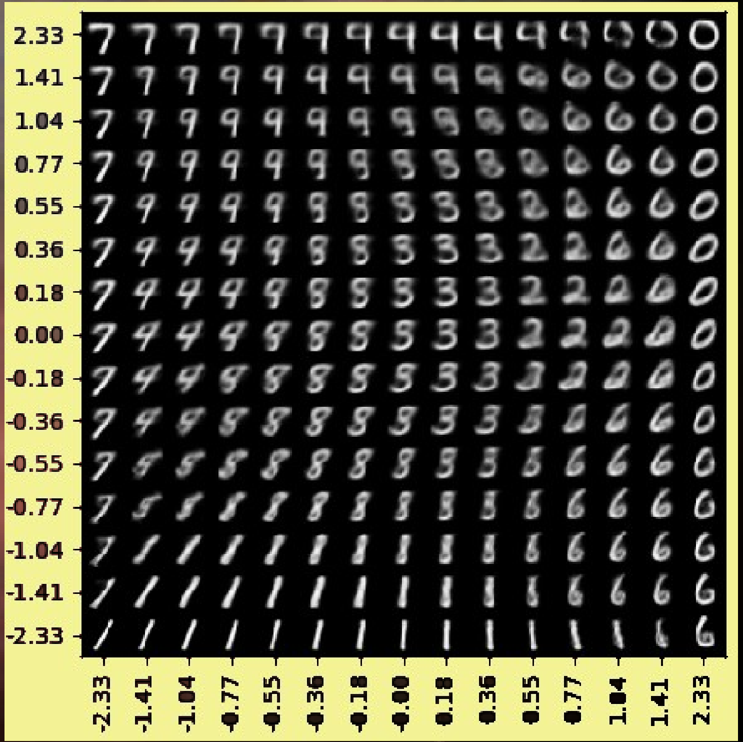




Variational Autoencoders

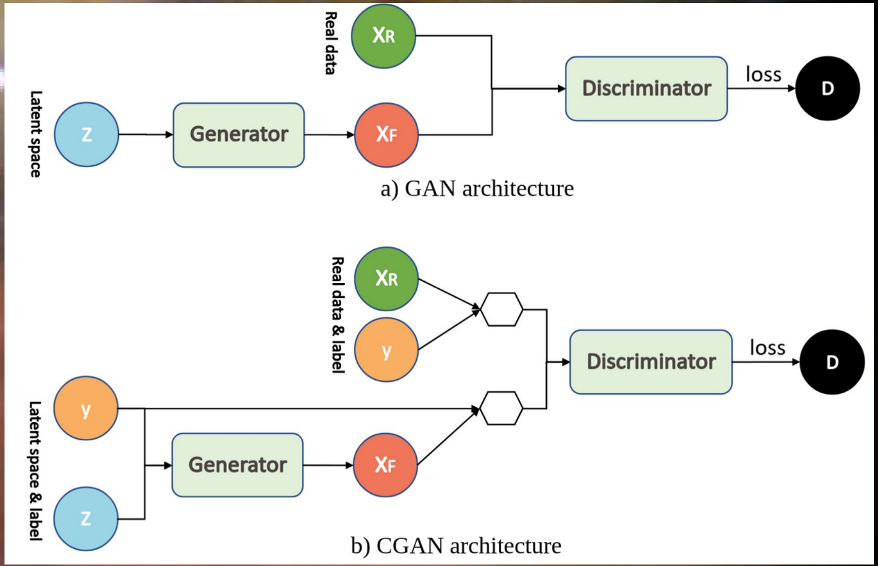
- VAE provides smooth object transformation unlike GANs which deal with classes.
- They are better suited for continuous parameters such as energy.





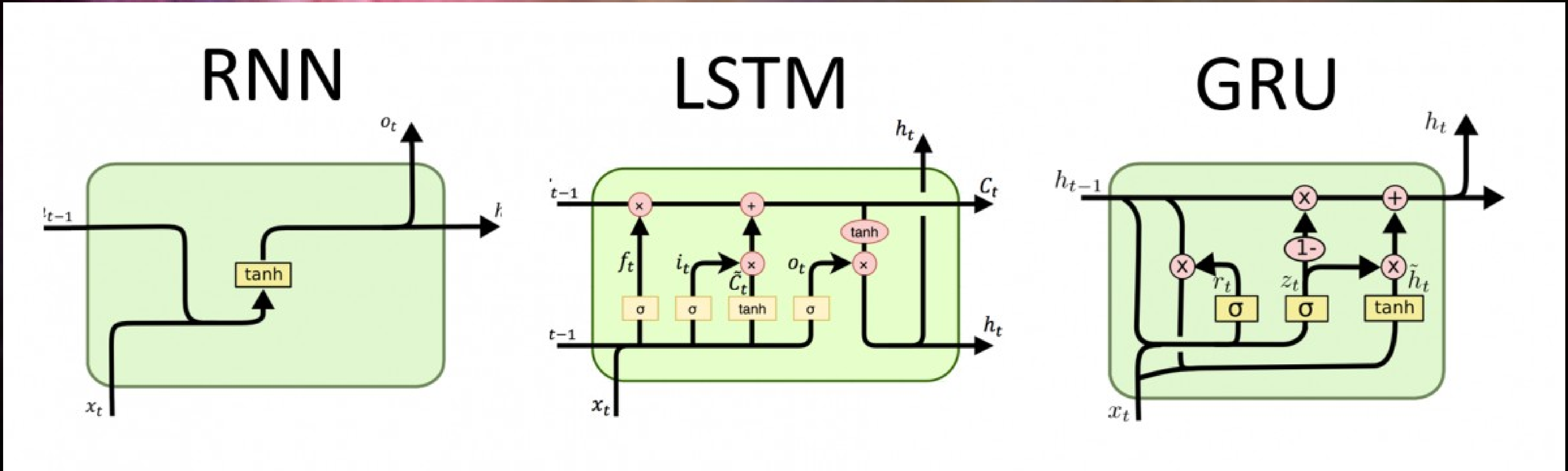
Generative adversarial network

- The training procedure for G is to maximize the probability of D making a mistake. This framework corresponds to a minimax two-player game.
- G and D are both trained simultaneously.
- Parameters for G are trained to minimize log(1-D(G(z)), and parameters for D are trained to minimize logD(x), following the above two-player min-max game with value function V(D,G).

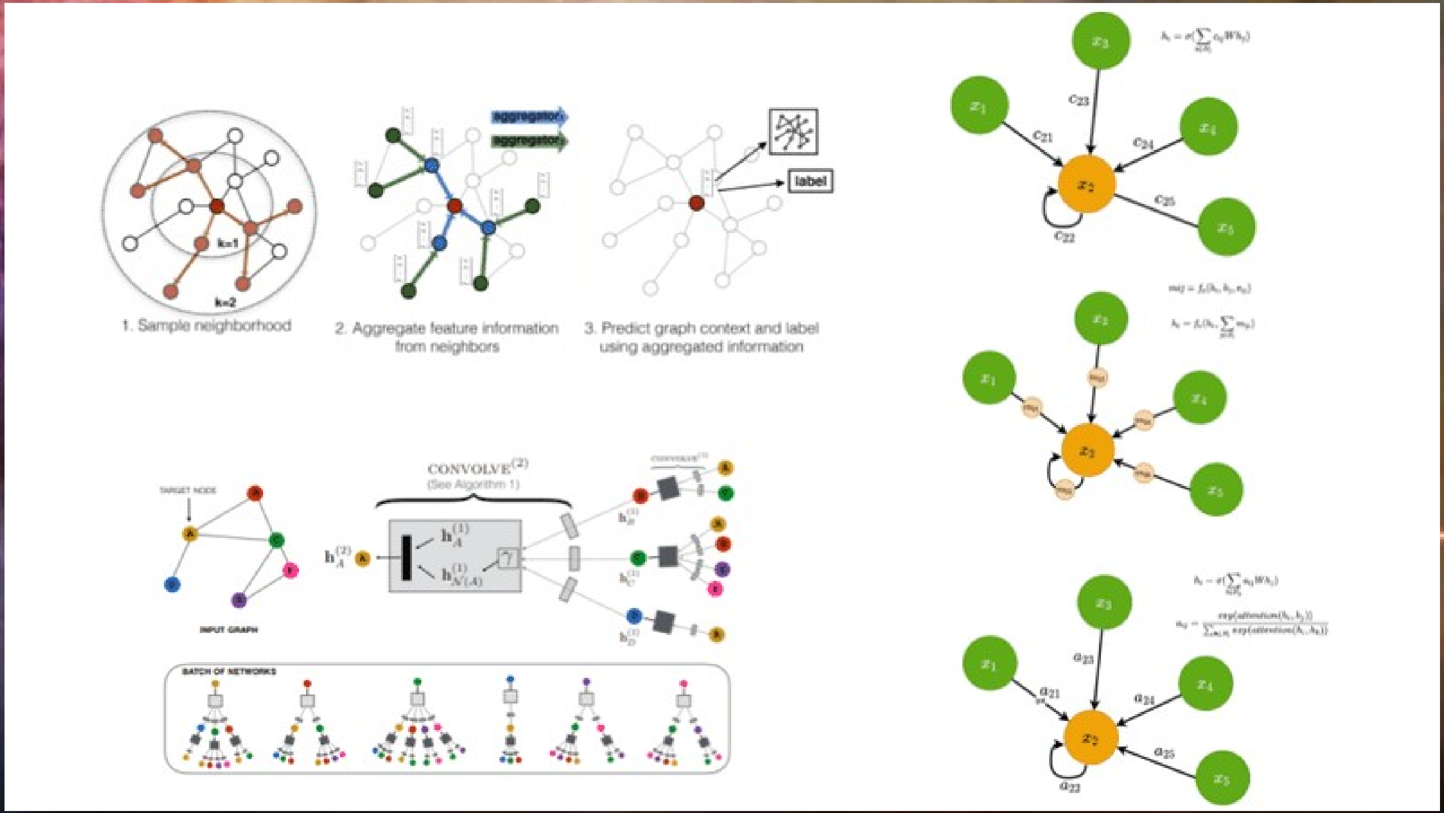


$$\min_G \max_D V(D,G) = \mathbb{E}_{\boldsymbol{x}\sim p_{\text{data}}(\boldsymbol{x})}[\log D(\boldsymbol{x})] + \mathbb{E}_{\boldsymbol{z}\sim p_{\boldsymbol{z}}(\boldsymbol{z})}[\log(1 - D(G(\boldsymbol{z})))].$$

