

Unsupervised Domain Transfer for Science: Exploring Deep Learning Methods for Translation between LArTPC Detector Simulations with Differing Response Models

Yi Huang, Dmitrii Torbunov, Brett Viren, Haiwang Yu, Jin
Huang, Meifeng Lin, Yihui Ren

Brookhaven National Laboratory, Upton, NY, USA

{dtorbunov,yhuang2,htseng,hyu,jhuang,sjyoo,mclin,bviren,yren}@bnl.gov

Abstract. Deep learning (DL) techniques have broad applications in science, especially in seeking to streamline the pathway to potential solutions and discoveries. Frequently, however, DL models are trained on the results of simulation yet applied to real experimental data. As such, any systematic differences between the simulated and real data may degrade the model’s performance—an effect known as “domain shift.” This work studies a toy model of the systematic differences between simulated and real data. It presents a fully unsupervised, task-agnostic method to reduce differences between two systematically different samples. The method is based on the recent advances in unpaired image-to-image translation techniques and is validated on two sets of samples of simulated Liquid Argon Time Projection Chamber (LArTPC) detector events, created to illustrate common systematic differences between the simulated and real data in a controlled way. LArTPC-based detectors represent the next-generation particle detectors, producing unique high-resolution particle track data. This work open-sources the generated LArTPC data set, called Simple Liquid-Argon Track Samples (or SLATS), allowing researchers from diverse domains to study the LArTPC-like data for the first time.

1. Introduction

In recent decades, much has been revealed about the nature of neutrinos. However, some key questions remain, including: What is the neutrino mass hierarchy? Do neutrinos violate charge-parity symmetry, and, if so, by how much? What more can be learned from observing neutrinos produced by future supernovae? Resolving these and other neutrino-centric questions is the goal of the international flagship Deep Underground Neutrino Experiment (DUNE) [1]. The massive DUNE detector modules will employ Liquid Argon Time Projection Chamber (LArTPC) technology [2, 3, 4] to observe neutrino-argon interactions occurring in the detector volume.

Neutrino interactions in the volume of a LArTPC detector produce energetic, charged particles that leave tracks of ionization electrons. The ionized electrons are made to drift toward the plane of electrodes (wires) aided by an external electric field. Signals in the wires, induced by the drifting electrons, are detected, amplified, digitized, and read out as data. The goal of the neutrino data analysis is to use the readout signals to reconstruct the original particle tracks and make inferences about the neutrino that produced them.

The LArTPC community has amassed a substantial set of algorithms for simulation and reconstruction that are codified in various large software suites too numerous to list exhaustively [5, 6, 7]. The bulk of this software consists of conventional algorithms, heuristic functions, and other codes manually designed by experts. Here, we refer to algorithms not based on Deep Neural Network (DNN) techniques as “conventional.” Results from these suites have been well validated against real detector data, utilizing methods and metrics also based on conventional algorithms. On that solid basis, techniques based on deep neural networks (DNNs) are being applied. Their inclusion along with conventional algorithms represents a clear paradigm shift for the LArTPC community.

DNN techniques bring improvements in precision and provide replacement or augmentation to the laborious manual development. These are attractive and welcome changes, but DNN techniques also bring a substantial new problem that must be solved before their results can be accepted. This problem begins with training a neural network using samples from one domain, typically that of a physics and detector *simulation*, and then applying the trained network to infer about samples from another domain, typically that of a *real detector*. With labeled data, the discrepancy may be addressed by transfer learning and domain adaptation approaches in machine learning [8, 9]. However, annotating or assigning a ground truth, such as a particle’s energy, in LArTPC data is a challenge in itself.

In this work, we propose a more general technique to reduce the differences between the simulations and experiments. And the technique does not require labeled or paired data. We call it *unsupervised unpaired data translation across domains*. At its core, it requires a transformation of individual samples of a source domain “A” (e.g. simulations) such that a transformed sample is indistinguishable from other samples

in a target domain “B” (e.g. experiments). In the meantime, a transformed sample should be highly correlated to its source. These two constraints – realism (to the target domain) and consistency (to the source domain), are somewhat competing. Our approach to translating between two domains is based on the UNet Vision-Transformer Cycle-consistent Generative Adversarial Network (UVCGAN) [10] validated on open benchmark photographic data sets in the machine learning community. UVCGAN is specifically designed to satisfy both constraints with a combination of loss functions.

Our work has two major contributions to both the experimental physics community and the machine learning community. First, we have prepared and released a large two-domain data set of simple liquid-argon track samples (SLATS). The data set represents a simplified variant of the type of data commonly used in the LArTPC community and it may serve as a benchmark data set for unpaired image translation tasks in machine learning. Due to the high-resolution nature of its 2D images, the SLATS data set presents a different challenge compared to existing public data sets of photographic images. The evaluation of transformed images can be done in a paired fashion, which is not possible in most existing data sets. Second, we adapted the state-of-the-art image-to-image translation model UVCGAN to work on SLATS.

The rest of the paper is structured as the following: in Section 2, we describe the nature of LArTPC data, how the two domains differ, and how they are simulated and pre-processed; in Section 3, we review the UVCGAN architecture and its training procedures for the SLATS data set; in Section 4, we evaluate the quality of translated images produced by UVCGAN and competing models in their immediate form and after a standard LArTPC signal processing procedure is applied.

2. The two-domain SLATS dataset

The data set we release here covers two data domains. Each domain is populated by a variant of a simulation of the LArTPC detector used in the ProtoDUNE-SP experiment[11, 12]. The two domains differ in precisely one feature – *the detector response function*. The details of this variance, the simulation in general, and the idealizations chosen are described in Section 2.1. The steps used to pre-process the data into a form that can be input to UVCGAN are described in Section 2.2.

2.1. LArTPC simulation

The Liquid Argon Time Projection Chamber (LArTPC) detectors enclose a volume of liquid argon. As illustrated in Figure 1, energetic charged particles traversing the volume will ionize electrons from nearby argon atoms. These electrons remain ionized long enough to drift under the influence of an external uniform electric field to one side of the detector. There, they drift through a series of planes of parallel wires uniformly spaced and transverse to the drift direction. The planes are labeled as “U”, “V”, and “Y” in Figure 1. Electrons will induce an electric current in nearby wires as they pass.

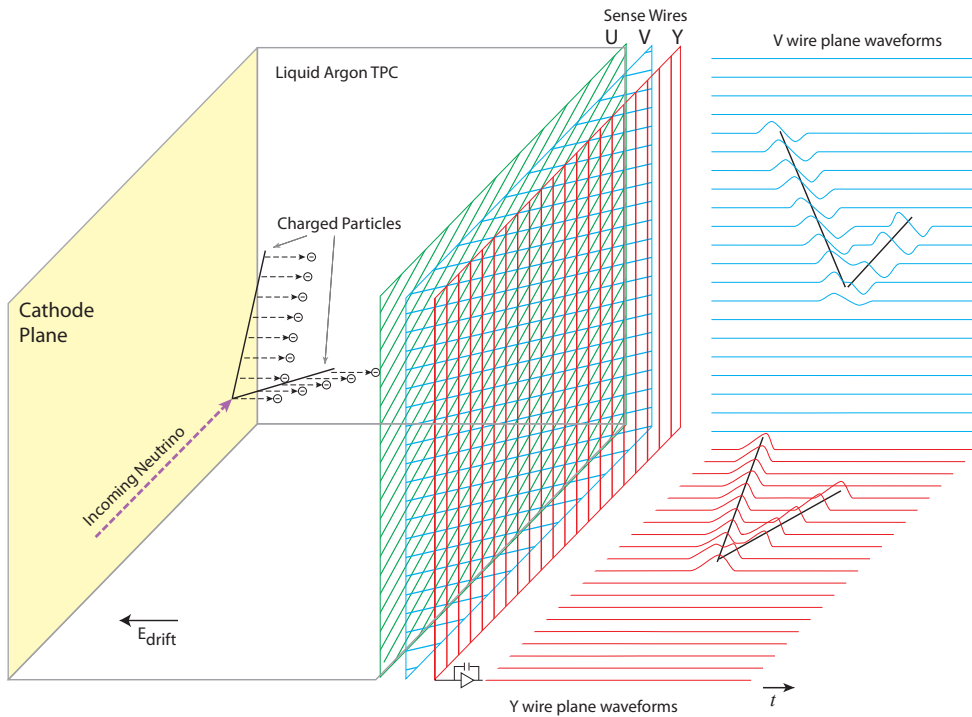


Figure 1: Illustration from [13] of a three-wire plane LArTPC and its signal formation.

