Unpaired Image-to-Image Translation with Cycle-Consistent GANs for Bridging Data-Simulation Discrepancies

Dmitrii Torbunov,
Brookhaven National Laboratory

August 7, 2022

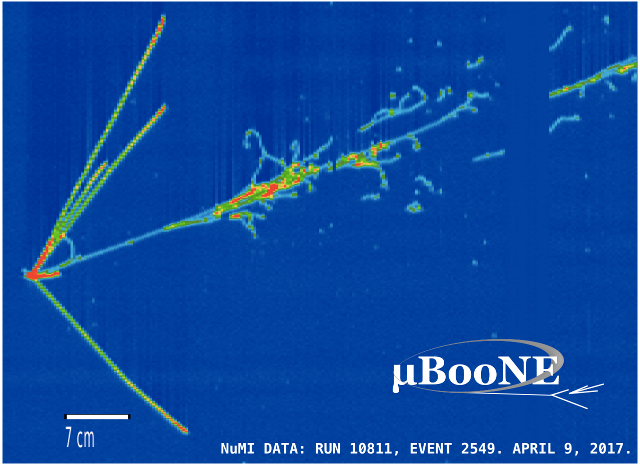
Introduction

- We attempt to reduce Data vs Simulation discrepancies in the High Energy Physics (HEP) with a help of unpaired image-to-image translation methods.

- Plan of the Talk:
  - Motivate the Problem
  - Discuss our Experiment and Dataset
  - Discuss our Method
  - Highlight some challenges that we faced

Motivation – No Labeled Data

- In High Energy Physics (HEP) experiments we study elementary particles as they pass through the detector.
- To refine our understanding of the laws of nature, we try to identify particles in the detector and estimate their energies.
- However, humans can perform such an identification only in a few special cases.
- Problem - where to get labels from in the general case?



Credits MicroBooNE Collaboration

Labeled Data. Standard Solution

- Since humans cannot perform data labeling, physicist are relying Machine Learning algorithms to perform the particle identification/energy estimation.

- However, the problem still remains – to train the ML algorithms one needs the labeled data.

- Solution – use Simulation:
    1. Perform simulation of particle interactions with the detector.
    2. Train ML algorithms on the simulation.
    3. Use ML algorithms on the real data.

- The usage of simulation solves the problem of getting the labeled data.
- Unfortunately, there are multiple **discrepancies present between the simulated and the real data**.
- These discrepancies raise serious concerns, regarding the validity of the usage of the ML algorithms on the real data, given that they are trained on the simulation.

Reality-Simulation Discrepancies, Solutions?

- Physicists spent a considerable amount of time trying to manually identify all possible discrepancies between the simulation and the real data.
- Yet, new discrepancies are still found continuously. They significantly limit the applicability of the ML methods in the HEP community and cause significant delays in data analysis.
- **We are trying to leverage the power of Deep Learning to automatically identify possible discrepancies between the simulation and the real data, and fix them**.

Bridging the Simulation-Reality Gap with Deep Learning

- We formulate the problem of reducing the simulation-reality discrepancies as the domain translation problem:
  1. **Domain A** – Simulated Data.
  2. **Domain B** – Real Data.

- Given an image X_A from the Simulation Domain we would like to translate it to the Domain of Real Data with a help of a Deep Network $\mathcal{G}_{A\to B}$, such that the translated image $\mathcal{G}_{A\to B}(X_A)$ is indistinguishable from the Real Data.

- Such a problem is known in the Deep Learning community as an **Image-to-Image translation** problem and has multiple solutions.

- In particular, since there is no simple way to pair Simulated to Real images, our problem maps to the **Unpaired Image-to-Image translation** problem in the Deep Learning field.

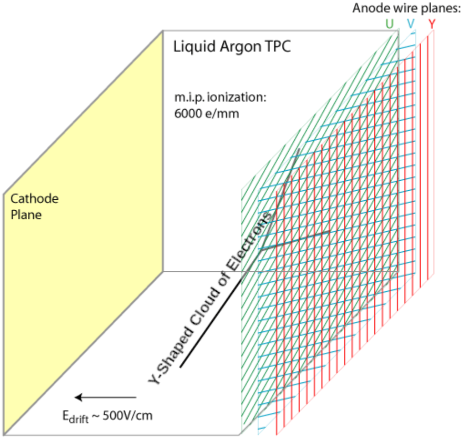
Milestone

- Plan of the Talk:
  - Motivate the Problem
  - **Discuss our Experiment and Dataset**
  - Discuss our Method
  - Highlight some challenges that we faced