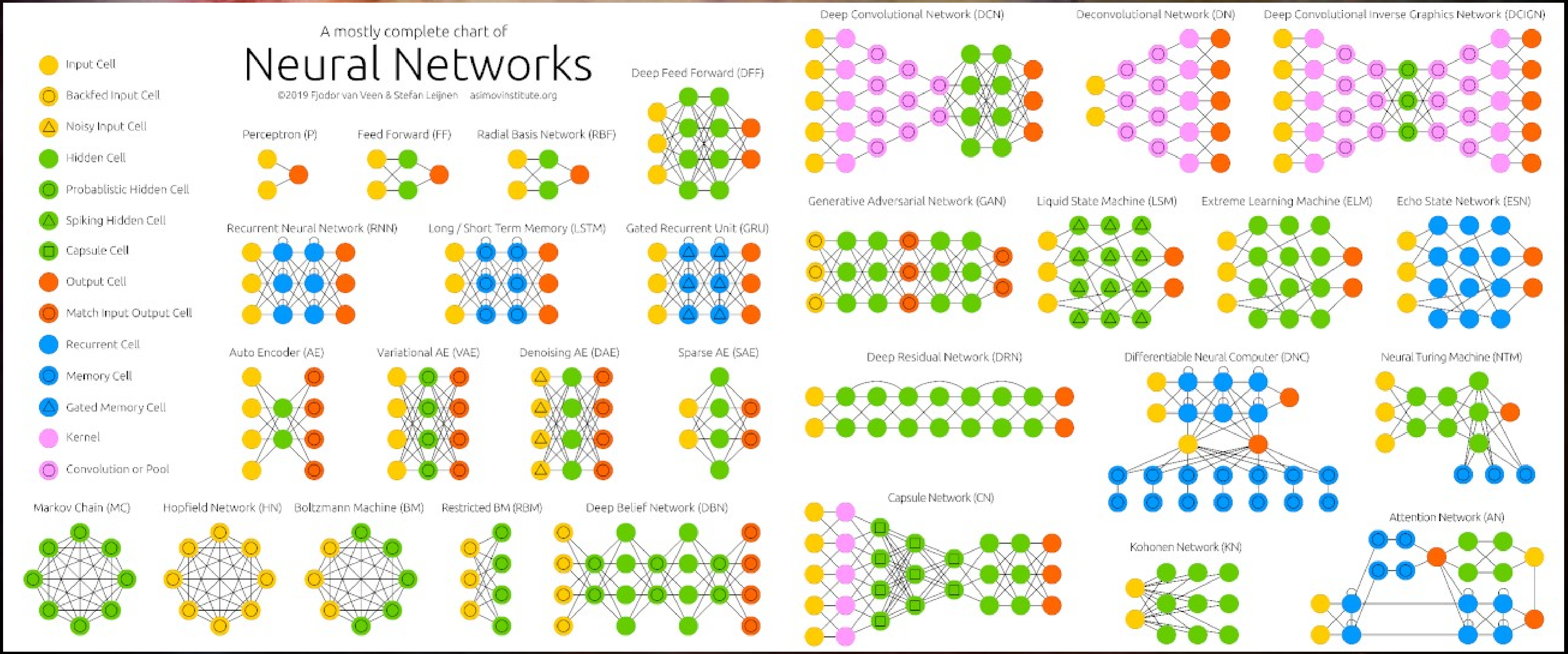
Recurrent NN



Graph NN


aggregator
aggregator
label
k=1
k=2
1. Sample neighborhood
2. Aggregate feature information from neighbors
3. Predict graph context and label using aggregated information
CONVOLVE$^{(2)}$
(See Algorithm 1)
TARGET NODE
INPUT GRAPH
BATCH OF NETWORKS

ANN chart


A mostly complete chart of
Neural Networks
©2019 Fjodor van Veen & Stefan Leijnen asimovinstitute.org
Input Cell
Backfed Input Cell
Noisy Input Cell
Hidden Cell
Probablistic Hidden Cell
Spiking Hidden Cell
Capsule Cell
Output Cell
Match Input Output Cell
Recurrent Cell
Memory Cell
Gated Memory Cell
Kernel
Convolution or Pool
Perceptron (P)
Feed Forward (FF)
Radial Basis Network (RBF)
Deep Feed Forward (DFF)
Recurrent Neural Network (RNN)
Long / Short Term Memory (LSTM)
Gated Recurrent Unit (GRU)
Auto Encoder (AE)
Variational AE (VAE)
Denoising AE (DAE)
Sparse AE (SAE)
Markov Chain (MC)
Hopfield Network (HN)
Boltzmann Machine (BM)
Restricted BM (RBM)
Deep Belief Network (DBN)
Deep Convolutional Network (DCN)
Deconvolutional Network (DN)
Deep Convolutional Inverse Graphics Network (DCIGN)
Generative Adversarial Network (GAN)
Liquid State Machine (LSM)
Extreme Learning Machine (ELM)
Echo State Network (ESN)
Deep Residual Network (DRN)
Differentiable Neural Computer (DNC)
Neural Turing Machine (NTM)
Capsule Network (CN)
Kohonen Network (KN)
Attention Network (AN)

Success stories (non-Physics)

- Natural language translation
- Self-Driving Cars
- Artificial generated objects
- Text-to-Image Translation

• Natural language translation

C. Glaser, S. McAleer, S. Stjärnholm, P. Baldi, S. W. Barwick. Deep learning reconstruction of the neutrino direction and energy from in-ice radio detector data // https://arxiv.org/abs/2205.15872

From abstract.
A large radio detector is currently being constructed in Greenland with the potential to measure the first UHE neutrino, and an order-of-magnitude more sensitive detector is being planned with IceCube-Gen2. For such shallow radio detector stations, we present an end-to-end reconstruction of the neutrino energy and direction using deep neural networks (DNNs). The DNN determines the energy with a standard deviation of a factor of two around the true energy, which meets the science requirements of UHE neutrino detectors.

Deep learning reconstruction of the neutrino direction and energy from in-ice radio detector data

C. Glaser[a], S. McAleer[b], S. Stjärnholm[a], P. Baldi[b], S. W. Barwick[c]

[a] Uppsala University Department of Physics and Astronomy, Uppsala SE-75237, Sweden.
[b] Department of Information and Computer Science, University of California, Irvine, CA 92697, USA.
[c] Department of Physics and Astronomy, University of California, Irvine, CA 92697, USA.

Abstract

Ultra-high-energy (UHE) neutrinos ($> 10^{16}$ eV) can be measured cost-effectively using in-ice radio detection, which has been explored successfully in pilot arrays. A large radio detector is currently being constructed in Greenland with the potential to measure the first UHE neutrino, and an order-of-magnitude more sensitive detector is being planned with IceCube-Gen2. For such shallow radio detector stations, we present an end-to-end reconstruction of the neutrino energy and direction using deep neural networks (DNNs). The DNN determines the energy with a standard deviation of a factor of two around the true energy, which meets the science requirements of UHE neutrino detectors. For the first time, we are able to predict the neutrino direction well for all event topologies including the complicated electron neutrino charged-current (ν_e-CC) interactions, a significant improvement compared to previous approaches. The obtained angular resolution follows a Gaussian distribution with $\sigma \approx 0.6^\circ (0.8^\circ)$ with extended tails that push the 68% quantile to $4^\circ (5^\circ)$ for non-ν_e-CC and ν_e-CC interactions, respectively. This highlights the advantages of DNNs for modeling the complex correlations in radio detector data, thereby enabling measurement of neutrino energy and direction.

1. Introduction

The detection of ultra-high-energy (UHE) neutrinos is a key to solving the 100-year-old mystery of the origin of cosmic rays and is one of the crucial milestones for astroparticle physics [1, 2]. Their detection gives access to the most violent phenomena in the universe, those that happen in the vicinity of supermassive black holes (active galactic nuclei), in neutron star mergers, or in gamma-ray bursts. Furthermore, it allows for fundamental measurements of neutrino cross-sections and flavor ratios at energies beyond the reach of Earth-based accelerators like the LHC [3, 4].

A cost-efficient way to measure these UHE neutrinos above 30 PeV of energy is via a sparse array of radio antenna stations installed, for instance, in the Arctic or Antarctic ice [5, 6, 7, 8, 9, 10]: A neutrino interaction in the ice generates a few-nanoseconds-long radio flash that can be detected from kilometer-long distances due to the large attenuation length of radio signals in ice. Because of the small flux, no UHE neutrino has been observed yet, but the technology has already been shown to work reliably with small test-bed arrays such as ARA and ARIANNA [8, 5]. With the Radio Neutrino Observatory in Greenland (RNO-G) a much larger detector is being constructed at the moment [11] and an order-of-magnitude more sensitive radio detector is foreseen for IceCube-Gen2 [9, 10].

With the first detection of a UHE neutrino on the horizon for the next years, the development of reconstruction methods becomes increasingly important. In addition, a good estimation of the energy and pointing resolution for different detector designs is crucial for planning IceCube-Gen2, which is happening at the moment. Two different station designs have been established: In the first, a *deep* design (as explored by ARA [12]) antennas are placed into narrow boreholes down to a depth of up to 200 m, thereby increasing the sensitivity to neutrinos per detector station but also increasing the costs per station and limiting the choice of available antennas due to the narrow borehole. The second design is a *shallow* detector station (as explored by ARIANNA [13]) with high-gain LPDA antennas installed a few meters below the surface. The Radio Neutrino Observatory in Greenland (RNO-G) combines both designs into hybrid detector stations. The radio detector of IceCube-Gen2 foresees a hybrid array of shallow-only stations interspersed with hybrid stations [10].

This work focuses on a shallow station design as shown in Fig. 1, which has been explored by the ARIANNA test-bed detector on the Ross Ice Shelf and at the South Pole [8]. Each station consists of 4 LPDA antennas installed at a depth of just a few meters below the snow surface, and 1 dipole antenna installed at a depth of 10 m to 15 m in a narrow borehole. These antennas observe the ice below for neutrino interactions. The dipole antenna was added

Email addresses: christian.glaser@physics.uu.se (C. Glaser), smcaleer@uci.edu (S. McAleer), Sigfrid.Stjarnholm.2141@student.uu.se (S. Stjärnholm), pfbaldi@uci.edu (P. Baldi), sbarwick@uci.edu (S. W. Barwick)

Preprint submitted to Journal of Astroparticle Physics *June 1, 2022*

• Natural language translation

En	A large radio detector is currently being constructed in Greenland with the potential to measure the first UHE neutrino, and an order-of-magnitude more sensitive detector is being planned with IceCube-Gen2. For such shallow radio detector stations, we present an end-to-end reconstruction of the neutrino energy and direction using deep neural networks (DNNs). The DNN determines the energy with a standard deviation of a factor of two around the true energy, which meets the science requirements of UHE neutrino detectors.
Ru	В настоящее время в Гренландии строится большой радиодетектор, который может измерять первое сверхвысокоэнергетическое нейтрино, а с помощью IceCube-Gen2 планируется создать на порядок более чувствительный детектор. Для таких неглубоких станций радиодетекторов мы представляем сквозную реконструкцию энергии и направления нейтрино с использованием глубоких нейронных сетей (ГНС). DNN определяет энергию со стандартным отклонением в два раза относительно истинной энергии, что соответствует научным требованиям детекторов нейтрино UHE.
Rev-En	A large radio detector is currently being built in Greenland that can measure the first ultra-high-energy neutrino, and IceCube-Gen2 is planned to be an order of magnitude more sensitive detector. For such shallow radio detector stations, we present end-to-end reconstruction of neutrino energy and direction using deep neural networks (DNNs). DNN determines energy with a standard deviation of twice the true energy, which is in line with the scientific requirements of UHE neutrino detectors.