Each plane is oriented at an angle with respect to the others and thus represents one tomographic view of the ionization electron distribution as a function of time and space. The space here is in the wire’s pitch direction perpendicular to the wire’s direction and the drift direction. This process explains the “time projection” part of the name LArTPC.

To simulate the process described above, we need a long and complex chain of simulations. Such a simulation chain provides neutrino flux models, interaction cross sections, nuclear transport effects, production of ionization electrons, and finally, the electron drift and the electric current response at the wire planes. Most LArTPC experiments and prototypes today utilize components of the Wire-Cell toolkit[14, 15] to simulate the last step. The toolkit is also responsible for simulating degradations happening during the drift, the induction of electric current, and the amplification and digitization of the signal.

Although a comprehensive simulation would provide us with the most realistic data, at this stage of our study, we need a more idealized data set to allow us better control and understanding of features a neural translator must learn. To construct the idealized SLATS data set, we make the following simplifications to the simulation chain. First, the steps prior to the Wire-Cell toolkit are replaced with a simple generator of idealized

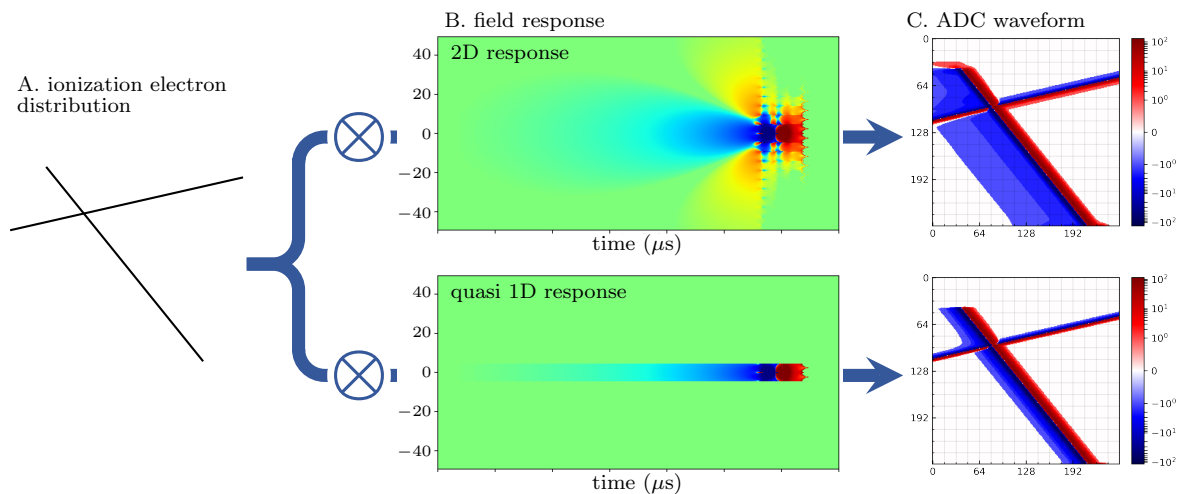


Figure 2: **Field responses and waveforms.** The ionization electron distribution (Panel A) is convolved with two types of responses to produce the two domains we translate between. 2D response (Panel B top) is used to produce domain “B” samples, and Quasi-1D response (Panel B bottom), domain “A” samples. The electronics response function and example of the final, digitized per-channel ADC waveforms (input to the translation algorithm) are shown in Panel C.

particle tracks. The tracks are in the form of line sources of uniform ionization and are distributed uniformly over the detector in time, space, and direction. Such tracks approximately mimic those of minimum-ionizing particles, such as cosmic-ray muons, which are abundant for all LArTPC detectors placed above ground. Second, inside the Wire-Cell toolkit simulation, we omit all noise models to focus purely on signals from ionization as described next.

The Wire-Cell toolkit applies the effects of electron diffusion[16, 17] and absorption while transporting the ionization electrons through a uniform drift field in the bulk of the detector volume. Near to the wire planes the ionization electrons are drifted through a far more complex electric field governed by the locations and sizes of the sense wires. This detailed drift field, as well as the associated Ramo weight fields[18] are provided to the toolkit as input. Here we use fields calculated by the GARFIELD [19] software package using a 2D model[20] of the detector electrode arrays. The two domains of the SLATS data set are distinguished by the differing nature of their fields.

Shown in Figure 2B are the electric current responses in a wire due to a nearby drifting electron in each of the two variant field models. We define samples in domain “B” as being produced with the full 2D field response model shown at the top of Figure 2B. On the other hand, samples from domain “A” are produced with a related but different response shown at the bottom of Figure 2B. It is constructed by masking the 2D response so that all contributions from regions near neighboring wires are removed. The term “quasi” is used to indicate that the response is actually still two-dimensional in the remaining narrow region near the central wire.

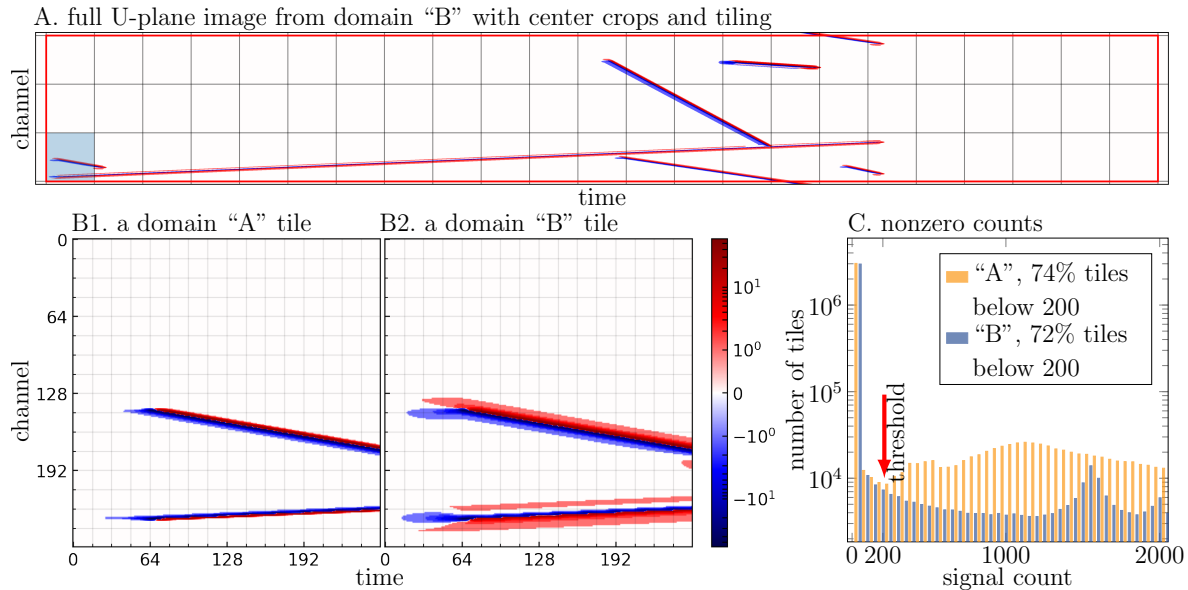


Figure 3: **Pre-processing for the SLATS dataset.** Panel A: An example full image of the U plane from domain “B” (2D). The full image has dimension (channel, time) = (800, 6000). The part bounded by the red box is the center crop of dimension (768, 5888). The center crop is divided into 3×23 tiles of size (256, 256) shown as the gray grid. Panel B1 and B2: A pair of tiles from domain “A” (quasi-1D) and “B” (2D), respectively. The tile in B2 corresponds to the highlighted tile in Panel A. Panel C: The Distribution of the number of nonzero pixels in the tiles.