Dataset

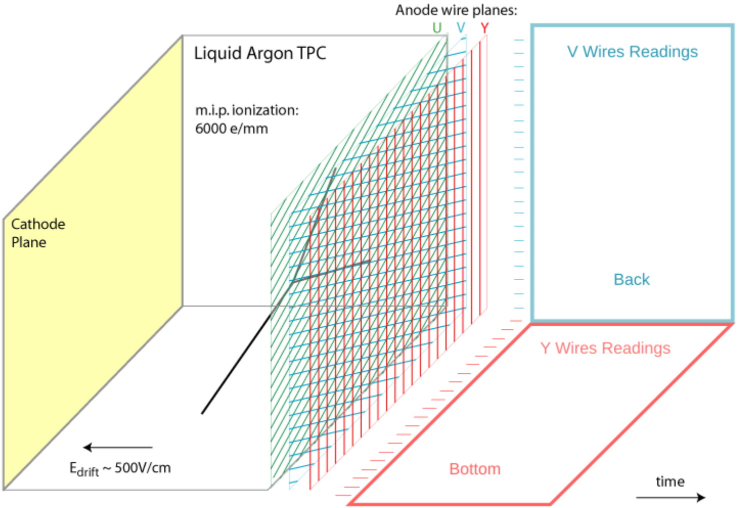
- Before attempting to solve the domain translation problem we need to agree on the dataset to use.
- As a starting point, we decided to consider **L**iquid **Ar**gon **T**ime **P**rojection **C**hamber (LArTPC) based experiments.
- The LArTPCs are the next generation particle detectors:
  - The LArTPC detector is simply a container filled with liquid Argon.
  - When charged particles travel though the liquid Argon they create trails of electrons.
  - These trails can be used to detect particles.

Liquid Argon Detector, 1

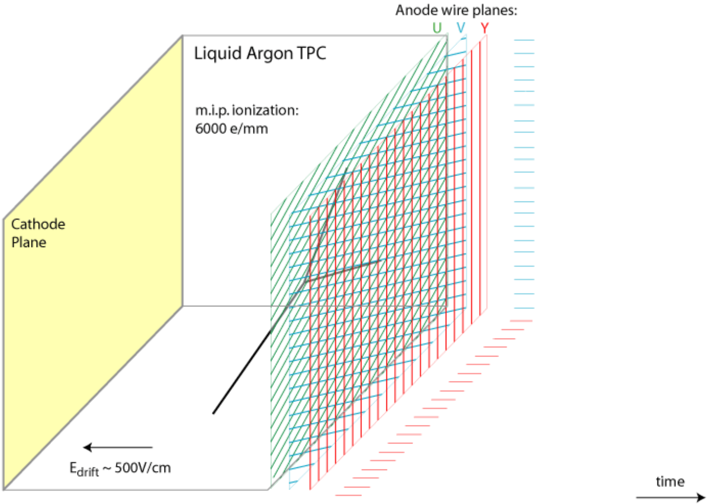


- Charged particles leave clouds of electrons in the detector.
- The clouds of electrons slowly drift towards the wire-planes under the influence of the strong electric field E_{drift}.
- Metallic wires record electric excitations caused by the bypassing electrons.