• Natural language translation

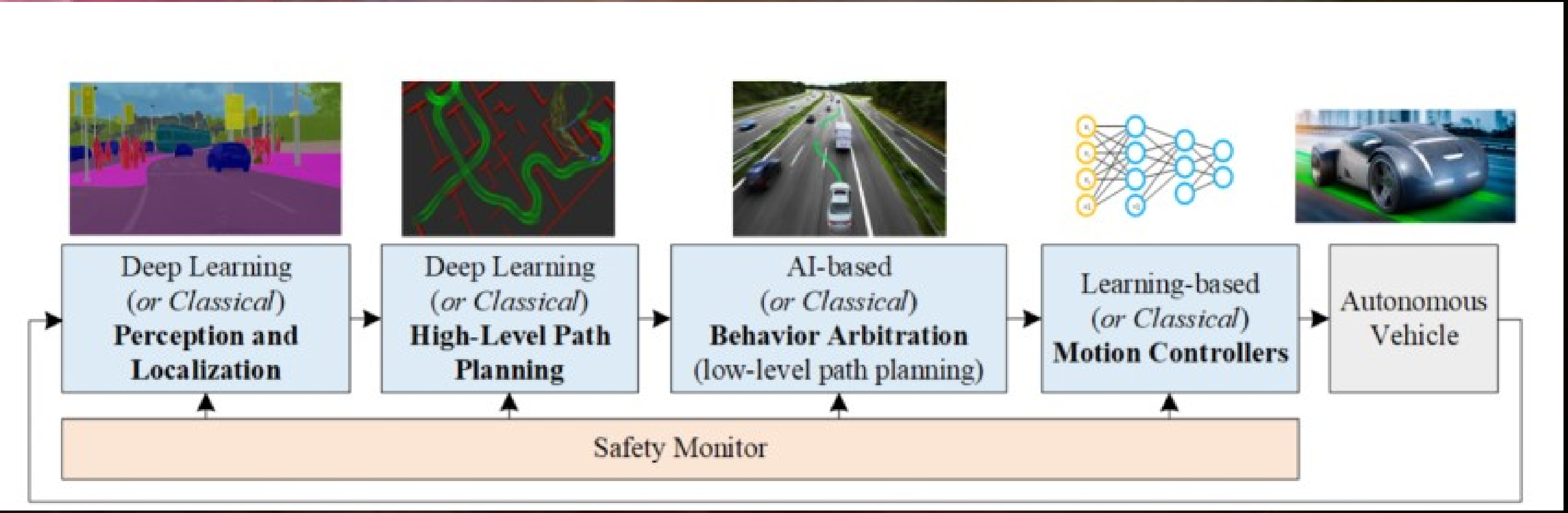
En	A large radio detector is currently being constructed in Greenland with the potential to measure the first UHE neutrino, and an order-of-magnitude more sensitive detector is being planned with IceCube-Gen2. For such shallow radio detector stations, we present an end-to-end reconstruction of the neutrino energy and direction using deep neural networks (DNNs). The DNN determines the energy with a standard deviation of a factor of two around the true energy, which meets the science requirements of UHE neutrino detectors.
Cn	目前正在格陵蘭建造一個大型無線電探測器，有可能測量第一個 UHE 中微子，並且正在計劃使用 IceCube-Gen2 建造一個數量級的更靈敏的探測器。 對於這樣的淺無線電探測器站，我們使用深度神經網絡 (DNN) 提出了中微子能量和方向的端到端重建。 DNN 確定能量的標準差為真能量的兩倍，滿足 UHE 中微子探測器的科學要求。
Rev-En	A large radio detector is currently under construction in Greenland with the potential to measure the first UHE neutrinos, and plans are underway to build an order of magnitude more sensitive detector using IceCube-Gen2. For such shallow radio detector stations, we propose an end-to-end reconstruction of neutrino energy and orientation using a deep neural network (DNN). The standard deviation of the energy determined by DNN is twice the true energy, which meets the scientific requirements of UHE neutrino detectors.

• Natural language translation

En	A large radio detector is currently being constructed in Greenland with the potential to measure the first UHE neutrino, and an order-of-magnitude more sensitive detector is being planned with IceCube-Gen2. For such shallow radio detector stations, we present an end-to-end reconstruction of the neutrino energy and direction using deep neural networks (DNNs). The DNN determines the energy with a standard deviation of a factor of two around the true energy, which meets the science requirements of UHE neutrino detectors.
Ar	ويتم التخطيط لكاشف أكثر حساسية بترتيب من حيث ، UHE يجري حاليًا إنشاء كاشف لاسلكي كبير في جرينلاند مع إمكانية قياس أول نيوترينو بالنسبة لمحطات الكشف الراديوي الضحلة ، نقدم إعادة بناء شاملة لطاقة النيوترينو واتجاهها باستخدام .IceCube-Gen2 الحجم باستخدام الطاقة بانحراف معياري لعامل اثنين حول الطاقة الحقيقية ، والتي تلبي المتطلبات العلمية لأجهزة DNN تحدد .(DNNs) الشبكات العصبية العميقة الكشف عن النيوترينو UHE.
Rev-En	A large radio detector is being built in Greenland with the ability to measure the first UHE neutrino, and an order of magnitude more sensitive detector is planned with IceCube-Gen2. For shallow radio detection stations, we present a comprehensive reconstruction of neutrino energy and direction using deep neural networks (DNNs). DNN determines the energy with a standard deviation of a factor of two about the true energy, which meets the scientific requirements of UHE neutrino detectors.

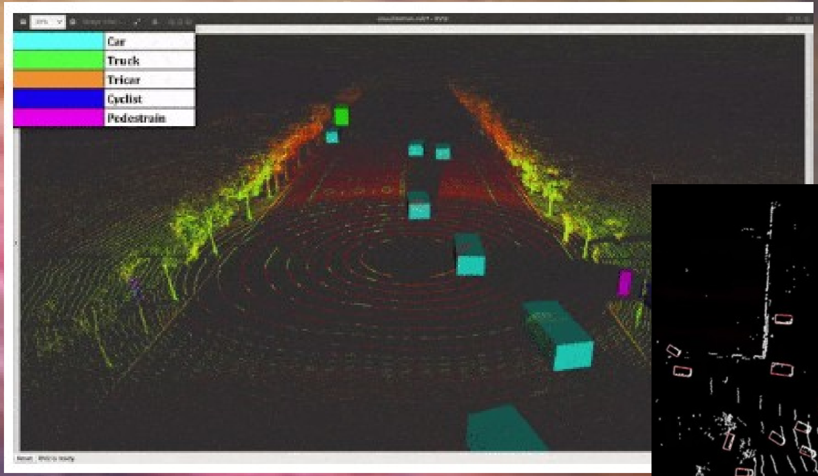
Self-driving car

- Perception
- Localization
- Prediction
- Decision Making



Self-Driving Cars

- Perception
  - Camera
  - LiDAR
  - RADAR
- Sensor data is fed into deep neural networks. The car uses the result of neural networks to predict the actions of objects or vehicles that are near it.



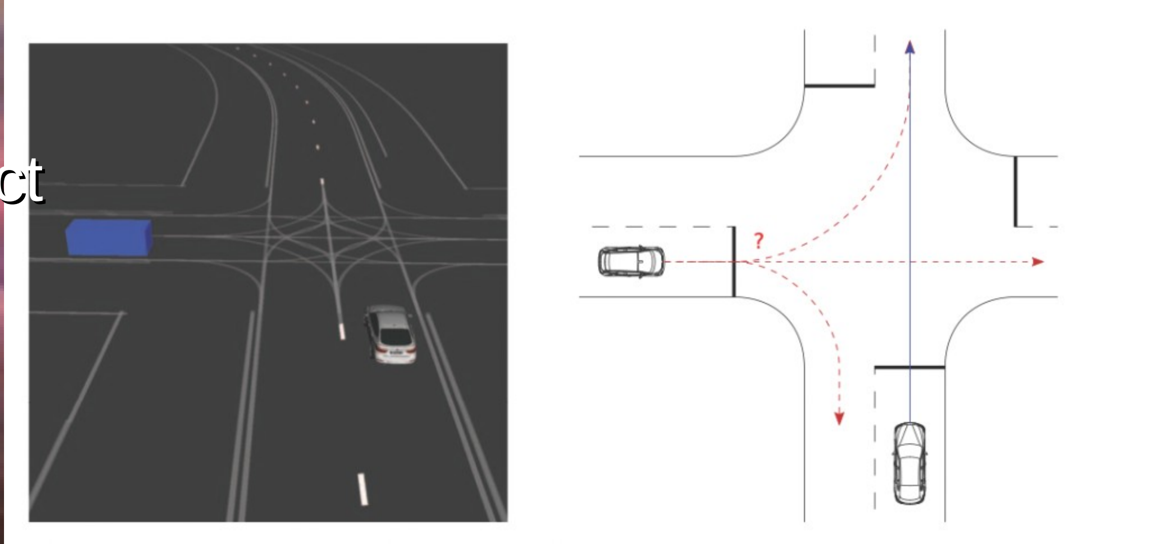
Localization

- Localization algorithms in self-driving cars calculate the position and orientation of the vehicle as it navigates – a science known as Visual Odometry (VO).

Prediction and decision-making

- Predict the next actions of drivers or pedestrians nearby.
  - This is very important for road safety.

- In order to make a decision, the car should have enough information so that it can select the necessary set of actions. Deep learning algorithms are used for localization and prediction.



Artificial generated objects

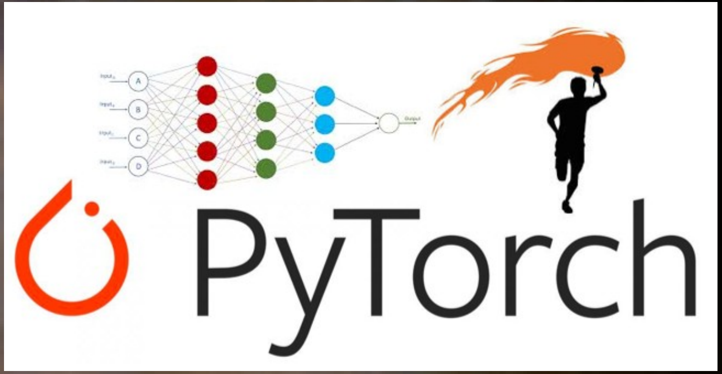
- Generate Photographs of Human Faces.
  Tero Karras, et al. “Progressive Growing of GANs for Improved Quality, Stability, and Variation”

Text-to-Image Translation

- Han Zhang, et al. in their 2016 paper titled “StackGAN: Text to Photo-realistic Image Synthesis with Stacked Generative Adversarial Networks” demonstrate the use of GANs to generate realistic looking photographs from textual descriptions of simple objects like birds and flowers.

