Generation of Artificial Neutral Hadron Events using Cycle-Consistent Neural Networks.

Jack Rolph

University of Hamburg

September 8, 2021

SPONSORED BY THE

Federal Ministry of Education and Research

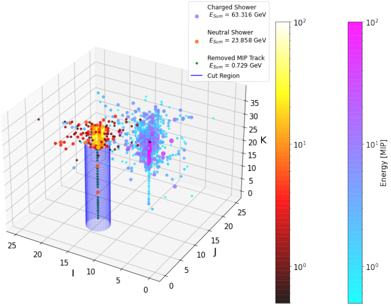
CALICE

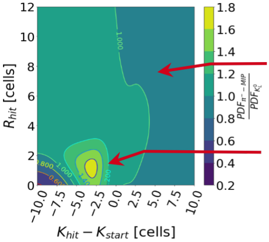
Universität Hamburg

DER FORSCHUNG | DER LEHRE | DER BILDUNG

Motivation.

- The AHCAL has measured **no neutral particle data**, and thus **neutral hadron showers can only be simulated**;
- **Particle Flow clustering algorithms** must be validated using measured data;
- Currently achieved using **selection criteria** to **remove ionising track**:
  - $E_{hit} \geq 3$ MIP
  - $r_{hit} \geq 60$ mm / 2 cells
  - $K_{hit} - K_{start} \geq 0$
- **Ratio of estimated density functions** of K_L^0 and $\pi^- - MIP$ reveal **MIP cut removes too many hits close to shower start**
- **Motivation: shower separation depends on quality of neutrals $\rightarrow$ is it possible to produce a superior alternative?**





After K_{Start}, distributions of K_L^0 to $\pi^- - MIP$ $\sim$ equal.

Before K_{Start}, too many hits removed by $\pi^- - MIP$

Motivation.

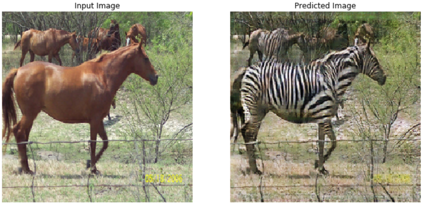
Goals:

- **Train an algorithm** to **perform the conversion** of a **simulated π^- hadron shower** to a **simulated K_L^0 hadron shower**, as observed with AHCAL, using a neural network solution, to produce **artificial** K_L^0 hadron showers;
- **Quantify agreement of method with simulated kaons** compared to CALICE MIP Cut.
- **Then, apply to simulation/data in exactly the same shower separation study shown in Spring 2021 CALICE Meeting.**

Note: link to my last meeting contribution available here.

CycleGAN.

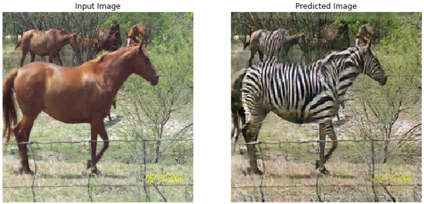
- **CycleGAN** is a special class of neural network designed to **turn data of Domain A into data of Domain B**;
- **Conversion achieved** between **classes of regular, structured images of the same size**: for instance, horse to zebra.
- **Does not require matched examples**, only examples of A and B;
- Powerful tool to examine differences between classes without prior knowledge.



Jun-Yan Zhu et al. *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks.* 2020. arXiv: `1703.10593 [cs.CV]`.

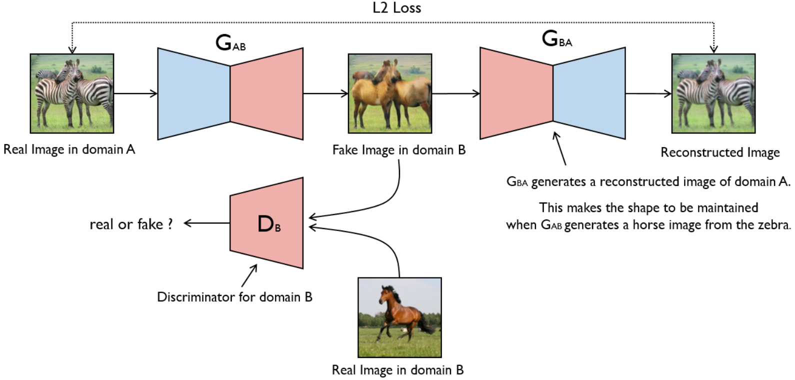
CycleGAN.