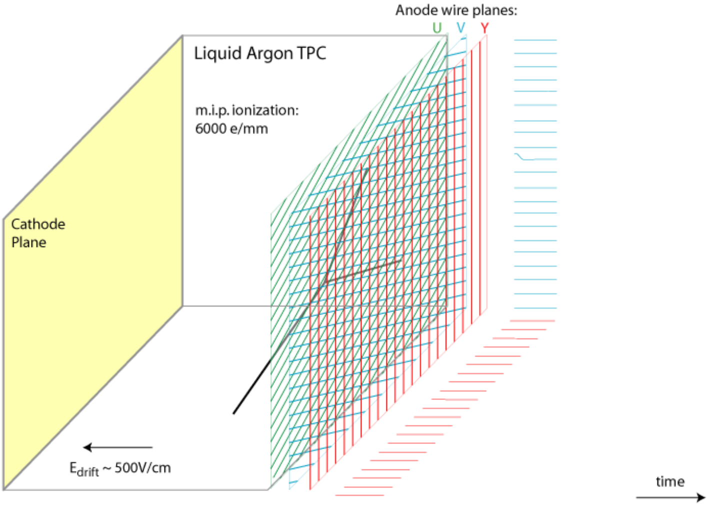
Liquid Argon Detector, 2



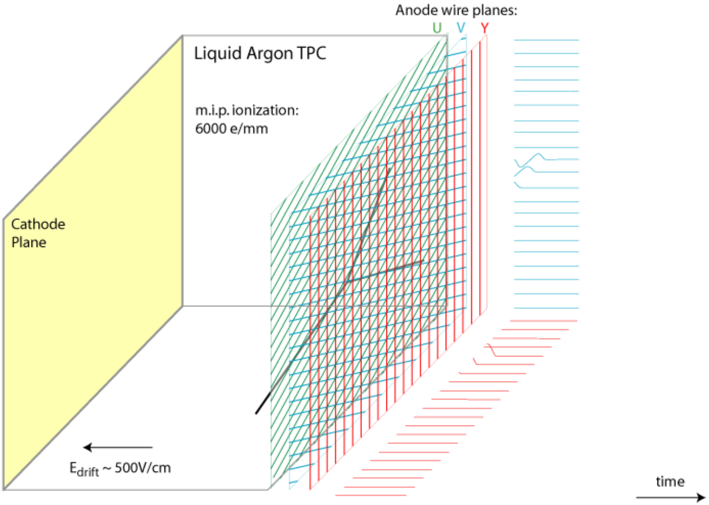
Liquid Argon Detector, 3



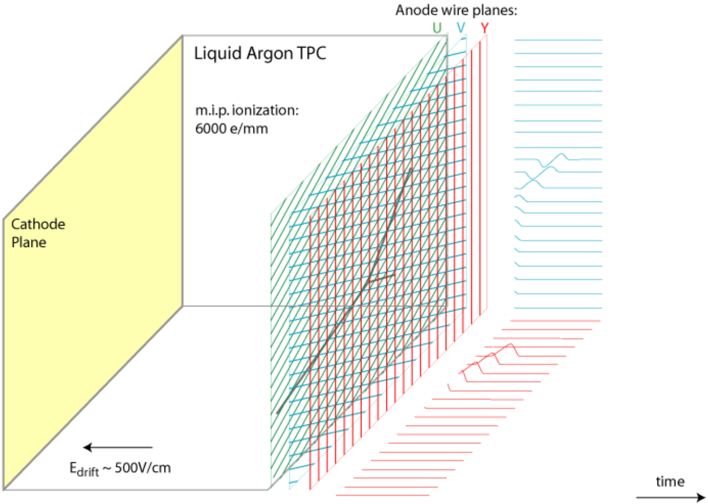
Liquid Argon Detector, 4



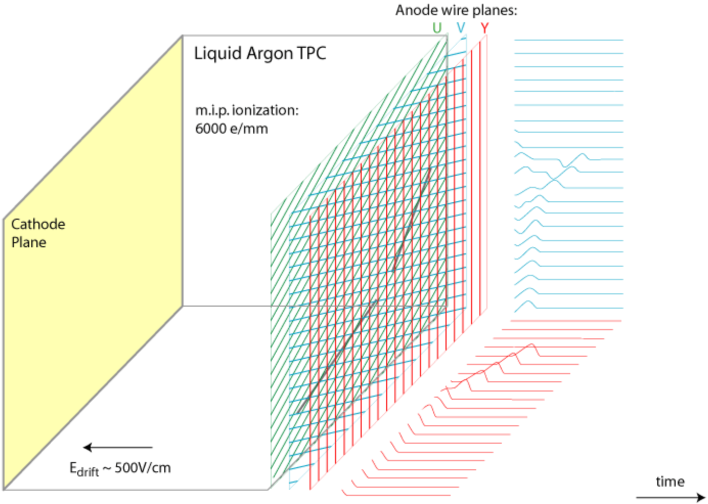
Liquid Argon Detector, 5



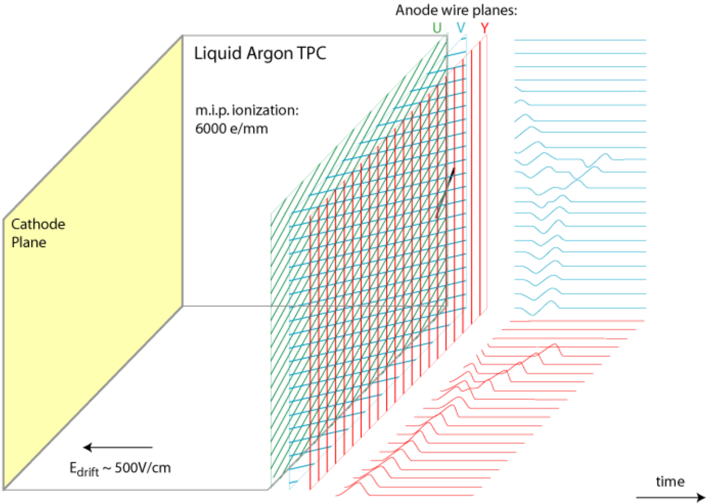
Liquid Argon Detector, 6



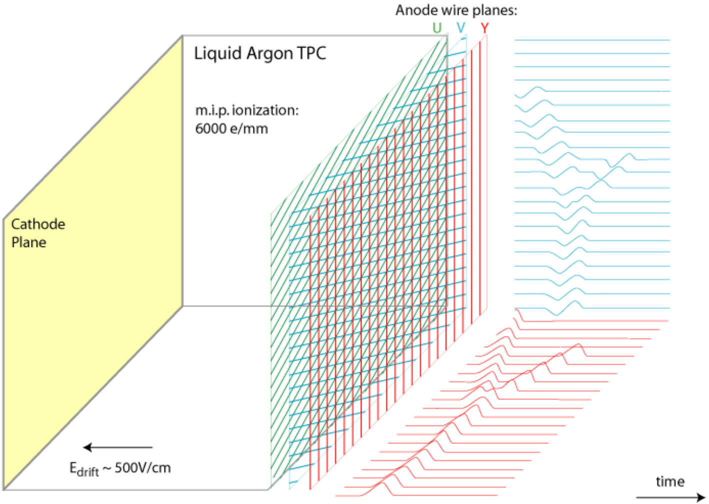
Liquid Argon Detector, 7



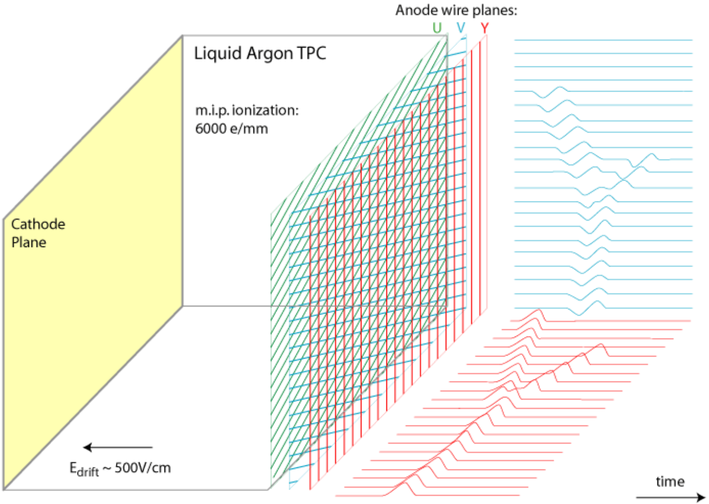
Liquid Argon Detector, 8



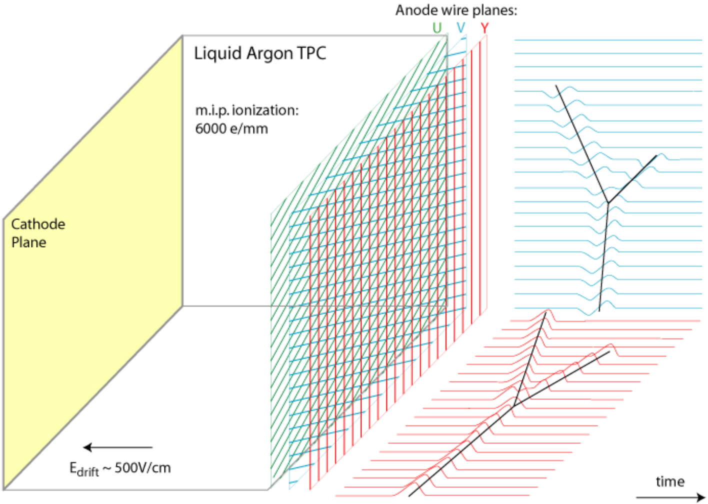
Liquid Argon Detector, 9



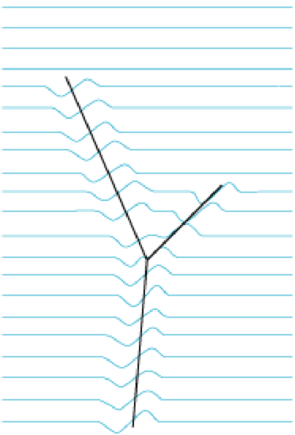
Liquid Argon Detector, 10



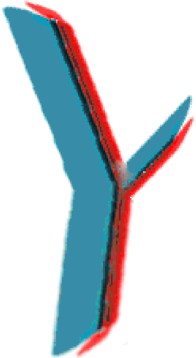
Liquid Argon Detector, 11



Liquid Argon Detector, 12



(a) Schematic Waveforms



(b) Actual Data Sample

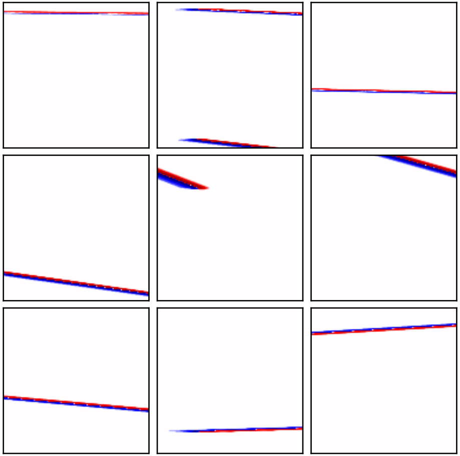
Benchmark Data

- Before tackling the **Simulation** to **Real Data** translation we focused on a simpler problem of translating **Simulation A** into **Simulation B**.

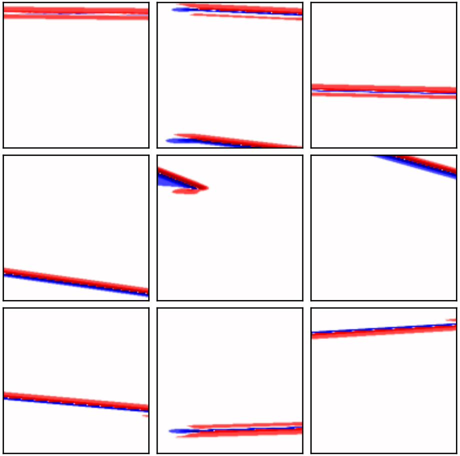