After the electric current response discussed above, an electronics response and digitization model (linear scaling and truncation to 12-bit integer) are applied. The final output from the simulation is the Analog-to-Digital Converter (ADC) count illustrated in Figure 2C and it serves as the input to a neural translator.

2.2. Data generation and pre-processing

For generating the SLATS data set, the simulation runs produced 10010 events, each with 10 ideal line sources at the minimum-ionization energy equivalent for muons. Each event results in a 2D ADC readout image for each wire plane. For this data release, we will focus only on the “U” plane – the first plane the electrons encounter during their drift. Since a detector has six active regions, each with a separate set of wire planes, the runs produced a total of 60060 “U”-plane images. The image from the “U” plane is 800 pixels high and 6000 pixels wide. The height of the image represents the readout channels and the width represents the time steps.

From each full readout image of shape (800, 6000) (Figure 3A), we take a center crop of shape (768, 5888). The shape of the center crop is chosen so it can be divided into tiles of shape (256, 256) for being used as input to a neural translator. Figure 3A shows an example of the center crop (red box) of a full “U”-plane image along with the

tiles (gray grid).

In the conventional practice of analyzing LArTPC readout images, it is common to apply similar crops for physical reasons such as removing activity from background interactions originating outside the detector or providing a size more optimal for fast-Fourier transforms. Nevertheless, future work will investigate how to avoid this loss of information at the edge of the readout image.

In some instances, the randomness of placing the 10 ideal particles across the entire detector leads to one or more of the six active regions containing no ionization electrons. The resulting “empty” center crops of readout images are neglected, leaving us with 56,253 non-empty center crops (93.7%). From these, 1000 crops are reserved for the test and 55,253, for training.

Similarly, the sparseness of activity leads to a majority of tiled patches being fully or nearly empty. To reject these tiles we form the distribution of the number of nonzero pixels in tiles from both domains as shown in Figure 3C. We chose a threshold of 200 pixels around the first local minimum of the distribution for domain “A”. And to keep the tiles paired, we drop a pair if either domain “A” tile or its domain “B” counterpart falls below the threshold. After the filtering, we have 1,065,870 tile pairs for training and 18,887 for testing.

Figure 3B1 and B2 show a pair of tiles from domain “A” and “B”. We can observe two major differences between the domains. First, the domain “B” track exhibits long-range induction effects in both the longitudinal (time, horizontal) and transverse spatial (channel, vertical) axes while the “A” track shows variation only in the former. This leads to domain “B” tracks exhibiting generally broader, less compact features than domain “A”. In particular, larger “lobe” structures are seen at the end of domain “B” tracks while domain “A” tracks end more abruptly. This can also lead to features in domain “B” tiles that are not present in the corresponding tile from domain “A” as shown in the small red lobe between the two tracks in Figure 3. Second, domain “B” has a larger neighborhood where the ionization distribution can lead to interference patterns of the bipolar response. This is seen in the red lobe above the bottom track in the figure.

3. Deep generative models for unpaired image translation

The model, UVCGAN, we adapted for unpaired translation on the SLATS data set is a CycleGAN-like model. We chose a CycleGAN-like model not only because CycleGAN [21] is a simple model that demonstrated promising performance on photographic data sets but also because it is a *deterministic* model. As we will review in Section 3.4, many algorithms designed for unpaired image-to-image translation introduce artificial randomness in image generation. Randomness is the source of diversity and hence is indispensable in data synthesis. It is also helpful in boosting diversity in photographic image translation tasks as there is not one correct translation for an input. However, the application of artificial randomness needs to be carefully

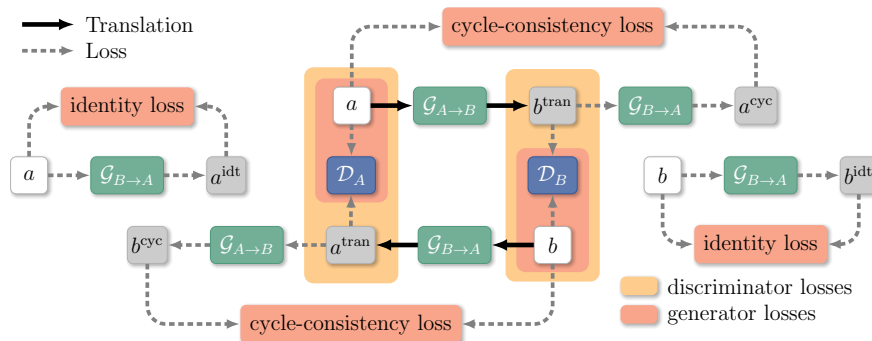


Figure 4: **Summary of the CycleGAN [21] model.** The CycleGAN consists of two pairs of GANs, $(\mathcal{G}_{A \rightarrow B}, \mathcal{D}_B)$ and $(\mathcal{G}_{B \rightarrow A}, \mathcal{D}_A)$. The discriminators, \mathcal{D}_A and \mathcal{D}_B , try to distinguish translations from real images while the generators (or translators), $\mathcal{G}_{A \rightarrow B}$ and $\mathcal{G}_{B \rightarrow A}$, try to produce realistic translations that are consistent with the source.

justified in scientific data translation. Specifically, for this study on the idealized SLATS data set, we assume that the map between the two domains should be *one-to-one* in nature, and hence a deterministic model is an appropriate choice.

In Section 3.1, we discuss the CycleGAN-like models. In Section 3.2, we detail the innovative UVCGAN generator. The pre-training and training procedures of UVCGAN are discussed in Section 3.3. And finally in Section 3.4, we review other DNN models for unpaired image-to-image translation, and the benchmarking models we compare UVCGAN to.

3.1. Cycle-consistency for preserving source information

A generative adversarial network (GAN) consists of two networks, a generator (\mathcal{G}) and a discriminator (\mathcal{D}). The two networks are pitted in a minimax game that the generator tries to synthesize data to fool the discriminator; whereas the discriminator aims to discern the synthesized data from the real ones. GANs are used wildly in data synthesis where they produce highly authentic images, soundtracks, text, etc. from random input [22, 23, 24, 25, 26]. But there is a problem when applying GANs to translation because the input is no longer random and the translation should not only resemble those from the target domain but also carry information from the input. A solution provided by CycleGAN-like model [21] for translating with GANs is to use *two* GANs, one for each translation direction, as shown in Figure 4. With the pair of GANs, CycleGAN-like models can form loops of translation so that the information loss during the translation could be measured and consequently penalized to encourage consistency to the source.