- **Hadron shower data** is **much more challenging** than images:
  - **Number of elements not constant;**
  - **Number of dimensions much greater;**
  - **Computationally unfeasible to work with 3D 'hadron shower images';**
  - **No 1:1 mapping between elements.**



Jun-Yan Zhu et al. *Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks.* 2020. arXiv: `1703.10593 [cs.CV]`.

CycleGAN


L2 Loss
G_{AB}
G_{BA}
Real Image in domain A
Fake Image in domain B
Reconstructed Image
D_B
real or fake ?
Discriminator for domain B
Real Image in domain B
G_{BA} generates a reconstructed image of domain A.
This makes the shape to be maintained
when G_{AB} generates a horse image from the zebra.

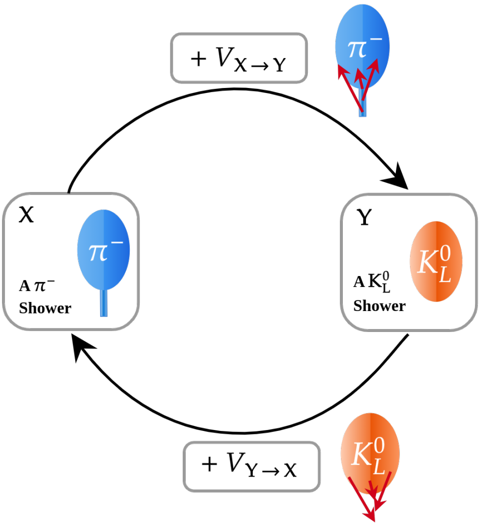
Vector-Modified CycleGAN.

Main Modifications to State-of-the-Art:

- Use a k-**Nearest Neighbours** graph as input $\rightarrow$
  - **reduces dimension** of input data;
  - provides emphasis on **local structure**;
  - **memory efficient.**

- **Output a vector** for **each point** to **change class** $\rightarrow$
  - **measurements cannot be created or destroyed, only displaced;**
  - **shower properties strongly conserved;**
  - **data of differing cardinalities can be compared;**
  - **Output vectors can be studied** $\rightarrow$ **transformation can be studied** on a **case by case basis.**

Vector-Modified CycleGAN



Vector-Modified CycleGAN - An Analogy.

- This solution is built around the **nature's solution** to this problem: **camouflage**.
- Oxford Dictionary camouflage definition:
  - *'the way in which an animal's colour or shape matches what is around or near it and makes it difficult to see'*

Vector-Modified CycleGAN - An Analogy.

- An **animal cannot 'become' its environment**: it has to **change its existing structure** to become **'more like' its environment**.
- With a **vector transformation**,

  the CycleGAN must act similarly:
  - it is **biased not to modify** the **intrinsic properties of the input**;
  - it **allows conversion** between classes of object of **intrinsically different shapes and sizes**.

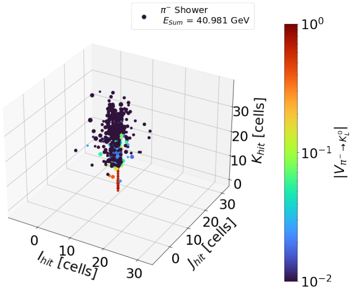
Simulation Information

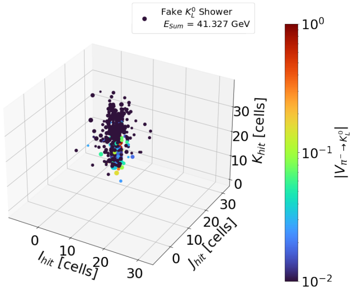
Simulation of π^-/K_L^0 hadronic showers **using Geant4** in the AHCAL were used:

- **full detector simulation** (inc. SiPM saturation/noise thresholds etc.)
- Physics list: **QGSB_BERT_HP**
- Based on **June 2018 CALICE Testbeam** taken at SPS;
- Simulated particle energies: 10, 20, 30, 40, 50, 60, 70, 80 GeV
- Training events: $\sim 2 \times 10^5$ pairs
- Validation events: $\sim 4 \times 10^5$ events $(2 \leq K_{start} < 15)$;
- Model information in backup slides.