- Both simulated datasets use the same simulation of particle propagation through the detector.

- The only difference between the two simulated datasets is the way the wires respond to the clouds of electrons:
    1. **Domain A** – simplified detector response, where a cloud of electrons is read only by the nearest wire.
    2. **Domain B** – realistic detector response, where a cloud of electrons can produce excitations in multiple wires.

Sample of Generated Images



(a) **Domain A**



(b) **Domain B**

Milestone

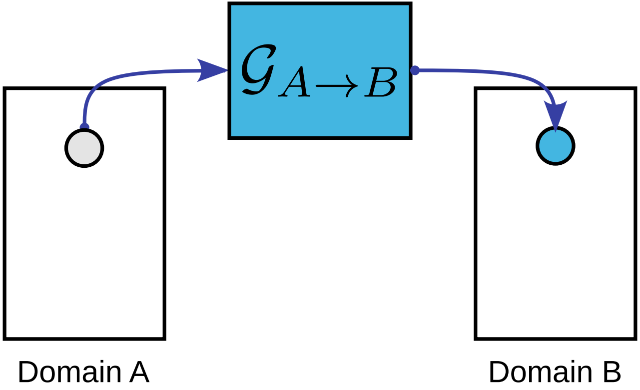
- Plan of the Talk:
  - Motivate the Problem
  - Discuss our Experiment and Dataset
  - **Discuss our Method**
  - Highlight some challenges that we faced

Unpaired Image-to-Image Translation

- Once the benchmark dataset has been chosen we can start exploring image translation methods.
- The most common image-to-image translation methods rely on Generative Adversarial Networks.
- Of the possible GAN image-to-image translation variants there is one model that has drawn our attention due to its inherent properties – the CycleGAN model [1].