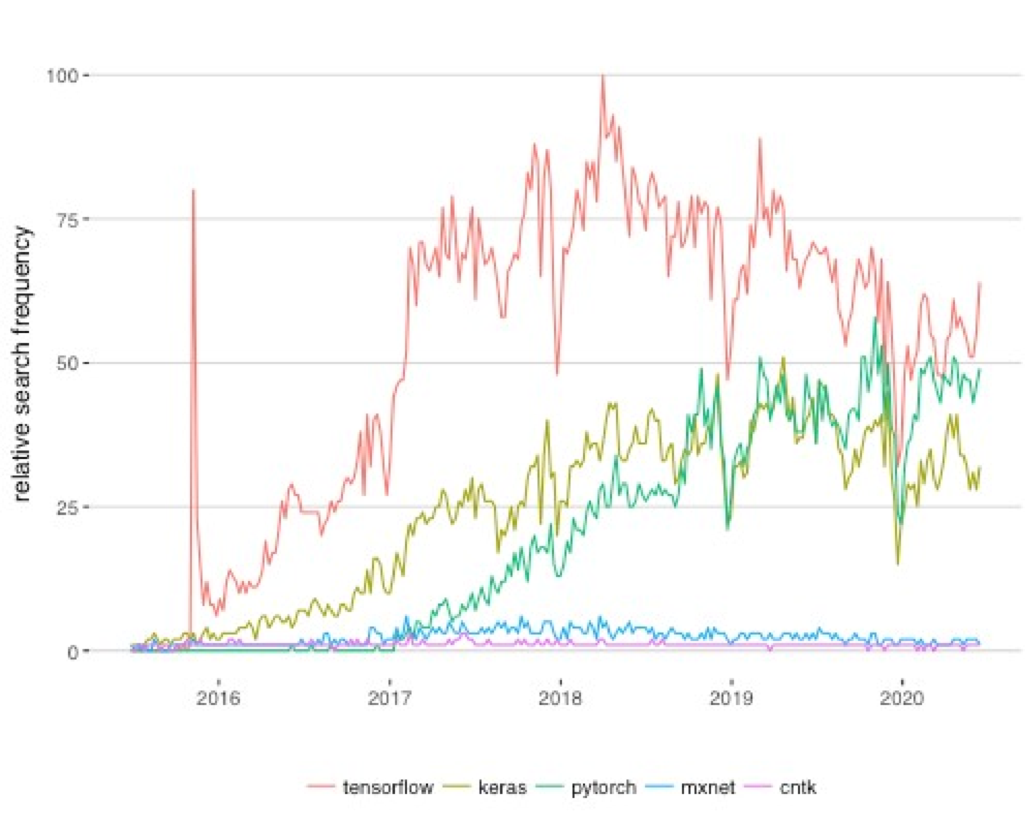
Tools for ML

- Mainly used Python libraries: PyTorch, TensorFlow+Keras
- PyTorch and TensorFlow have
  - easy GPU implementation
  - automated gradient computation for various operations
  - building blocks of commonly used models
  - online available implementations of many (trained) models
  - tools for visualization
  - lot's of online tutorials and documentation



Tools for ML

- Differences:
  - dynamic (PyTorch) vs. static (TensorFlow) graph definition
  - different ways of parallelization
  - PyTorch offers better development and debugging experience
- PyTorch or TensorFlow is a question of taste



Tools for ML

	Keras	TensorFlow	PyTorch
Level of API	high-level API[1]	Both high & low level APIs	Lower-level API[2]
Speed	Slow	High	High
Architecture	Simple, more readable and concise	Not very easy to use	Complex[3]
Debugging	No need to debug	Difficult to debugging	Good debugging capabilities
Dataset Compatibility	Slow & Small	Fast speed & large	Fast speed & large datasets
Popularity Rank	1	2	3
Uniqueness	Multiple back-end support	Object Detection Functionality	Flexibility & Short Training Duration
Created By	Not a library on its own	Created by Google	Created by Facebook[4]
Ease of use	User-friendly	Incomprehensive API	Integrated with Python language
Computational graphs used	Static graphs	Static graphs	Dynamic computation graphs[5]

Physicds applications

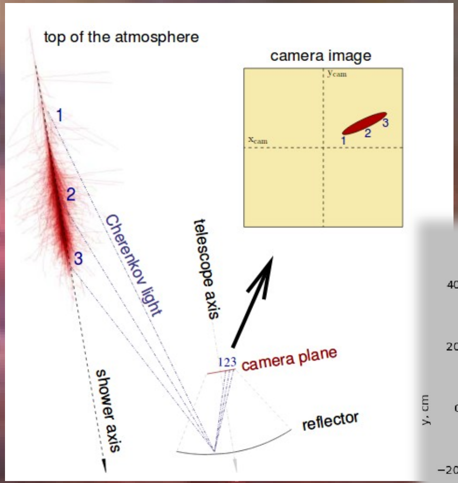
- Gamma astronomy (TAIGA)
- Neutrino physics (JUNO)
- Theory

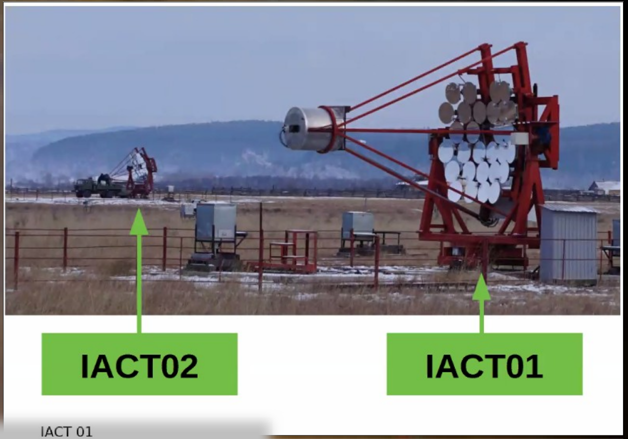
Gamma astronomy

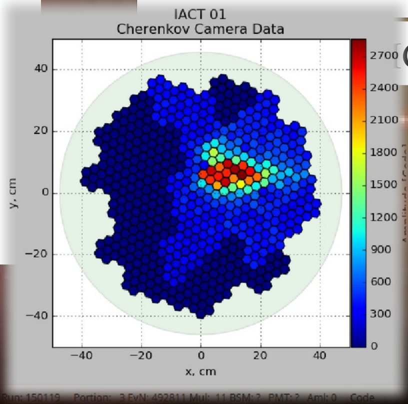
- Scientific problem: determination and study of sources of high-energy (energy of the order of tens of TeV) gamma radiation.
- Measurement of the flux, energy spectrum, direction of arrival of gamma rays helps to understand the mechanisms of generation of high energy gamma radiation and the morphology of these sources.

TAIGA-IACT

- TAIGA-IACTs are located in The Tunka valley of the republic Buryatia. Three telescopes have been installed and are operating.
- Telescopes detect Cherenkov radiation created by the Extensive Air Shower (EAS).



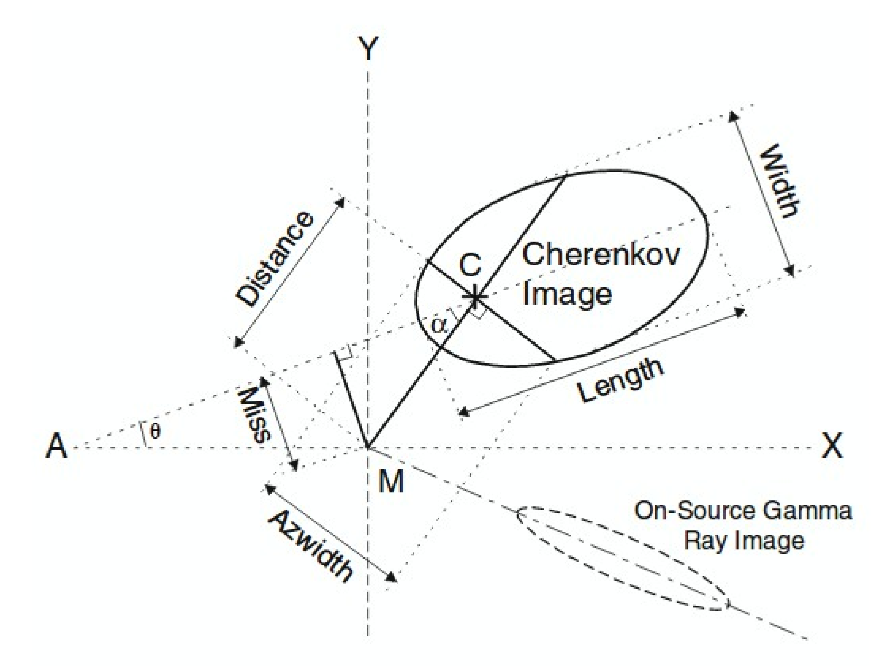




GA-IACT

Traditional image processing method

- Hillas parameters – description of the image by an ellipse with certain parameters.