Also, more information about model/training in upcoming paper!

Results: Example 1, 40 GeV.

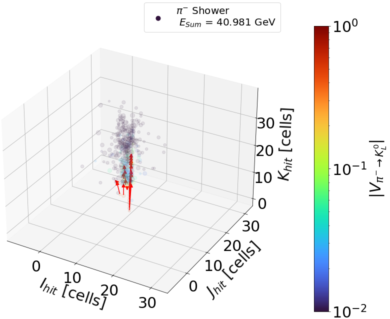




What one learns:

- **Colour axis** describes **length of transformation vector**;
- **Longer vector** $\rightarrow$ **hit less likely to be in domain of** K_L^0;
- **MIP track easily distinguishable** in an event display.

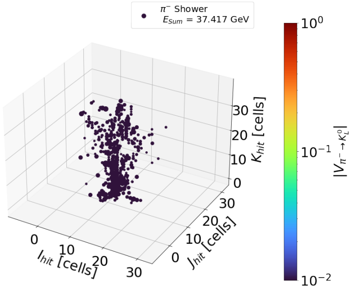
Results: Example 1, 40 GeV.

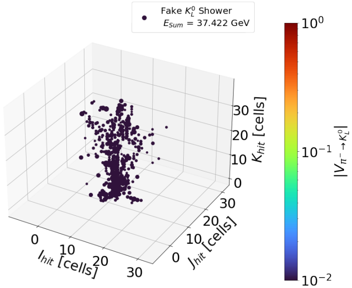


What one learns:

- **MIP hits** are typically **merged** with the **hadron shower core**;
- **Total shower energy** is **re-weighted** to **compensate** for **'missing' MIP track energy**.
- **Important: neural network can allocate energy to an already active cell. This energy is resummed after the conversion has taken place.**

Results: Example 2, 40 GeV.

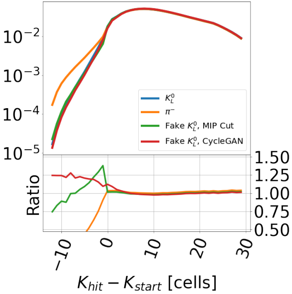


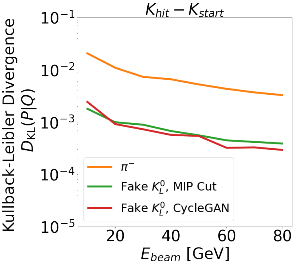


What one learns:

- **CycleGAN** does **nothing** when no **MIP track is present**.

Comparison Plots: $K_{hit} - K_{start}$





What one learns:

- Slightly better agreement overall using CycleGAN method than with CALICE MIP Cut;
- Particularly, improvement in agreement in distributions at the close to starting position.
- However, very close agreement overall.

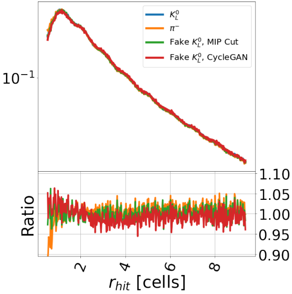
Notes:

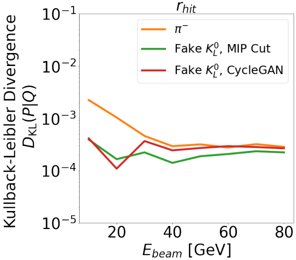
- Top figure is comparison of distributions at incident energy of 40 GeV;
- Kullback-Leibler Divergence measures *'information lost using one probability distribution to describe another'*:

$$D(P|Q) = \sum_{i=0}^{N} P_i \log(\frac{P_i}{Q_i}) \tag{1}$$

- Binning is optimised using Knuth's Rule where appropriate.
- Range chosen: left-tailed 95% confidence interval.