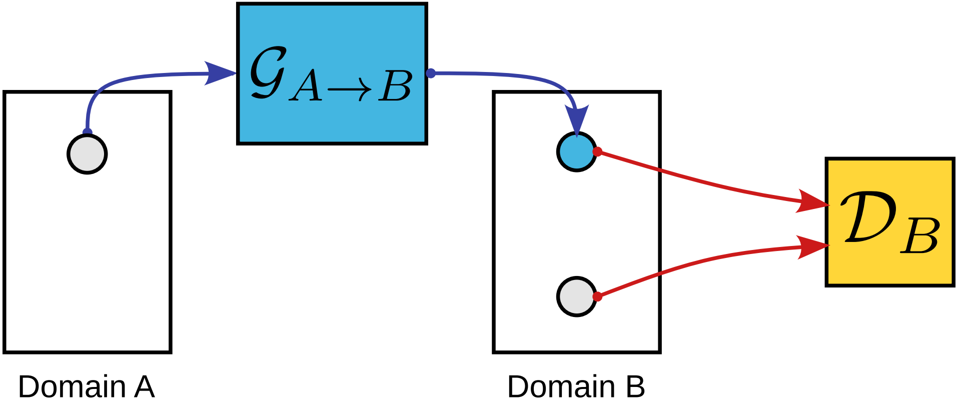
[1] `arXiv:1703.10593`

CycleGAN Basics



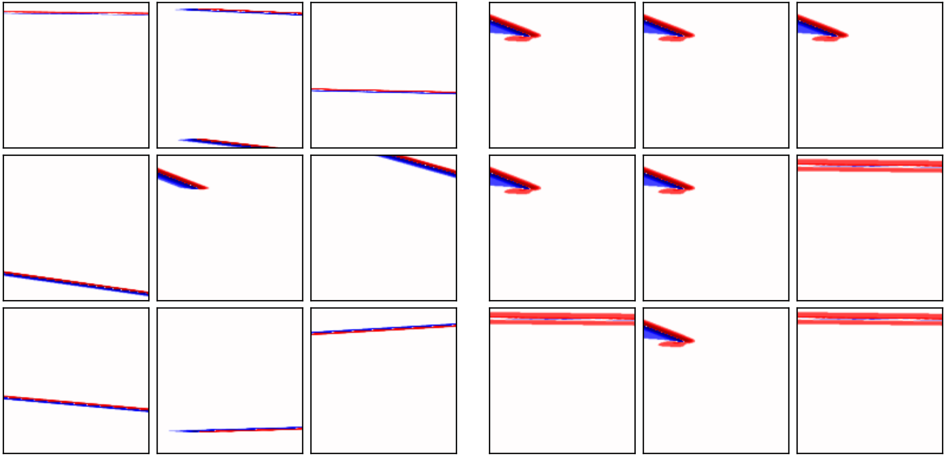
At the core of CycleGAN lies a traditional GAN architecture with a generator $\mathcal{G}_{A\to B}$ that translated images from domain A to B.

CycleGAN Basics



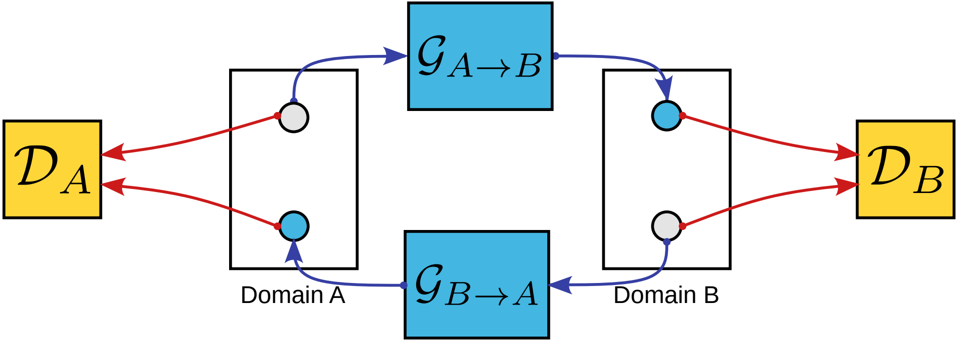
To train the generator $\mathcal{G}_{A\to B}$, GAN uses an adversarial min-max game against a discriminator $\mathcal{D}_B$