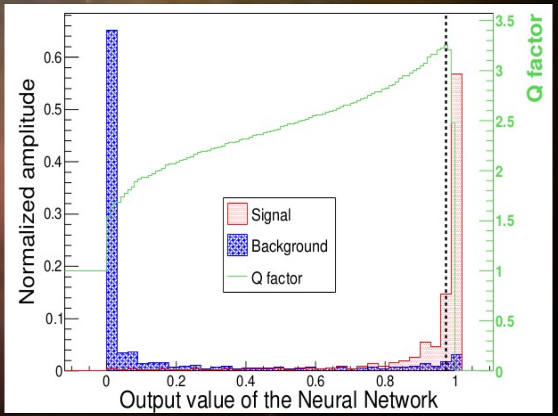


Classification of primary particles

Used NN	N_{total}	N_g	N_h	N_{g_g}	N_{h_g}	S_{after}	$Q=S_{after}/S_{before}$
User CNN	40 000	20 000	20 000	11677	180	275,6	2,22
	4 182	58	4 124	25	21	4,71	5,22
	36 783	35	36 748	13	187	0,94	5,15
ResNet	36 783	35	36 748	18	279	1,07	5,84
GoogLeNet	36 783	35	36 748	19	262	1,16	6,35

where

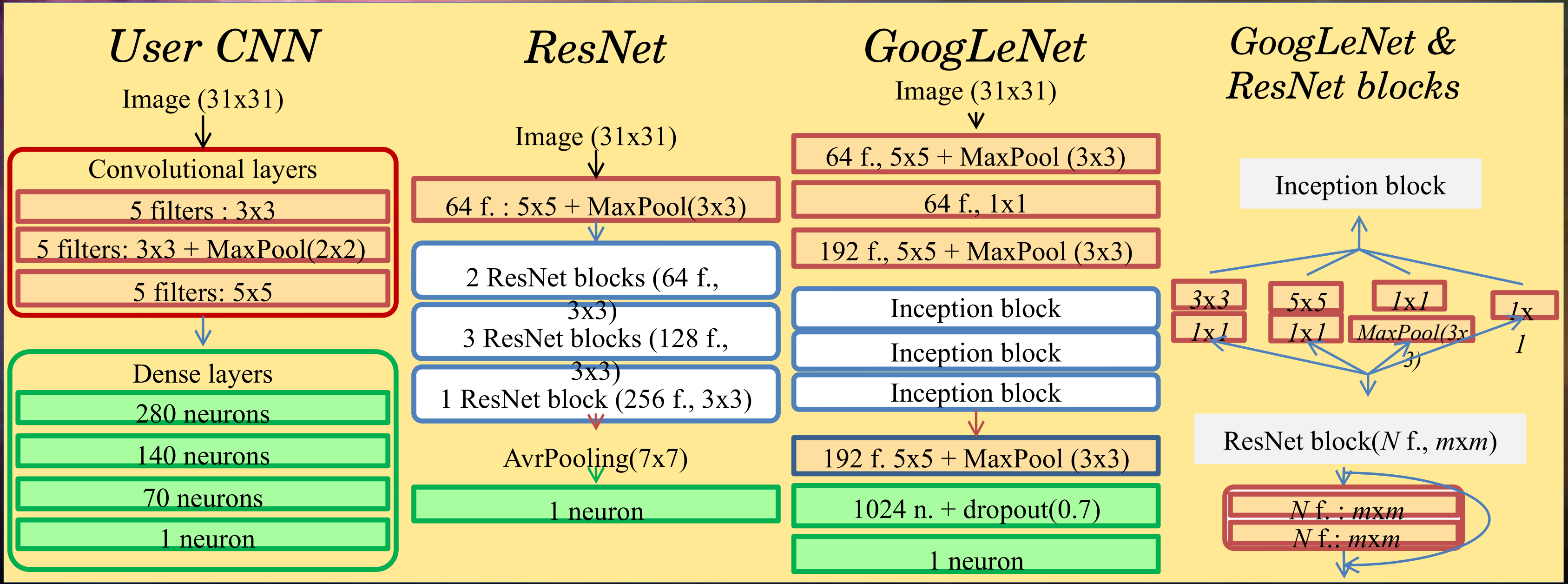
$$S_{after} = \sqrt{2\left((N_{g\_g} + N_{h\_g})\ln\left(1 + \frac{N_{g\_g}}{N_{h\_g}}\right) - N_{g\_g}\right)}$$



- Right: R. Alfaro and et.al. Gamma/Hadron Separation with the HAWC Observatory // ArXiv: 2205.12188

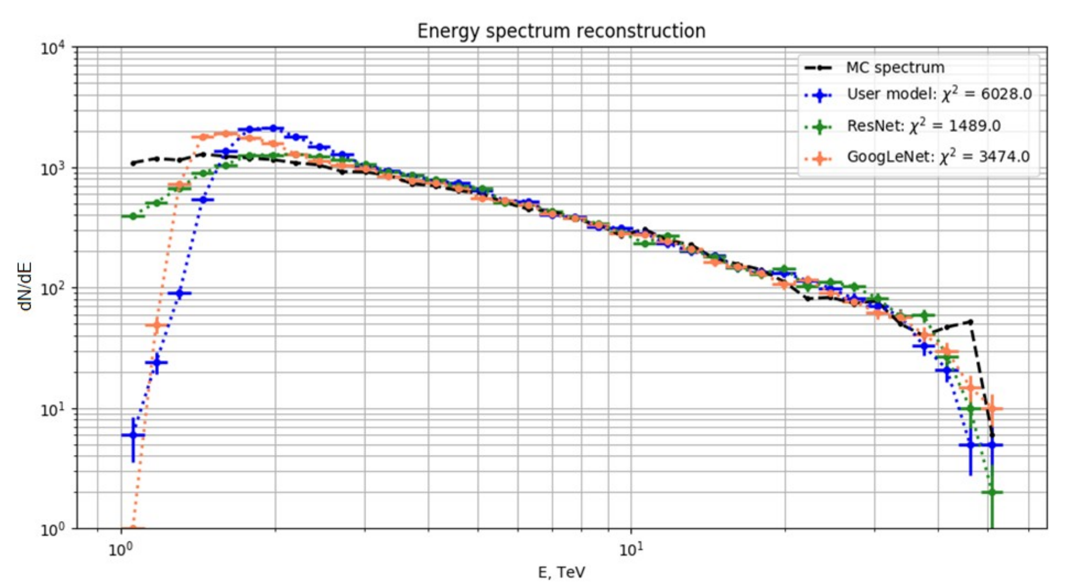
CNN architectures

- For an adequate comparison, ResNet and GoogLeNet were simplified in such a way that the number of weight coefficients for CNN networks approximately coincided. In this case, their number is ~2 millions.



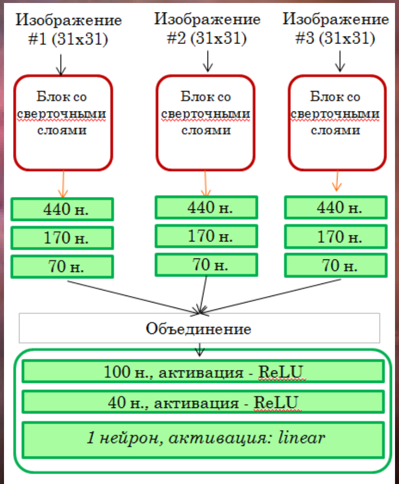
Energy spectrum reconstruction of only gamma quanta events with different CNNs

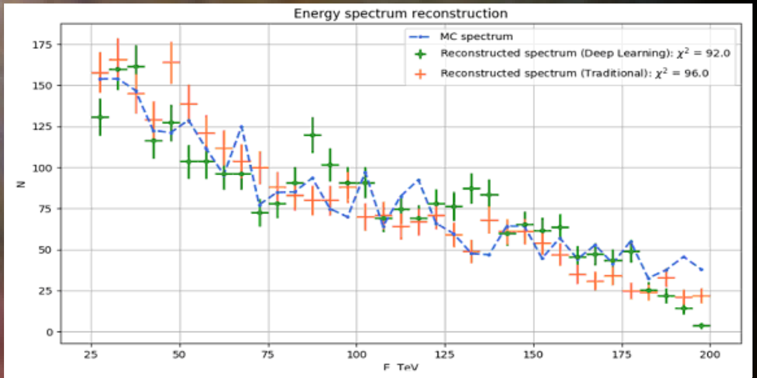
Set	Total events (gamma/proton)	Train/validation separation	Energies
Mono-mode	200 000 (100 000 / 100 000)	160 000 / 40 000	Protons: 5-100 TeV Gammas: 2-50 TeV

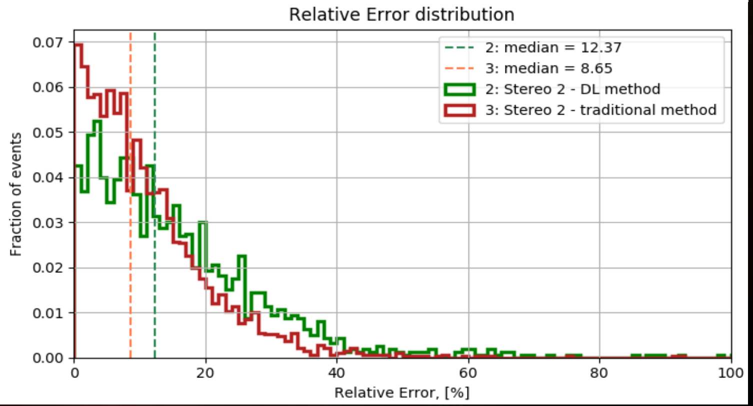


Gamma ray energy spectra. DL vs. Hillas parameters

- A comparison was made on a sample consisting of gamma rays with an energy of 25-200 TeV.
- Traditional energy recovery method: for each telescope, approximation by a function depending on some Hillas parameters (spot brightness, size, distance) and EAS characteristics (EAS maximum height).
- Deep learning method: A custom two-channel CNN (Stereo2) was chosen.