Comparison Plots: r_{hit} Distribution





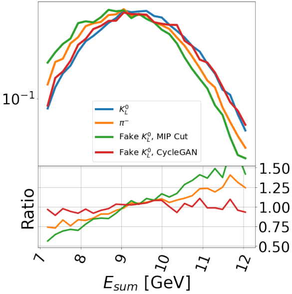
What one learns:

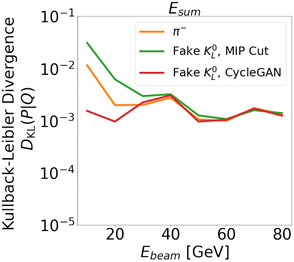
- Slightly worse agreement overall using CycleGAN method than with CALICE MIP Cut;
- However, very close overall.

Notes:

- Top figure is comparison of distributions at incident energy of 40 GeV;
- Range chosen: central 90% confidence interval. item Else, as in previous slide;

Comparison Plots: E_{sum} Distribution





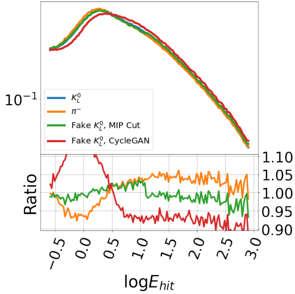
What one learns:

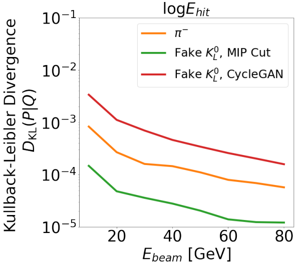
- **Main improvement: machine learning method capable of re-weighting shower energy;**
- **This energy was observed to be 'added back' to the electromagnetic core; energy is lost when removed from the hadron shower.**
-
- **Very significant effect** for **'low' energy** particles, **less so** for **'high' energy particles;**

Notes:

- Top figure is comparison of distributions at incident energy of 10 GeV;
- Else, as in previous slide;

Comparison Plots: $\log E_{hit}$ Distribution





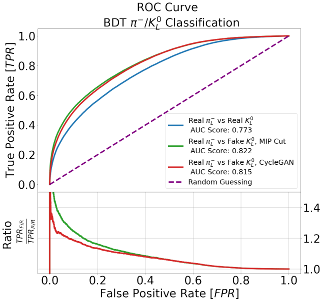
What one learns:

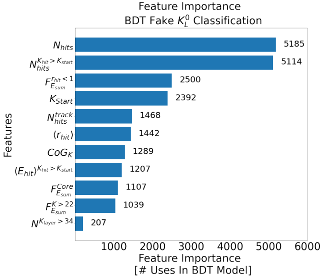
- **Main flaw: the hit energy distribution is systematically shifted to higher energy.**
- **Agreement** in this case **is poor**, since the **total shower energy cannot be increased without** either:
  - **increasing average hit energy**;
  - **increasing the number of hits** by re-sampling;
- Likely resolvable with more sophisticated traditional model zoo: i.e. CycleGAN, TravelGAN + memory.

Notes:

- As in previous slide;

Comparison Plots: PID performance.



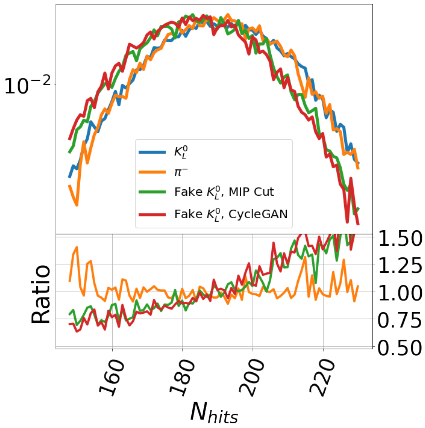


- **Vladimir's PID used to validate further;**
- **Boosted Decision Tree** first trained to **distinguish simulated** π^- **and** K_L^0 **hadron showers**, from half validation sample;
- **ROC curve** was measured **w.r.t remaining validation sample**, using simulated, MIP Cut, and CycleGAN, artificial K_L^0;
- **Hypothesis**: the **more similar the ROC curve** for **artificial** K_L^0/π^- **classification** is to the **ROC curve** for **simulated** K_L^0/π^- **classification**, the **more convincing** the **artificial** K_L^0 **showers** must be.