Traditional GAN training is prone to Mode Collapse



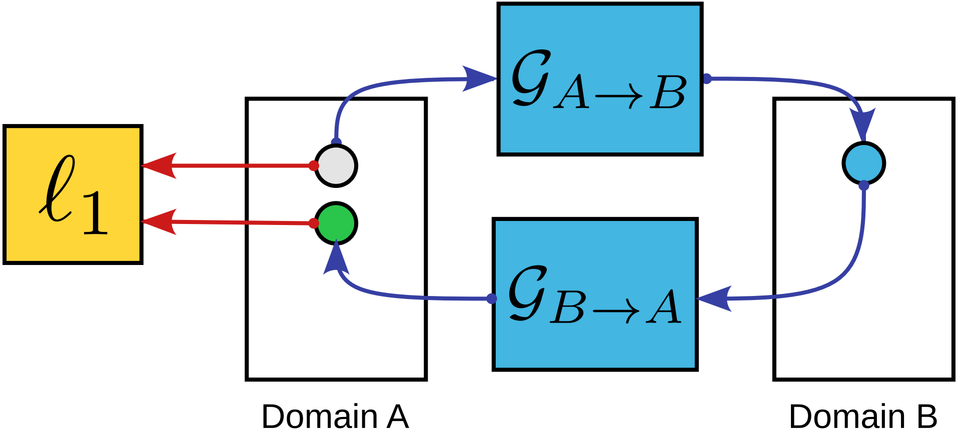
(a) **Domain A** (b) **Translated to Domain B**

CycleGAN Architecture



CycleGAN uses two generators $\mathcal{G}_{A\to B}$, $\mathcal{G}_{B\to A}$ and two discriminators $\mathcal{D}_A$ $\mathcal{D}_B$.

CycleGAN Cycle Constraint

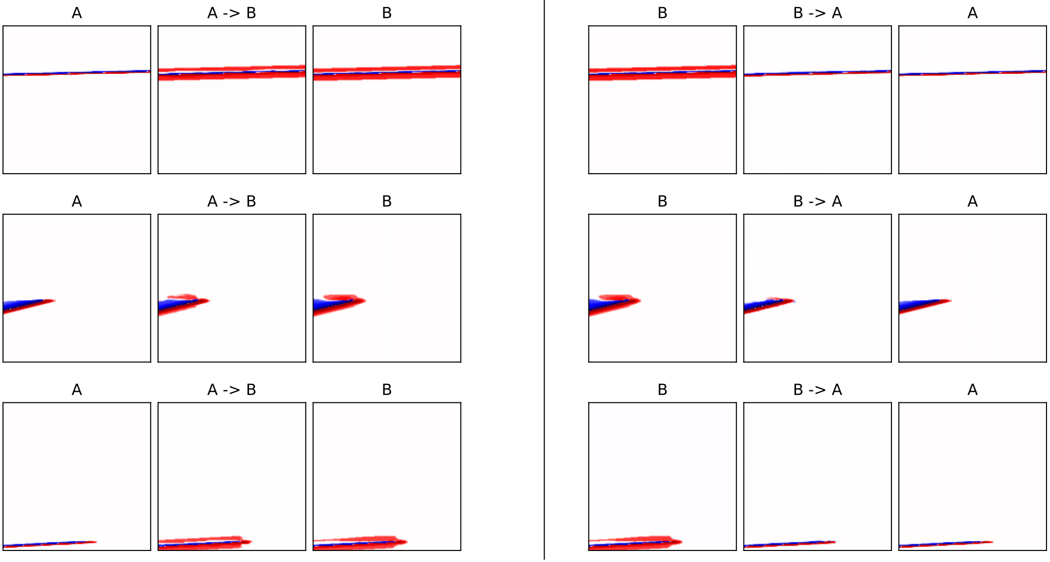


To solve the Mode Collapse problem, CycleGAN enforces a cycle-consistency constraint: $\mathcal{G}_{B\to A}(\mathcal{G}_{A\to B}(X_A)) == X_A$, and symmetrically in for X_B.

CycleGAN Solution of the Mode Collapse Problem

- **CycleGAN** is one of the earliest successful Unpaired Image-to-Image transfer architectures. It solves the mode collapse problem by imposing the **cycle-consisntency constraint**.
- **Better models** exist in terms of the image translation quality. However, these models **relax the cycle-consistency** constraint, allowing generator to gain and loose information during the translation.
- We believe that maintaining the cycle-consistency is important goal, since we do not want the physics to be altered going from domain A to B and back.
- So, we have attempted to use the default CycleGAN architecture on our benchmarking dataset.

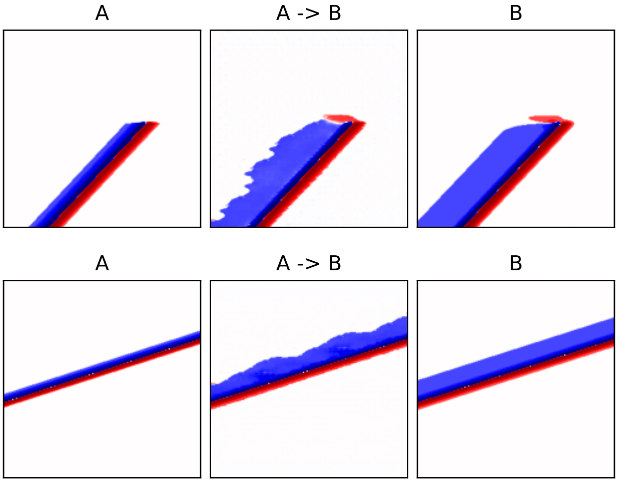
Initial CycleGAN Usage Attempt



Milestone

- Plan of the Talk:
  - Motivate the Problem
  - Discuss our Experiment and Dataset
  - Discuss our Method
  - **Highlight some challenges that we faced**