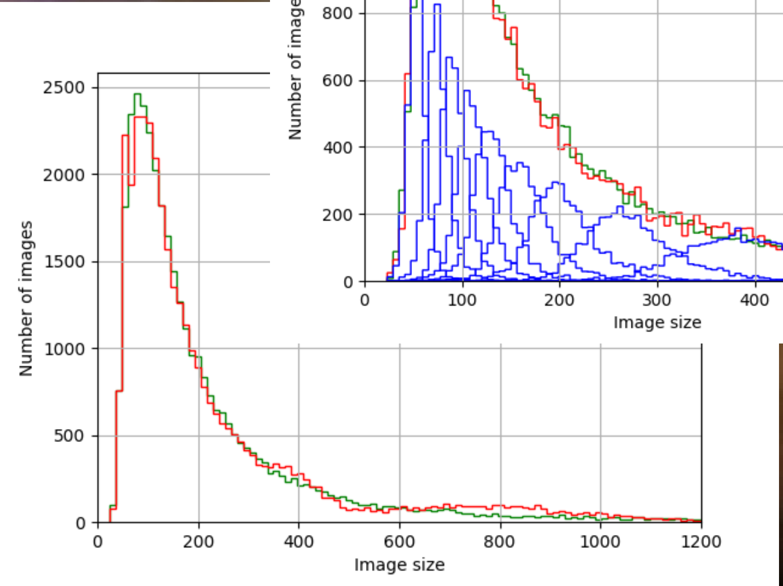


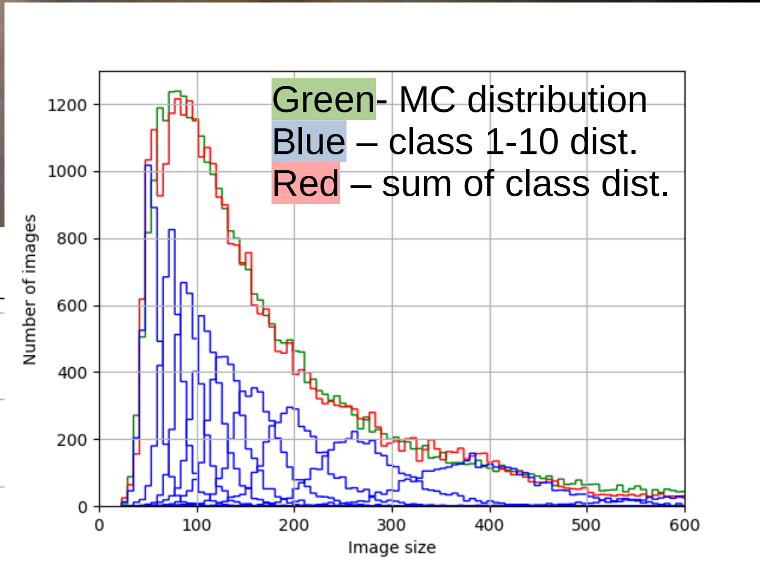




IACT event modelling.(c)GAN

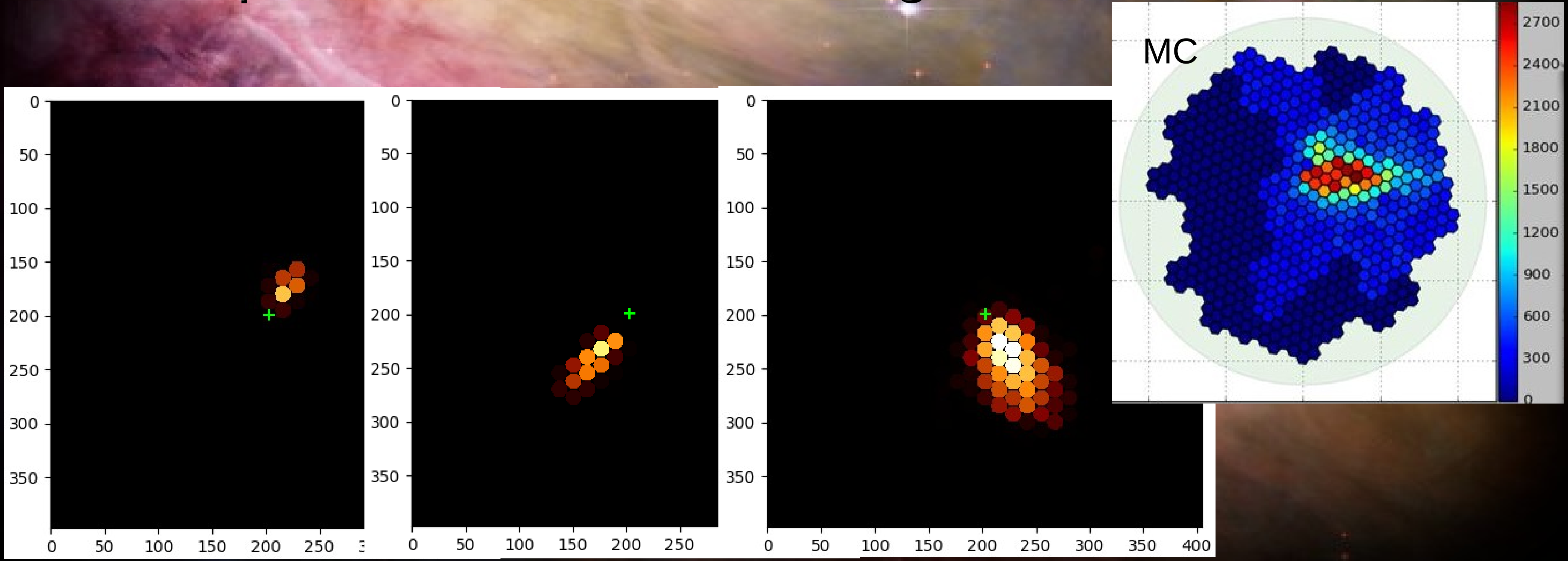
- Conditional GAN
- During training, all events were sorted by energy and divided into 10 equal parts (about 3500 events per each). Each part was considered a separate class, information about which was used on the training sample.
- When generating events by the trained network, the same number of events of each class was generated.
- Generation speed of about 5000 images per second



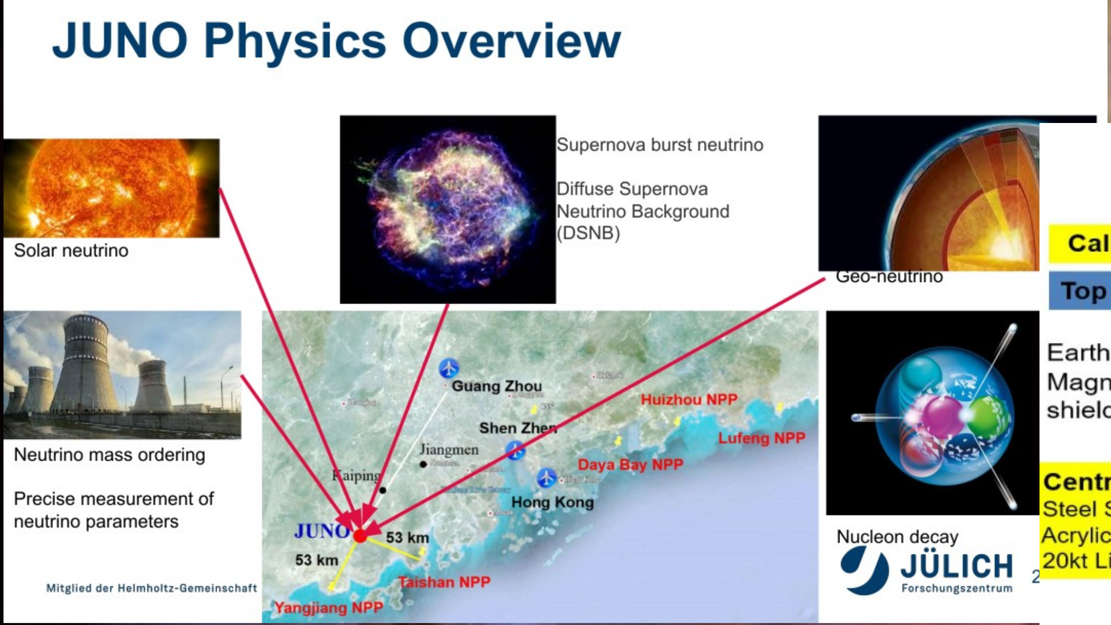


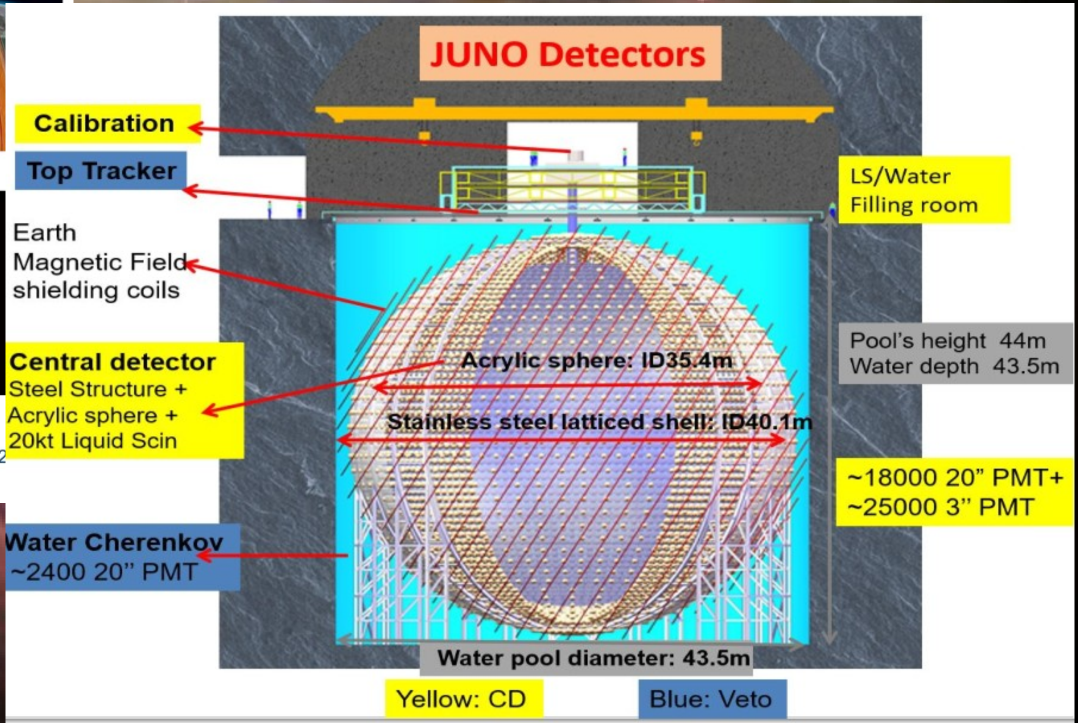
IACT event modelling. (c)VAE

- Examples of simulated IACT gamma events.



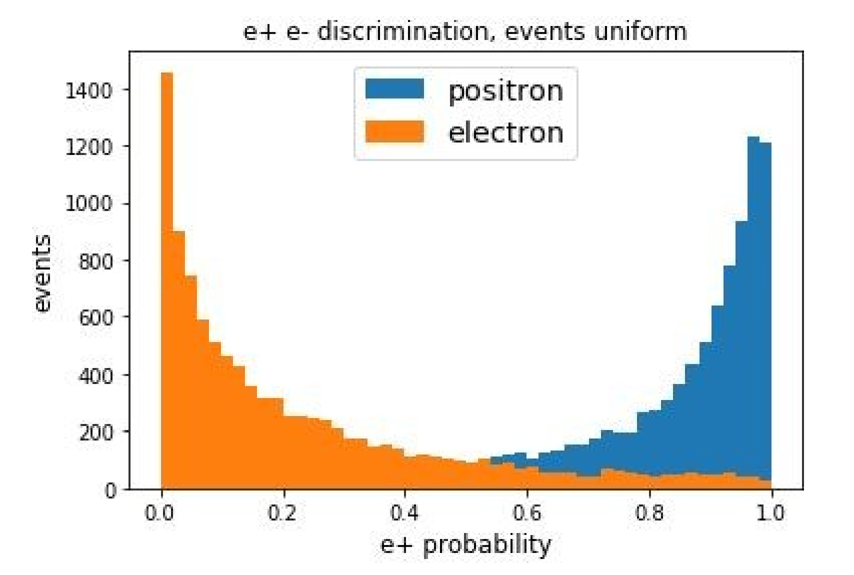
Neutrino experiments (JUNO)