What one learns:

- **Both MIP Cut and CycleGAN** produce **similar ROC Curves;**
- **The methods** produce **more easily classifiable** K_L^0 **hadron showers** than in **simulation;**
- Reason: **both methods** systematically **reduce** the **number of active cells** in a π^- shower **too greatly**.

Comparison Plots: N_{hits}



- **systematically lower numbers of active cells** in **10 GeV artificial K_L^0 hadron showers** than for **simulated π^- hadron showers**;
- π^- **and** K_L^0 have **similar numbers of active cells** during an event - **removing these elements/allocating energy back to the hadron shower** results in **artificial events with too few active hits**;
- **Only solution is re-sampling event**:
  - **Theoretically possible** with **ordinary CycleGAN**;
  - **Practically**, requires **greater computing resources** and **strong constraints on model output** (i.e. to enforce shower start position);

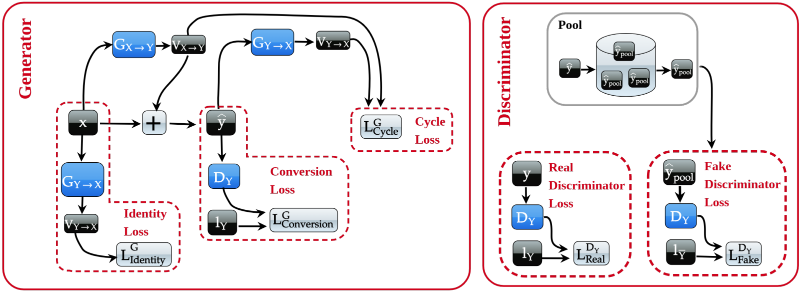
Conclusion and Outlook.

- **A method** for determining a **charged-neutral conversion** has been deduced using **modified CycleGAN**:
- Is this **better than CALICE MIP Cut** at **producing artificial neutrals in simulation**?
  - **Pros: total reconstructed energy re-weighted** to account for missing MIP track and produces a more consistent $K_{hit} - K_{start}$ distribution.
  - **Cons: hit energy distribution** must **shift to higher energies**;
  - **PID validation** shows the methods produce **similarly distinguishable** K_L^0 **hadron showers**;
  - Reason: **both methods reduce the overall number of active cells too much** $\rightarrow$ **re-sampling is therefore mandatory.**
  - **Long-term Solution**: a **traditional cycle-consistent method**, when the **resources become available to train one**.
- Next steps: use CycleGAN neutrals in shower separation study.