Initial CycleGAN Usage Attempt – Chunky Tracks

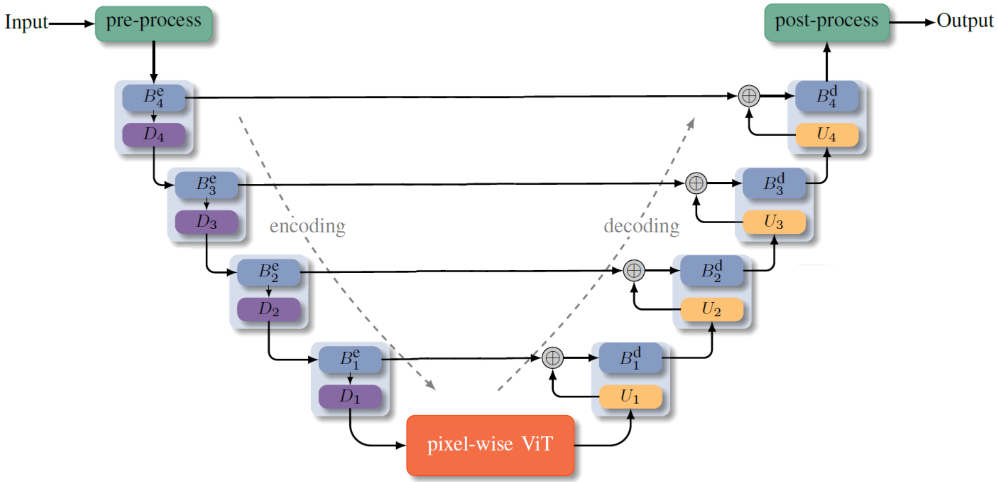


The default CycleGAN experiences difficulties with a translation of wide tracks

CycleGAN Translation Issues

- The default CycleGAN gives good translation on average, yet produces noticeable artifacts for some cases.
- Analysing the artifacts, we believe that they are the result of the generator architectures of the CycleGAN (UNet or ResNet).
- The CycleGAN generators rely on CNN layers, which are very local in nature. So, perhaps the network simply translated image patches locally, and they are globally mismatched.
- We tried to augment the CNN generator to add a global matching step.

UNet-ViT Generator



We augmented UNet by a Vision Transformer bottleneck to handle long-range dependencies.

UNet-ViT Fixing Bad Translations

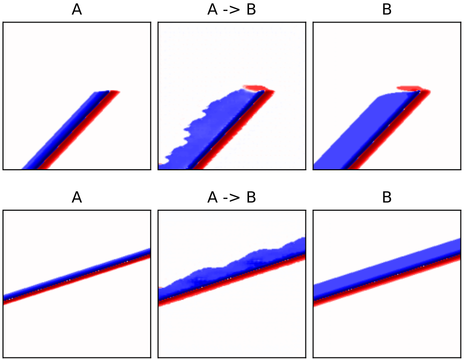


Figure: Default CycleGAN Generator

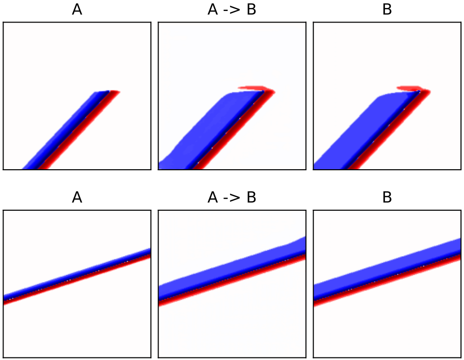


Figure: New UNet-ViT Generator

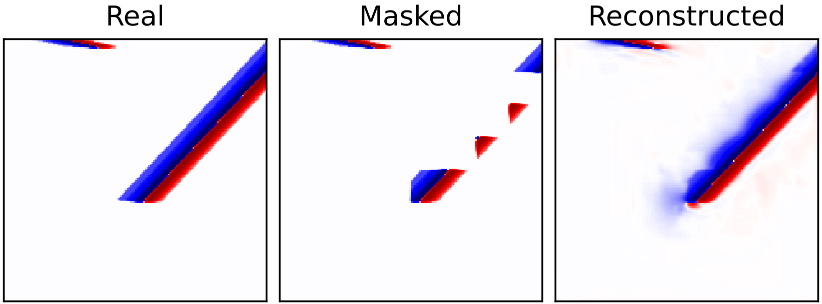
Improving CycleGAN Performance

- By modifying the generator architecture we were able to improve the translation quality of CycleGAN.
- However, the vanilla CycleGAN is still outperformed by the more advanced models that relax the cycle-consistency constraint.
- We have tried to further improve the CycleGAN performance to match the one of the more advanced algorithms.

CycleGAN Modifications, 1

- To further improve the CycleGAN performance we have tried to pre-train its generators.
- To pre-train the generators we focused on a task of a **self-supervised image inpainting**.

Pre-Training Generators by Image Inpainting



For the image inpainting task, some parts of the image are masked. The network is tasked with recovering the masked parts of the image from their surroundings

CycleGAN Modifications, 2

- Unfortunately, the pretraining alone did not seem to improve the image translation quality.
- So, we have implemented a discriminator regularization by penalizing the magnitude of its gradients.
- Together, the **self-supervised pre-training** and **gradient penalty** allowed the CycleGAN to achieve the performance of more advanced algorithms.