Denote the two domains by A and B and let $(\mathcal{G}_{A \rightarrow B}, \mathcal{D}_B)$ and $(\mathcal{G}_{B \rightarrow A}, \mathcal{D}_A)$ be the two GANs. For domain A , the cycle-consistency loss equals $\|a - a^{\text{cyc}}\|$ where $a^{\text{cyc}} = \mathcal{G}_{B \rightarrow A}(\mathcal{G}_{A \rightarrow B}(a))$, and similarly for domain B . The heuristics behind the cycle-consistency loss is as the following. Let $a \in A$ be a source image, since the translation

$\mathcal{G}_{A \rightarrow B}(a)$ is supposed to resemble those from domain B , it is in general unclear how to compare the translation $\mathcal{G}_{A \rightarrow B}(a)$ to a directly. However, if we can recover a to some extent from $\mathcal{G}_{A \rightarrow B}(a)$ by translating it back to A via $\mathcal{G}_{B \rightarrow A}$ (i.e. require $\mathcal{G}_{B \rightarrow A}(\mathcal{G}_{A \rightarrow B}(a))$ to be close to a), we can trust that $\mathcal{G}_{A \rightarrow B}(a)$ carries substantial information from a .

In addition to the cycle-consistency loss, for some translation tasks, we can also use the identity loss to encourage the generator to retain features of the source which are also present in the target domain. Concretely, for domain A , the identity loss equals $\|a - a^{\text{idt}}\|$, where $a^{\text{idt}} = \mathcal{G}_{B \rightarrow A}(a)$, and similarly for domain B .

3.2. UVCGAN: U-Net Vision-transformer CycleGAN

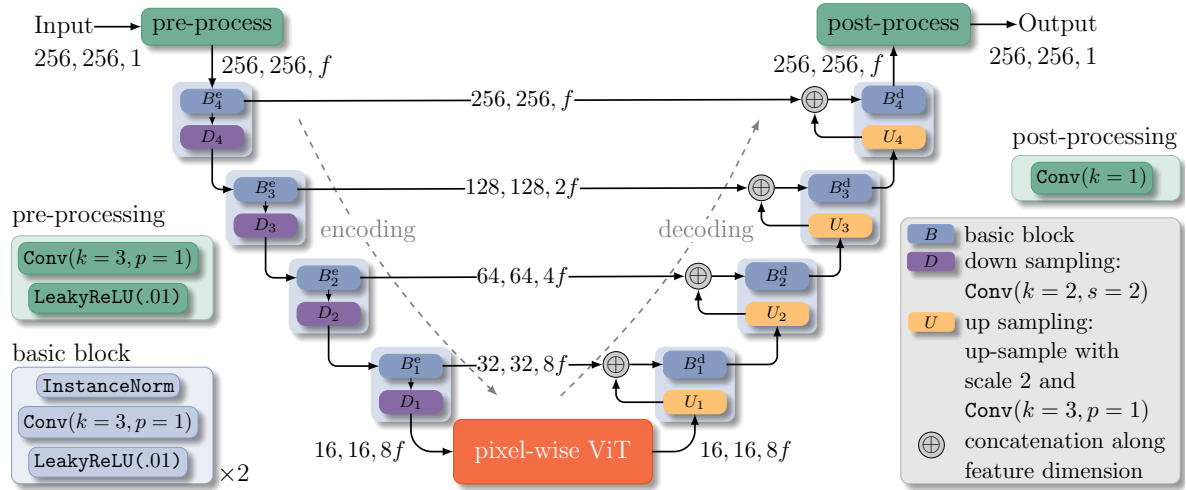


Figure 5: **UVCGAN Generator.** The generator of the UVCGAN is a U-Net [27] with a Vision Transformer (ViT) [28] bottleneck.

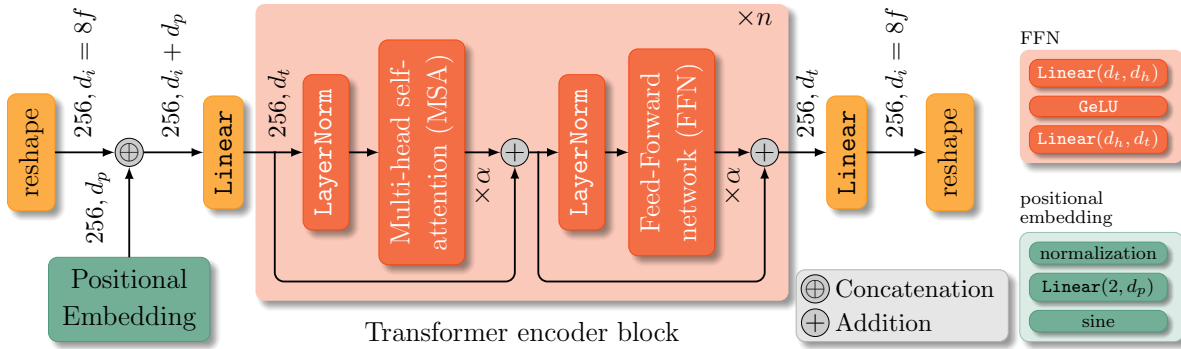


Figure 6: **Pixel-wise Vision Transformer.** The main body of the ViT is a stack of Transformer encoder blocks.

The model U-Net Vision-transformer CycleGAN (UVCGAN) improves upon the CycleGAN model by introducing a vision-transformer block at the bottleneck of the U-

Net generator. A U-Net consists of a series of convolutional layers capturing local short-ranged patterns [28]. Activations at different scales are passed directly to the decoding part via skip connections. Convolution layers are in general ineffective in modeling long-range dependencies, which are common in scientific applications. On the contrary, Transformer-based [29] architectures excel at capturing long-range dependencies but may struggle with modeling local pixel patterns [28]. These observations motivated the design of a hybrid generator architecture for UVCGAN that combined the best of both worlds – a U-Net for handling short-range pixel patterns with a Transformer-based bottleneck for modeling long-range dependencies. We describe the working of the UVCGAN generator briefly below and list the exact corresponding parameter values used for SLATS tiles in parentheses.

The pre-processing layer of the U-Net (Figure 5) maps a single-channel (12-bit gray-scaled) SLATS tile to a tensor with the same spatial dimension and f (48) features (or channels). Each successive downsampling layer (D_4, \dots, D_1) halves the width and height of the input, while each of the last three basic blocks (B_3^e, B_2^e, B_1^e) doubles the features. The output from the encoding path of the U-Net is a tensor of dimension $(16, 16, 8f)$ and it serves as the input to the ViT. The pixel-wise ViT (Figure 6) first flattens an image encoding along the spatial dimensions to form a sequence of tokens. The sequence has length 16×16 , and each token is a vector of length $d_i = 8f$ (384). It then concatenates each token with its two-dimensional Fourier positional embedding [30] of length d_p (384) and linearly maps the result to a vector of length d_t (384). The main body of the pixel-wise ViT is a stack of n (12) Transformer encoder blocks [31], each of which consists of a multi-head attention [29] with 6 heads and a feed-forward network with $d_h = 4d_t$ (1536) hidden units. To accelerate the Transformer convergence, rezero regularization [32]) is used (a trainable scalar α) to modulate the magnitude of the main path of the residual blocks. The output from the Transformer stack is linearly mapped to have dimension d_i and reshaped back to a tensor of shape $(16, 16, d_i)$. Finally, on the decoding path of the UNet, the upsampling layers (U_1, \dots, U_4) double the width and height while the basic blocks (B_1^d, \dots, B_4^d) and the post-processing layer shrink the feature dimension to form a translated tile.

3.3. UVCGAN pretraining and training procedures

3.3.1. Pretraining. Typically, image translation generators are randomly initialized. The training procedure moves these generators from random states into the final configurations, capable of domain translation. However, similar to the findings of UVCGAN, we discovered that initializing the generators by pretraining them on a simpler task provides an advantage over random initialization.