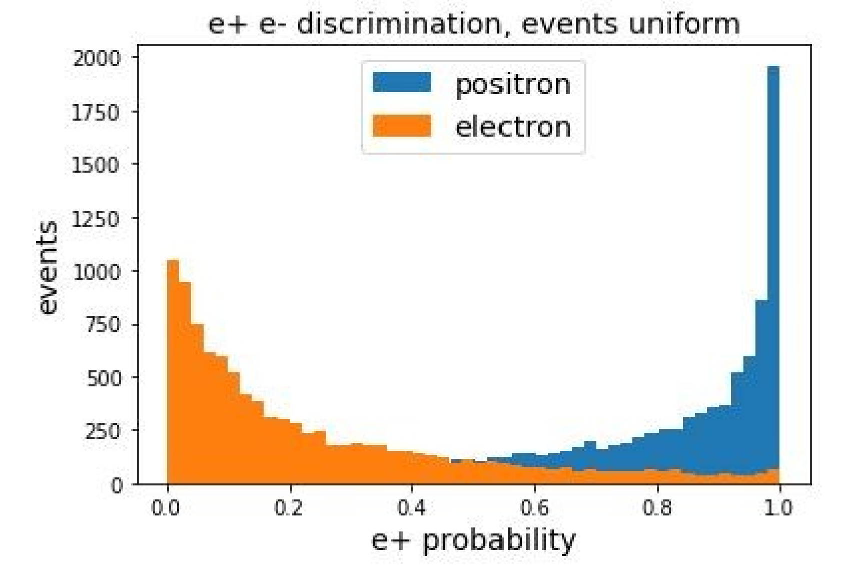


JUNO: e+/e- descrimination

Type 1: FCNN
Accuracy 87.5%



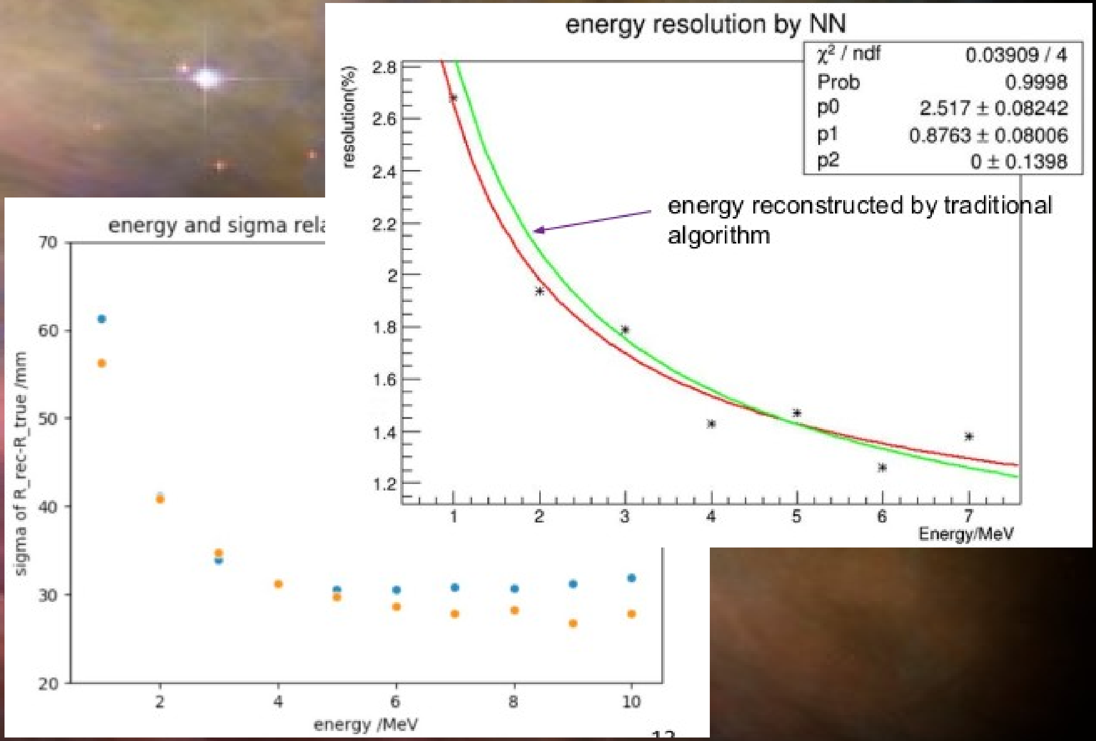
Type 1: CNN
Accuracy 83.0%



From: M.Gromov. Machine learning applications for JUNO

JUNO: Vertex and Energy Reconstruction (low energy)

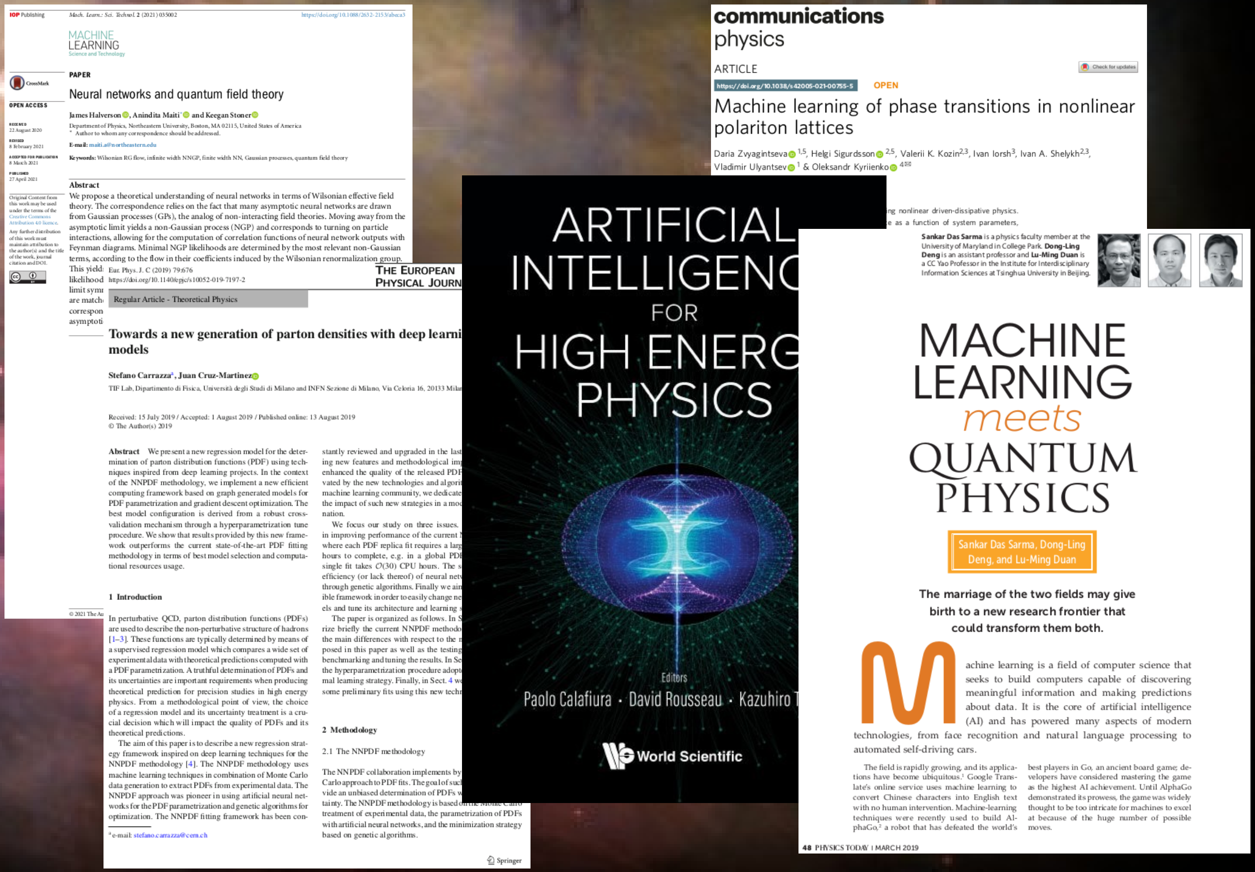
- Training data sample: 2M e + uniform in the detector, kinetic enegy continuous in (1,10) MeV
- Test data sample: 10k e + events at each discrete energy points
- No TTS, no dark noise
- Vertex resolution: 5.6cm@1MeV
- Energy resolution: 2.88%



From: Yu Xu. Machine Learning methods for JUNO Experiment

ML and theortical physics

- Pure quantum field theory
- Parton distribution function
- Statistical physics and phase transitions

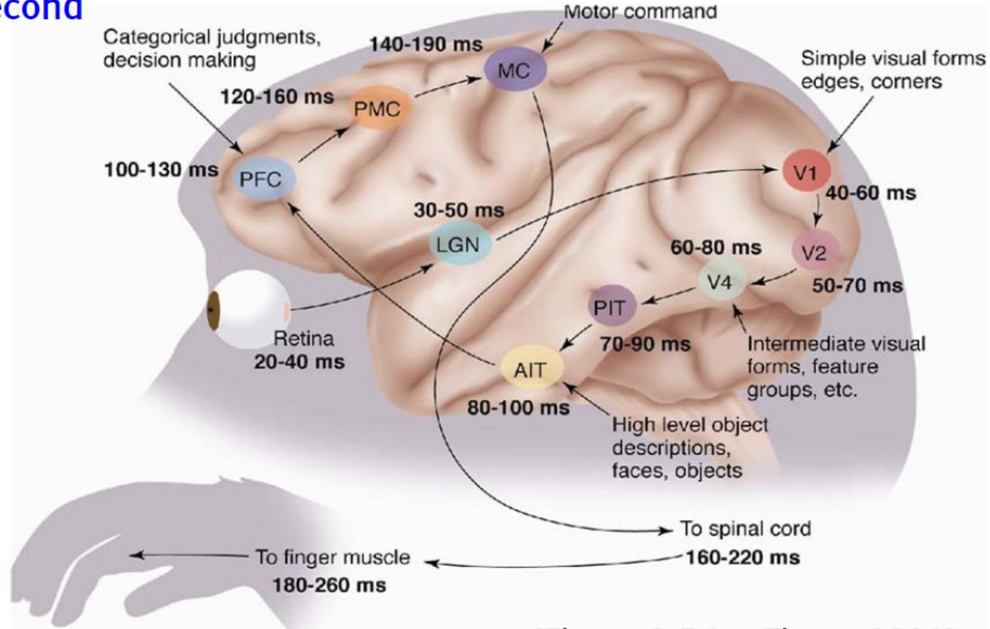


Artificial intelligence
vs
Human brain

The Brain: an Amazingly Efficient "Computer"

Y LeCun

- 10^{11} neurons, approximately
- 10^{4} synapses per neuron
- 10 "spikes" go through each synapse per second on average
- 10^{16} "operations" per second
- 25 Watts
  - Very efficient
- 1.4 kg, 1.7 liters
- 2500 cm2
  - Unfolded cortex