Modified CycleGAN Benchmarking

- We have compared our modified CycleGAN (called **UVCGAN**) vs more advanced models:
  1. **ACL-GAN** `arXiv:2003.04858`
  2. **Council-GAN** `arXiv:1911.10538`
  3. **U-GAT-IT** `arXiv:1907.10830`
- We performed the comparison on the standard Image-to-Image benchmark datasets (Selfie-to-Anime, Male-to-Female, Glasses Removal).
- Our **UVCGAN** outperforms more advanced models on 4 out of 6 benchmarks and achieves competitive results on the other two.

Table 2. FID and KID scores. Lower is better.

	Selfie to Anime		Anime to Selfie	
	FID	KID (×100)	FID	KID (×100)
ACL-GAN	99.3	3.22 ± 0.26	128.6	3.49 ± 0.33
Council-GAN	91.9	2.74 ± 0.26	126.0	2.57 ± 0.32
CycleGAN	93.4	2.96 ± 0.27	129.4	2.91 ± 0.39
U-GAT-IT	95.8	2.74 ± 0.31	**108.8**	$\mathbf{1.48 \pm 0.34}$
UVCGAN	**79.0**	$\mathbf{1.35 \pm 0.20}$	122.8	2.33 ± 0.38
	Male to Female		Female to Male	
	FID	KID (×100)	FID	KID (×100)
ACL-GAN	**9.4**	$\mathbf{0.58 \pm 0.06}$	19.1	1.38 ± 0.09
Council-GAN	10.4	0.74 ± 0.08	24.1	1.79 ± 0.10
CycleGAN	15.2	1.29 ± 0.11	22.2	1.74 ± 0.11
U-GAT-IT	12.6	0.88 ± 0.08	23.1	1.91 ± 0.12
UVCGAN	9.6	0.68 ± 0.07	**13.9**	$\mathbf{0.91 \pm 0.08}$
	Remove Glasses		Add Glasses	
	FID	KID (×100)	FID	KID (×100)
ACL-GAN	16.7	0.70 ± 0.06	26.6	2.26 ± 0.17
Council-GAN	37.2	3.67 ± 0.22	19.5	1.33 ± 0.13
CycleGAN	24.2	1.87 ± 0.17	19.8	1.36 ± 0.12
U-GAT-IT	20.9	1.39 ± 0.13	20.0	1.16 ± 0.09
UVCGAN	**14.4**	$\mathbf{0.68 \pm 0.10}$	**13.6**	$\mathbf{0.60 \pm 0.08}$

UVCGAN Intermediate Results

- By modifying the generator architecture and improving the training procedure, we were able to make the CycleGAN performance competitive with more advanced models.
- We have also obtained reasonably good translation quality on our benchmarking LArTCP dataset.
- Generic Image-to-Image translation results are published in `arXiv:2203.02557`[2], and LArTCP results are in preparation.

[2] Code can be found here `https://github.com/LS4GAN/uvcgan`

UVCGAN Next Steps

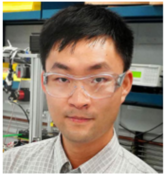
- Now our work focused on three directions:
  1. Simulate a more advanced LArTPC dataset.
  2. Extend UVCGAN to work on much larger images (6000×1000).
  3. Explore other improvements to the UVCGAN training.

Our Team

Dmitrii Turbunov

Yi Huang

Haiwang Yu

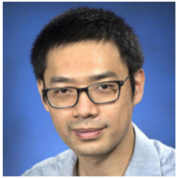


Jin Huang

Shinjae Yoo

Meifeng Lin

Brett Viren



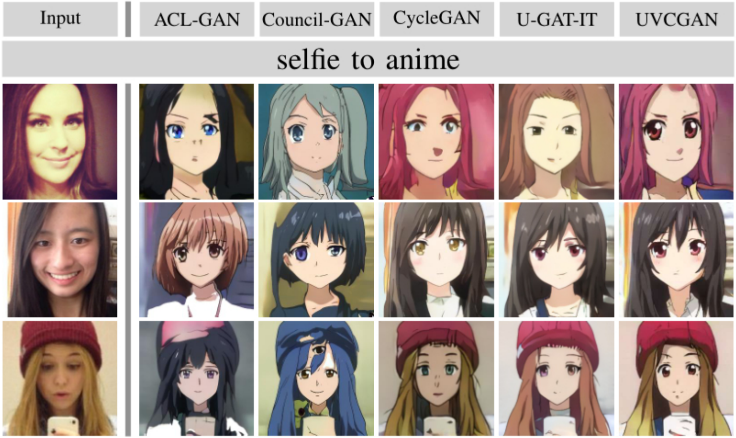
Yihui Ren

Backups

Natural Image Translation Quality, FID and KID

- For natural images, well-established metrics are avaialble to assess the quality of translation: **FID** and **KID**.
- These metrics compare translated images $X_{A\to B} \equiv \mathcal{G}_{A\to B}(X_A)$ to the real images in the **Domain B** X_B.
- Evaluation of all metrics begins by applying Inception Net classifier, trained on the ImageNet dataset over all images $\{X_{A\to B}\}$ and $\{X_B\}$.
- For each image, a number of features is extracted from a selected set of deep layers of the Inception Net.
- **FID** and **KID** scores compare the resulting sets of features between the distribution of translated and real images.

UVCGAN Cycle-Constraint Effects



Unlike other models, UVCGAN maintains strong correlation between input and output