As a simpler task to pretrain the generators, we use an image inpainting task. For this task, SLATS images are tiled into a grid of patches of size 32×32 pixels. Then, each patch in the grid is randomly masked by zeros with a probability of 40%. The generators are pretrained to recover the masked regions via ℓ_2 loss. This pretraining

allows the generators to learn non-trivial dependencies between different parts of the SLATS image, which may be helpful for the subsequent domain translation.

We pretrain both $\mathcal{G}_{A \rightarrow B}$ and $\mathcal{G}_{B \rightarrow A}$ for 16,384,000 iterations on the image inpainting task configured similarly to [10]. We use a smaller learning rate of 6.25×10^{-6} since SLATS data has a larger range compared to natural images.

However, we found that the pretrained generators fail to recover the full width of the tracks, filling masked regions with skinnier tracks. We speculated this happens because pixel values away from the track cores are very small compared to the cores. Therefore, their proper reconstruction gives a very small benefit in terms of the ℓ_2 loss. On the other hand, before the network learns to reconstruct these small-valued pixels properly, it is going to make a lot of mistakes. These mistakes are costly in terms of the ℓ_2 loss. The high cost of mistakes compared to the small benefit of proper reconstruction creates some sort of potential barrier to learning the full width of the tracks.

To lower the learning barrier, we modify the ℓ_2 loss function. We reduce the penalty for the network to incorrectly overwrite zeros by α . Let y be an image from either domain ‘‘A’’ or ‘‘B’’ and let \hat{y} be the inpainting output, the reconstruction loss is defined as follows:

$$L_{\text{reco}}(\hat{y}, y) = \frac{\alpha \cdot \sum_{y_{i,j}=0} \hat{y}_{i,j}^2 + \sum_{y_{i,j} \neq 0} (\hat{y}_{i,j} - y_{i,j})^2}{H \times W} \quad (1)$$

where H and W are the height and width of the image.

During the pretraining, we kept α at 0 for the first 819,200 iterations, allowing the network to freely overwrite the empty space without penalty. Then, we linearly annealed α to 1 during the subsequent 2,457,600 iterations. When $\alpha = 1$, the loss function in Equation (1) reduces to the normal ℓ_2 and remains that way till the end of pretraining. An ablation study showed the modified ℓ_2 loss sped up the learning of the reconstruction of small-valued pixels. The generators trained with the modified ℓ_2 loss also achieved a $\sim 10\%$ lower reconstruction error than generators trained with the normal ℓ_2 .

3.3.2. Domain Translation Training. The translation on the SLATS data set was trained for 200 epochs with 5000 randomly selected tiles per epoch (hence 1,000,000 iterations in total). We kept the learning rate constant for the first half of the training and linearly annealed it to zero during the second half. We found out that using slightly unequal initial learning rates for generators (10^{-5}) and discriminators (5×10^{-5}) improved the performance. We also performed a small-scale hyperparameter optimization on coefficients of cycle-consistency loss, λ_a and λ_b , and the gradient penalty parameters, λ_{GP} and γ . The results reported in the evaluation (Section 4) were produced with the best model found in the optimization with $\lambda_a = \lambda_b = 1$, $\lambda_{\text{GP}} = 1$, and $\gamma = 10$. Identity loss was also used for SLATS training with coefficients kept at half of those for cycle consistency. See [10] for a more detailed discussion on loss coefficients and gradient penalty.

3.4. Alternative generative models for unpaired image translation

Many recent models for unpaired image-to-image translation have been developed based on photographic image data sets. The models can be grouped with respect to how translation is generated and how consistency is enforced. For example, according to image generation, CycleGAN [21], ACL-GAN [33], U-GAT-IT [34], Council-GAN [35], and UVCGAN [10] were GAN-based methods, CUT [36] adopted the contrastive learning approach, LETIT [37] utilized the energy transport on the latent feature space of the images, and EGSDE [38] and ILVR [39] were based on diffusion models. According to consistency enforcement, CycleGAN, ACL-GAN, U-GAT-IT, UVCGAN, CUT, and EGSDE imposed explicit consistency constraints via loss functions, while other methods did so implicitly. Another key feature that concerns us specifically is whether randomness is used in image generation. Among the aforementioned models, CycleGAN, U-GAT-IT, UVCGAN, and CUT are deterministic models, while Council-GAN, EGSDE, and ILVR inject randomness into the image generation process.

Given the limitation on time and computing power, we decided to benchmark the performance of UVCGAN on three GAN-based algorithms: CycleGAN, ACL-GAN, and U-GAT-IT. All three benchmarking algorithms explicitly enforce consistency and two of them, CycleGAN and U-GAT-IT, are deterministic. In general, we didn't consider models injecting randomness into image generation, but we kept ACL-GAN for benchmarking to demonstrate the adverse effect of randomness (see Section 4).

To make the algorithms designed for photographic images run on SLATS, we turned off data normalization and augmentation, changed the input channel to one (SLATS contains gray-scaled images while photographic images normally have three channels – RGB), and removed the final \tanh activation.

For each benchmarking model, we did a small-scaled hyperparameter (HP) tuning. The results from the best performers can be found in Section 4 and more details on HP tuning and full results with all HP settings can be found in Appendix A.

4. Evaluation

In this section, we evaluate the performance of UVCGAN and compare it with those of three benchmarking neural translators. The paired SLATS data set allows us to do a direct comparison of the translated image with the target image. In Section 4.1, we evaluate the translation quality with pixel-level mean absolute error (ℓ_1) and mean squared error (ℓ_2) and visualize common defects that appear in the images produced by the benchmarking algorithms. In Section 4.2, we evaluate the translations after applying a signal processing procedure.

4.1. Evaluating translation quality at pixel level

To evaluate the quality of the translation, we exploit the pairing between samples in the source and target domain. We use ℓ_1 (mean absolute error) and ℓ_2 (mean squared

Table 1: **Translation performance comparison with ℓ_1 and ℓ_2 differences.** The differences are produced with the best performer of each benchmarking algorithm.

algorithm	“A” to “B”		“B” to “A”	
	ℓ_1	ℓ_2	ℓ_1	ℓ_2
ACL-GAN	0.083	0.566	0.039	0.121
CycleGAN	0.074	0.180	0.061	0.159
U-GAT-IT	0.078	1.187	0.073	1.161
UVCGAN	0.030	0.033	0.025	0.027