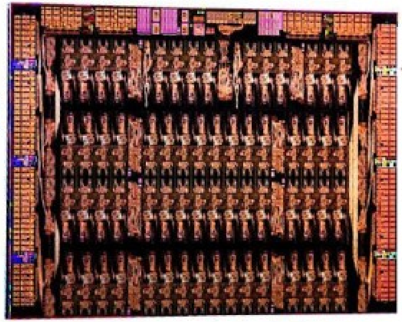


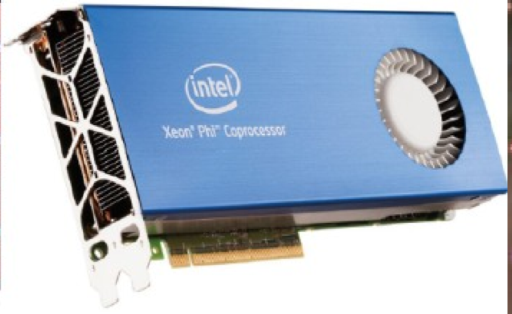
[Thorpe & Fabre-Thorpe 2001]

Fast Processors Today

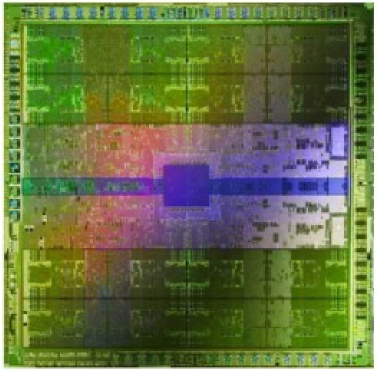
Y LeCun

- **Intel Xeon Phi CPU**
  - 2×10^{12} operations/second
  - 240 Watts
  - 60 (large) cores
  - $3000





- **NVIDIA Titan-Z GPU**
  - 8×10^{12} operations/second
  - 500 Watts
  - 5760 (small) cores
  - $3000



- **Are we only a factor of 10,000 away from the power of the human brain?**
  - Probably more like 1 million: synapses are complicated
  - A factor of 1 million is 30 years of Moore's Law
  - 2045?

Conclusion

- ML is a powerful tool for data analysis in physics.
  - Especially for difficultly formalized tasks.
- For the task of classifying events, deep learning gives a very good result in gamma-ray astronomy.
  - Because of the strong suppression of proton events, neural networks are an important tool for extracting of signal events over background.
- In the problem of restoring the energy spectrum, CNN gives good results.
  - Better result is achieved in the case of simultaneous use of data from several Cherenkov telescopes. The results obtained in this mode are in good agreement with traditional methods based on the Hillas parameters.
- Very good prospects for methods based on generative networks for event simulation as an alternative to Monte Carlo simulation.
  - These methods make it possible to speed up the process of modeling good quality events with correct statistics hundreds and thousands of times.

Acknowledgment

- I want to express my deep gratitude to L. Kuzmichev and V, Prosin for numerous consultations and support.
- I am also grateful to the TAIGA collaboration for providing experimental and Monte Carlo data.
- Many thanks to A. Demichev, S. Polyakov, E. Gres, and A. Vlaskina for providing the results of the analysis of the data from the TAIGA experiment.
- Also to Professor A. Studenikin, without whose support this report would not have taken place.
- The work was partially performed at the UNU “Astrophysical Complex of MSU-ISU” (agreement 13.UNU.21.0007)

References

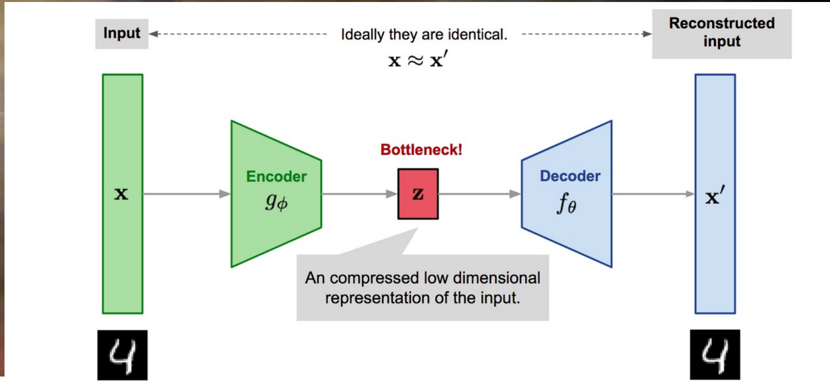
- Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning // Nature, v.521, 28 May 2015
- Ajay Sherestha, and Ausif Mahmood. Review of Deep Learning Algorithms and Architectures // IEEE ACCESS, 2019, 2912200
- Anamika Dhillon, Gyanendra K. Verma. Convolutional neural network: a review of models, methodologies and applications to object detection // Progress in Artificial Intelligence (2020) 9:85–112
- Stefan Funk. Ground- and Space-Based Gamma-Ray Astronomy // Annu. Rev. Nucl. Part. Sci. 2015. 65:245–77
- Mikaël Jacquemont. Cherenkov image analysis with deep multi-task learning from single-telescope data // 257p., https://tel.archives-ouvertes.fr/tel-03590369v2,
- Tobias Fischer. Convolutional Neural Networks for H.E.S.S. // Master’s Thesis, 2018
- Dan Guest, Kyle Cranmer, and Daniel Whiteson. Deep Learning and Its Application to LHC Physics // Annu. Rev. Nucl. Part. Sci. 2018. 68:161–81
- G. Carleo. and et. al. Machine learning and the physical sciences. Rev. Mod. Phys. 91, 045002 (2019)
- And more and more ...


Thank You
Questions

Back slides

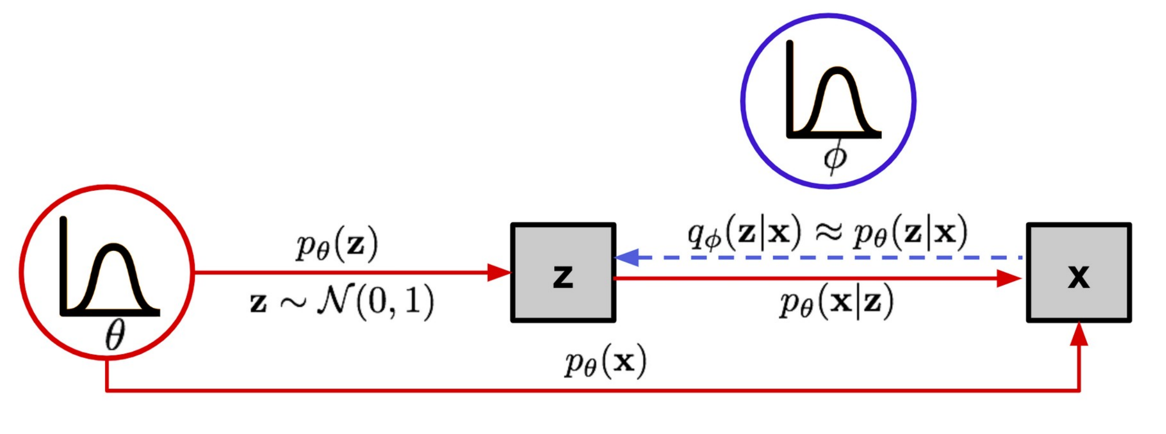
VAE

The model contains an encoder function $g(.)$ parameterized by ϕ and a decoder function $f(.)$ parameterized by θ. The low-dimensional code learned for input x in the bottleneck layer is $z = g_\phi(x)$ and the reconstructed input is $x' = f_\theta(g_\phi(x))$.

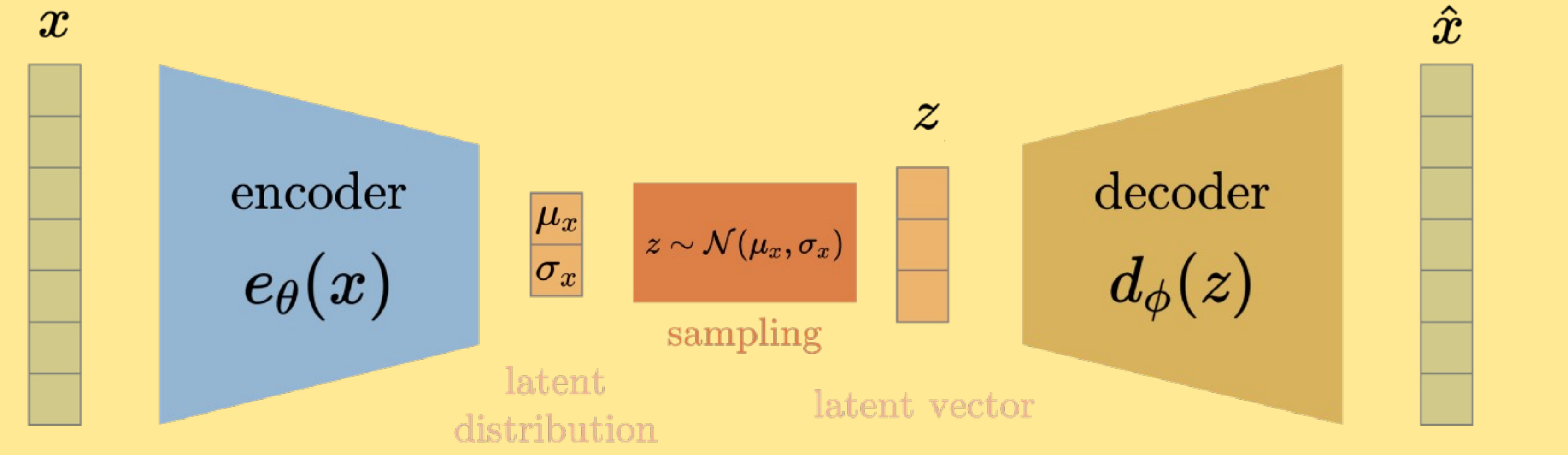


Now the structure looks a lot like an autoencoder:

- The conditional probability $p_\theta(x|z)$ defines a generative model, similar to the decoder $f_\theta(x|z)$ introduced above. $p_\theta(x|z)$ is also known as *probabilistic decoder*.
- The approximation function $q_\phi(z|x)$ is the *probabilistic encoder*, playing a similar role as $g_\phi(z|x)$ above.



VAE



$$\text{reconstruction loss} = \|x - \hat{x}\|_2 = \|x - d_\phi(z)\|_2 = \|x - d_\phi(\mu_x + \sigma_x \epsilon)\|_2$$

$$\mu_x, \sigma_x = e_\theta(x), \quad \epsilon \sim \mathcal{N}(\mathbf{0}, \mathbf{I})$$

$$\textit{similarity loss} = \textit{KL Divergence} = D_{KL}(\mathcal{N}(\mu_x, \sigma_x) \parallel \mathcal{N}(\mathbf{0}, \mathbf{I}))$$

$$\textit{loss} = \textit{reconstruction loss} + \textit{similarity loss}$$