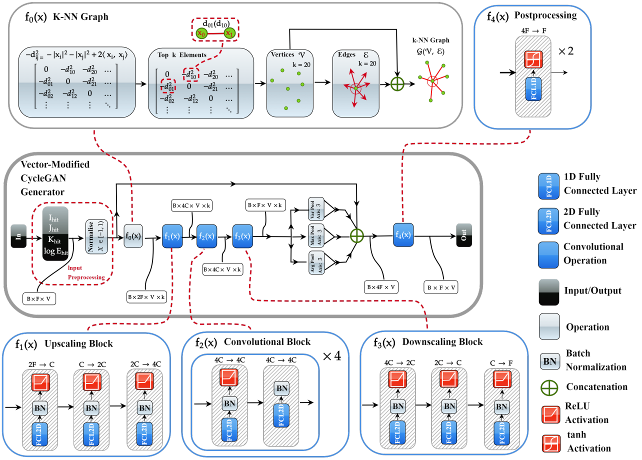
> Backup

Example Training Cycle

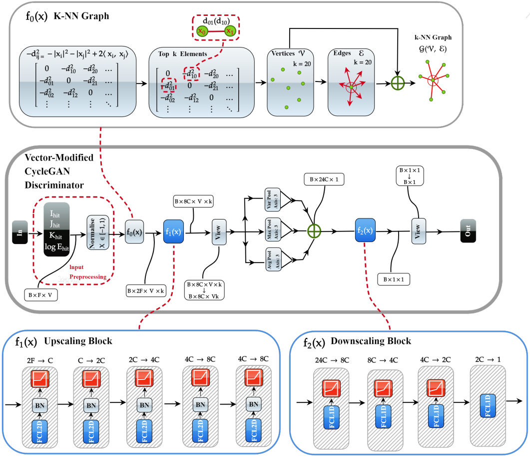
$X \rightarrow Y$



Generator Network



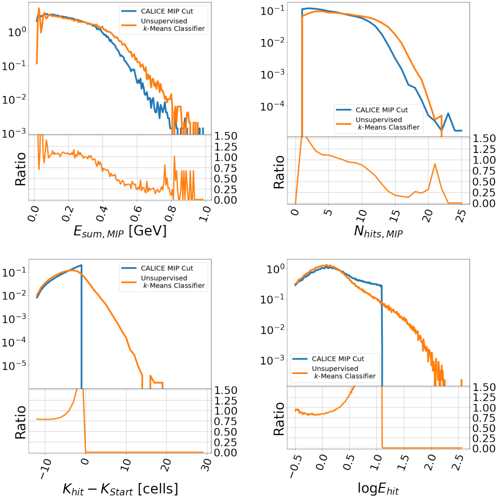
Discriminator Network.



Unsupervised Classification.

- **Suppose we knew nothing** about the **physics of hadron showers.**
- Using **only this model** and **example measurements of** π^- **and** K_L^0 **hadron showers**, how well could we **determine a MIP selection criteria**?
- i.e. **can we used this method for inventing/informing unsupervised selection cuts**?
- For instance: determining background/detector noise, systematic differences between data/simulation etc.

Unsupervised Classification.



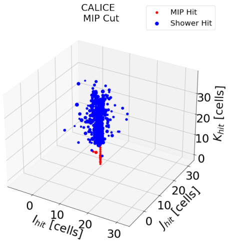
- An **unsupervised** k-means **clustering algorithm** was optimised to **cluster hits into two groups, using the vector elements as input**.
- The **only prior** is that there are **two groups**: K_L^0**-like hits (unmodified)** and **'other' hits (modified)**
- **No other information is provided.**
- **No ground truth available**: the **behaviour** of the **resulting classifier** was **compared** to the **existing CALICE MIP Cut**.

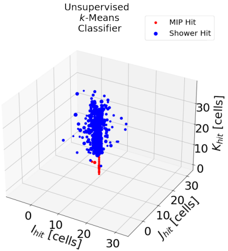
What one learns:

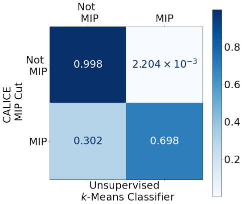
- **Distributions of the selected hits** suggest that the unsupervised classifier provides **a selection criteria** that classifies hits with similar properties, and of a similar number and total energy event-by-event.
- **Deviations** are however, **observed** at the

  tails of each distribution, where:
  - $N_{hits,MIP} > 10$;
  - $E_{sum,MIP} > 0.4$ GeV;
  - $K_{hit} > K_{start}$
  - $K_{start} > 1.6$ MIP;

Unsupervised Classification.





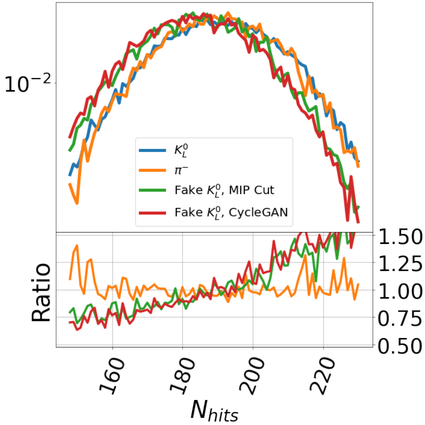


- An **unsupervised** k-means **clustering algorithm** was optimised to **cluster hits into two groups, using the vector elements as input**, using half the validation set, and tested on the other half.
- The **only prior** is that there are **two groups**: K_L^0**-like hits (unmodified)** and **'other' hits (modified)**
- **No other information is provided.**
- **No ground truth available**: the **behaviour** of the **resulting classifier** was **compared** to the **existing CALICE MIP Cut**.

What one learns:

- **70% of the same hits** are **classified as MIP** by the **unsupervised classifier** as the **standard cut**.

Comparison Plots: N_{hits}



- Example: **systematically lower numbers of active cells** in **10 GeV artificial K_L^0 hadron showers** than for **simulated π^- hadron showers**.