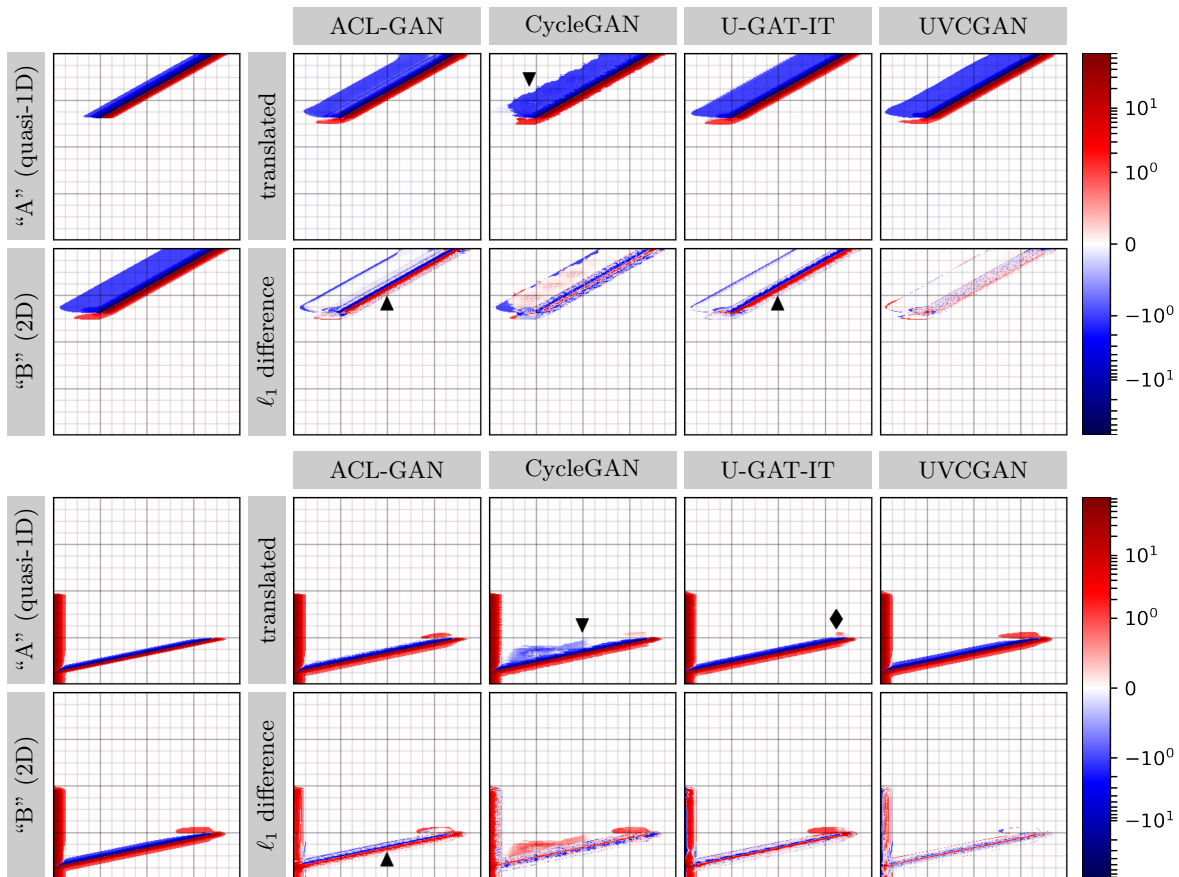


Figure 7: **Translation examples.** Here we show two examples of translation from “A” (quasi-1D) to “B” (2D). We mark defects appearing in the translations produced by the benchmarking algorithms: ▼ for rugged track edge; ▲ for big error in the core of the track where the signal is the strongest; ◆ for missing blob-shaped track tip.

error) to judge the quality of the translation and summarize the best result of each benchmarking algorithm and that of UVCGAN in Table 1. Full benchmarking results for all hyper-parameter settings can be found in Table A1 of Appendix A.

We show two sample translations in Figure 7. The images are also generated

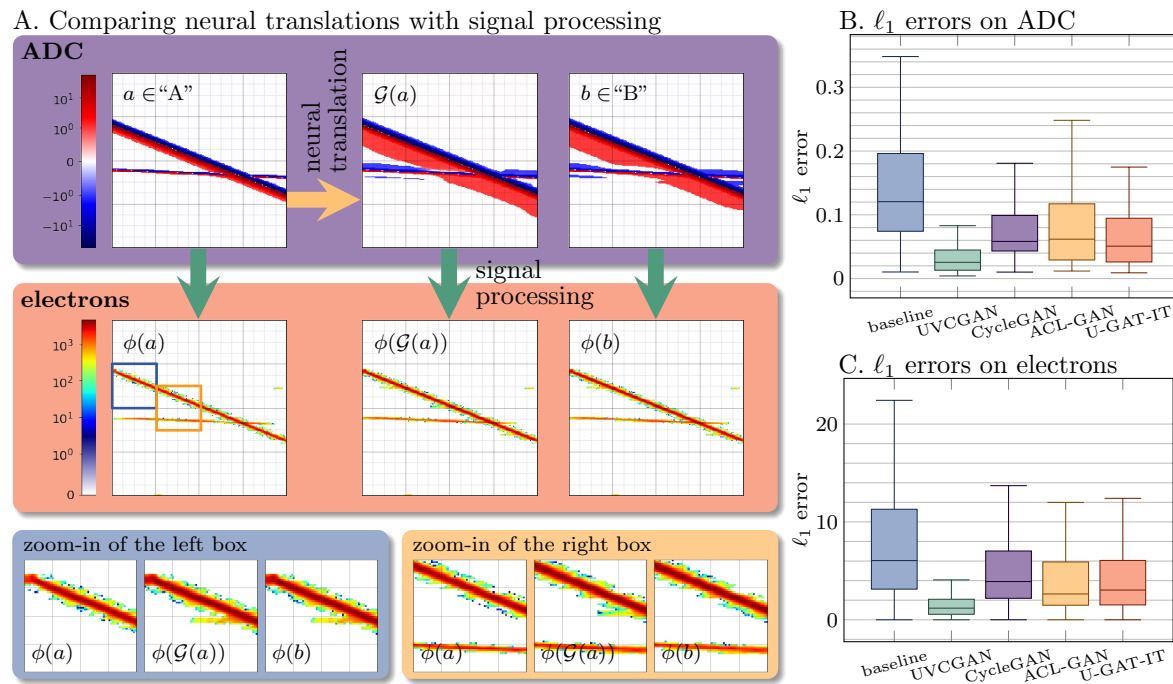


Figure 8: **Signal processing study.** Panel A. Diagram of the signal processing study. An image a from domain “A” is translated by a neural translation algorithm \mathcal{G} to an image $\mathcal{G}(a)$ that is supposed to resemble a ’s counterpart b from domain “B”. In this study, domain “A” is the ADC generated with the quasi-1D response, while domain “B” is ADC generated with the 2D response (See Section 2.1). We apply signal processing ϕ to a , $\mathcal{G}(a)$, and b so that the electrons can be reconstructed. We zoom in on two areas exhibiting artifacts that result from signal processing and neural translation, respectively. For the box on the left, we can see that the translation reproduced the mismatch caused by applying signal processing derived from domain “A” to domain “B”. For the box on the right, we can see that the translation introduced artifacts to the signal processing output. Panel B. Comparison of ℓ_1 errors on ADC values. Panel C. Comparison of ℓ_1 errors on electrons obtained with signal processing. For the ADC values, the baseline ℓ_1 error equals $\|a - b\|_1$, and the translation ℓ_1 error equals $\|\mathcal{G}(a) - b\|_1$. Those for the electrons are defined similarly. From B and C, we can see that UVCGAN shows even more advantage over the benchmarking algorithms with respect to the more sensitive measurements provided by signal processing.

with the best-performing parameters of each benchmarking algorithm. We point out that there are several obvious defects appearing in the translations produced by the benchmarking algorithms: rugged track edge, big error in the track center, and missing track tip.

4.2. Evaluation based-on downstream signal processing

In practice, the bipolar nature of LArTPC ADC waveforms obscures an accurate and precise measurement of the underlying distribution of ionization electrons. In order to reveal this distribution so that physically meaningful parameters about the original particles can be reconstructed (e.g. their momentum and mass) a procedure generically called *signal processing* is applied[20, 40]. In a nutshell, signal processing has two stages: deconvolution and high-pass filtering. First, it performs a deconvolution of an ADC readout image with a model of the same detector response used in the simulation but averaged over each region near a wire. The bipolar nature of the response causes the deconvolution to inevitably amplify low-frequency noise. To counter that, the second stage applies an adaptive high-pass filter known as signal region-of-interest (ROI) selection. Further details can be found in [20].

Due to the inevitable amplification of noise described above, signal processing is designed to contend with realistic detector noise by the application of various filters. The interplay of the input noise, the filters, and the thresholds to define ROI make the signal processing very sensitive to the presence of noise, or the lack thereof. Therefore, the simplifying assumption to neglect noise that we have made thwarts the application of signal processing. Post-processing of the noise-free ADC waveforms is thus performed in order to add a realistic noise component. To do this, we linearly scale ADC pixel values to be consistent with the voltage levels originally produced by the amplifiers in the electronics prior to digitization. We then add noise generated from a model that has been previously developed to match observations of LArTPC detectors. Finally, we rescale (re-digitize) the result back to ADC levels. The signal processing may then be correctly applied.

An example of the results of the procedure is shown in Figure 8A. The difference between the domains in the ADC data is broad and large. This is expected due to the fact that the domain “B” response spreads ionization distribution information across a larger region than does the domain “A” response. The result after signal processing compacts this diffuse information. As the signal processing uses a response derived from that used to simulate domain “A”, it produces an ideal solution when applied to that same domain. When applied to domain “B”, a sub-optimal yet reasonably good result is obtained. Spurious artifacts can be seen in the domain “B” result which is due to a conflation of the response mismatch and the imperfect filtering of inescapable noise. To see such an example, we zoom in to the areas of $\phi(a)$, $\phi(\mathcal{G}(a))$, and $\phi(b)$ as marked by the boxes in $\phi(a)$. The fact that the artifact appears in both $\phi(\mathcal{G}(a))$ and $\phi(b)$ attests to the realism of the translation. However, we can also observe an artifact from the zoom-ins of the box on the right which appears in $\phi(\mathcal{G}(a))$ only. Such an artifact is a result of the neural translation and indicates a need for further perfection of the algorithm.

In order to better benchmark the four algorithms, we form a sub-sample containing 100 tiles spanning a variety of track topologies and compare the ℓ_1 statistics on ADC values and signal processing results in Figure 8B and C. For the ADC values, the baseline

ℓ_1 error equals $\|a - b\|_1$, and the translation ℓ_1 error equals $\|\mathcal{G}(a) - b\|_1$. Those for the electrons are defined similarly. We demonstrate in Section 4.1 that UVCGAN produces higher quality translations than the benchmarking algorithms with respect to ADC. In Figure 8B and C, we can see that UVCGAN shows even more advantage over the benchmarking algorithms with respect to the more sensitive signal processing results. A visualized performance comparison on three examples can be found in Appendix B.

Conclusion

In this study, we demonstrated the potential of deep neural network-based algorithms for unpaired translation between simulated and detector data from neutrino physics facilities. To lay a solid foundation, we resorted to an idealized simulation to generate paired datasets with two types of field responses: quasi-1D and 2D. This allows us to validate our method using paired metrics such as pixel-wise ℓ_1 or ℓ_2 . The dataset generated with a more realistic 2D response serves as a proxy for real detector data. We pre-processed the datasets to form input to neural translation algorithms and made the original and derived datasets available for the use of the larger community. We proposed a neural translator, UVCGAN, that is based on the CycleGAN framework. We compared UVCGAN with other existing unpaired translation algorithms and demonstrated its advantage on pixel-level accuracy. Finally, we applied signal processing to reconstruct the particles from ADC readings and showed that UVCGAN exhibits even more advantage with respect to this more sensitive measurement.

In future, in order to apply our method in real-world detector data, several problems need to be addressed. First, because the two domains will be truly unpaired, how to measure the quality of translated data is a challenge. While there are widely accepted metrics on the realism for photographic data sets, such a metric is missing for scientific data sets. Second, how to measure and ensure the translation does not alter the physics meaning of the source is another difficulty. Hence, rigorous and physically meaningful consistency criteria and measurements need to be developed with neural translators designed and evaluated accordingly. Finally, how to strike a good balance between realism to the target and consistency to the source needs to be explored.

Acknowledgment

This work was supported by the Laboratory Directed Research and Development Program of Brookhaven National Laboratory, which is operated and managed for the U.S. Department of Energy Office of Science by Brookhaven Science Associates under contract No. DE-SC0012704.

- [1] B. Abi et al. Volume i. introduction to dune. *Journal of Instrumentation*, 15(08):T08008, aug 2020.
- [2] Carlo Rubbia. The liquid-argon time projection chamber: a new concept for neutrino detectors. Technical report, 1977.
- [3] W.J. Willis and V. Radeka. Liquid-argon ionization chambers as total-absorption detectors. *Nuclear Instruments and Methods*, 120(2):221–236, 1974.
- [4] Veljko Radeka, Hucheng Chen, Grzegorz Deptuch, Gianluigi De Geronimo, Francesco Lanni, Shaorui Li, Neena Nambiar, Sergio Rescia, Craig Thorn, Ray Yarema, and Bo Yu. Cold electronics for ”giant” liquid argon time projection chambers. *Journal of Physics: Conference Series*, 308(1):012021, jul 2011.
- [5] S. Agostinelli et al. Geant4—a simulation toolkit. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 506(3):250–303, 2003.
- [6] E. L. Snider and G. Petrillo. LArSoft: Toolkit for Simulation, Reconstruction and Analysis of Liquid Argon TPC Neutrino Detectors. *J. Phys. Conf. Ser.*, 898(4):042057, 2017.
- [7] MicroBooNE Collaboration et al. Wire-cell 3d pattern recognition techniques for neutrino event reconstruction in large lartpcs: algorithm description and quantitative evaluation with microboone simulation. *Journal of Instrumentation*, 17(01):P01037, jan 2022.
- [8] Sinno Jialin Pan and Qiang Yang. A survey on transfer learning. 22(10):1345–1359. Conference Name: IEEE Transactions on Knowledge and Data Engineering.
- [9] Mei Wang and Weihong Deng. Deep visual domain adaptation: A survey. 312:135–153.
- [10] Dmitrii Torbunov, Yi Huang, Haiwang Yu, Jin Huang, Shinjae Yoo, Meifeng Lin, Brett Viren, and Yihui Ren. Uvcgan: Unet vision transformer cycle-consistent gan for unpaired image-to-image translation. *arXiv preprint arXiv:2203.02557*, 2022.
- [11] B. Abi et al. The single-phase protodune technical design report, 2017.
- [12] B. Abi et al. First results on protodune-sp liquid argon time projection chamber performance from a beam test at the cern neutrino platform. *Journal of Instrumentation*, 15(12):P12004, dec 2020.
- [13] R. Acciarri et al. Design and Construction of the MicroBooNE Detector. *JINST*, 12(02):P02017, 2017.
- [14] X. Qian, C. Zhang, B. Viren, and M. Diwan. Three-dimensional imaging for large LArTPCs. *Journal of Instrumentation*, 13(05):P05032–P05032, may 2018.
- [15] Wire-Cell Team. Wire-Cell Toolkit. <https://wirecell.bnl.gov/>, 2023. [Online; accessed 06-Feb-2023].
- [16] Yichen Li, Thomas Tsang, Craig Thorn, Xin Qian, Milind Diwan, Jyoti Joshi, Steve Kettell, William Morse, Triveni Rao, James Stewart, Wei Tang, and Brett Viren. Measurement of longitudinal electron diffusion in liquid argon. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 816:160–170, 2016.
- [17] P. Cennini et al. Performance of a three-ton liquid argon time projection chamber. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 345(2):230–243, 1994.
- [18] S. Ramo. Currents induced by electron motion. *Proceedings of the IRE*, 27(9):584–585, 1939.
- [19] Rob Veenhof. Garfield, recent developments, 1998.
- [20] C. Adams et al. Ionization electron signal processing in single phase LArTPCs. part i. algorithm description and quantitative evaluation with MicroBooNE simulation. *Journal of Instrumentation*, 13(07):P07006–P07006, jul 2018.
- [21] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.
- [22] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair,

- Aaron Courville, and Yoshua Bengio. Generative adversarial networks. *Communications of the ACM*, 63(11):139–144, 2020.
- [23] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.
- [24] Weili Nie, Nina Narodytska, and Ankit Patel. Relgan: Relational generative adversarial networks for text generation. In *International conference on learning representations*, 2019.
- [25] Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, and Yi-Hsuan Yang. Musegan: Multi-track sequential generative adversarial networks for symbolic music generation and accompaniment. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- [26] Javier Nistal, Stefan Lattner, and Gael Richard. Drumgan: Synthesis of drum sounds with timbral feature conditioning using generative adversarial networks. *arXiv preprint arXiv:2008.12073*, 2020.
- [27] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.
- [28] Bichen Wu, Chenfeng Xu, Xiaoliang Dai, Alvin Wan, Peizhao Zhang, Zhicheng Yan, Masayoshi Tomizuka, Joseph Gonzalez, Kurt Keutzer, and Peter Vajda. Visual transformers: Token-based image representation and processing for computer vision. *arXiv preprint arXiv:2006.03677*, 2020.
- [29] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.
- [30] Ivan Anokhin, Kirill Demochkin, Taras Khakhulin, Gleb Sterkin, Victor Lempitsky, and Denis Korzhenkov. Image generators with conditionally-independent pixel synthesis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14278–14287, 2021.
- [31] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding, 2019.
- [32] Thomas Bachlechner, Bodhisattwa Prasad Majumder, Henry Mao, Gary Cottrell, and Julian McAuley. Rezero is all you need: Fast convergence at large depth. In *Uncertainty in Artificial Intelligence*, pages 1352–1361. PMLR, 2021.
- [33] Yihao Zhao, Ruihai Wu, and Hao Dong. Unpaired image-to-image translation using adversarial consistency loss. In *European Conference on Computer Vision*, pages 800–815. Springer, 2020.
- [34] Junho Kim, Minjae Kim, Hyeonwoo Kang, and Kwanghee Lee. U-gat-it: Unsupervised generative attentional networks with adaptive layer-instance normalization for image-to-image translation. *arXiv preprint arXiv:1907.10830*, 2019.
- [35] Ori Nizan and Ayellet Tal. Breaking the cycle-colleagues are all you need. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7860–7869, 2020.
- [36] Taesung Park, Alexei A Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive learning for unpaired image-to-image translation. In *European conference on computer vision*, pages 319–345. Springer, 2020.
- [37] Yang Zhao and Changyou Chen. Unpaired image-to-image translation via latent energy transport. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16418–16427, 2021.
- [38] Min Zhao, Fan Bao, Chongxuan Li, and Jun Zhu. Egsde: Unpaired image-to-image translation via energy-guided stochastic differential equations. *arXiv preprint arXiv:2207.06635*, 2022.
- [39] Jooyoung Choi, Sungwon Kim, Yonghyun Jeong, Youngjune Gwon, and Sungroh Yoon. Ilvr: Conditioning method for denoising diffusion probabilistic models. *arXiv preprint arXiv:2108.02938*, 2021.
- [40] C. Adams et al. Ionization electron signal processing in single phase LArTPCs. part II.

data/simulation comparison and performance in MicroBooNE. *Journal of Instrumentation*, 13(07):P07007–P07007, jul 2018.

Appendix A. More evaluation of translation quality

We compared the performance of UVCGAN with three algorithms of similar training time: ACL-GAN, CycleGAN, and U-GAT-IT. Since all benchmarking algorithms were originally designed for photographic image translation, they use \tanh at the final layer to limit the pixel value within $[-1, 1]$. However, since SLATS data is integer-valued and used directly as input without any pre-processing, we removed the final \tanh activations from all three algorithms.

For **ACL-GAN**, we used three hyper-parameter (HP) sets, one for each of the three unpaired translation tasks studied in the paper. The three tasks are selfie-to-anime, male-to-female, and eye-glasses removal. Since ACL-GAN does not train translations in both directions jointly, we trained a total of six models, one for each translation direction and parameter set. For each model, we trained with a batch size of 4 for 250000 iterations, which means a total of one million images are used for training. ACL-GAN can generate a variable number of outputs each with a randomly generated style. To compare directly with other algorithms, we generated one output for any input and use 1 as the random seed.

For **CycleGAN**, we did a grid search on two key HP values: generator architecture and the coefficient for the cycle consistency loss. We evaluated the ResNet generator with 9 blocks and the U-Net generator with size 256 input. The cycle consistency was evaluated at values 1, 5, and 10 (default). Since CycleGAN trains both translators jointly, we trained in total six models, one for each generator and cycle consistency. For each model, we trained on 5000 images (with batch size 4) for 200 epochs, which means a total of one million images are used for training.

For **U-GAT-IT**, we tried three cycle consistency (λ_{cyc}) values of 1, 5, and 10 (default). And by following the default of U-GAT-IT, we kept the identity consistency equal to the cycle consistency. Since U-GAT-IT also trains both translators jointly, we trained in total three models. For each model, we trained with a batch size of 4 for 250000 iterations, which means a total of one million images are used for training.

We evaluate translation quality with respect to ℓ_1 and ℓ_2 on all the benchmarking algorithms with different HP settings and listed the result in Table A1

Appendix B. More Signal Processing Results

In this section, we show additional signal processing results. In Figure B1, we compare the signal processing result on three events and evaluate the performance using a modified version of the percentage difference. For two scalars $x, y \geq 0$, the modified percentage difference (abbr. mpd) equals $100 \cdot (y - x)/(x + y)$ if $x + y > 0$ and 0 otherwise. Note that, the score mpd thus defined has a symmetric range, $(-100, 100)$. We list the mean absolute modified percentage difference (mampd) of the baseline (signal processed “A” v.s. signal processed “B”) as the header of the first column of Figure B1 and the mampds of the translations on the individual images. We can see that while all

Table A1: **Translation performance comparison.** Benchmarking of the evaluated algorithms trained on the LArTPC data sets in terms of ℓ_1 and ℓ_2 differences between translated and target.

algorithm	HP variant	“A” to “B”		“B” to “A”	
		ℓ_1	ℓ_2	ℓ_1	ℓ_2
ACL-GAN	anime HP	0.219	5.476	0.180	5.188
	gender HP	0.079	0.727	0.065	0.330
	glasses HP	0.083	0.566	0.039	0.121
CycleGAN	(ResNet, 1)	0.266	6.123	0.202	5.180
	(ResNet, 5)	0.171	2.947	0.235	5.449
	(ResNet, 10)	0.147	2.469	0.322	10.451
	(UNet, 1)	0.089	0.177	0.056	0.114
	(UNet, 5)	0.078	0.178	0.062	0.147
	(UNet, 10)	0.074	0.180	0.061	0.159
U-GAT-IT	$\lambda_{\text{cyc}} = 1$	0.086	1.367	0.069	0.997
	$\lambda_{\text{cyc}} = 5$	0.078	1.187	0.073	1.161
	$\lambda_{\text{cyc}} = 10$	0.079	1.404	0.075	1.217
UVCGAN		0.030	0.033	0.025	0.027

neural translators bring some improvement with respect to mampd over the baseline, translations produced by UVCGAN achieve the best performance.

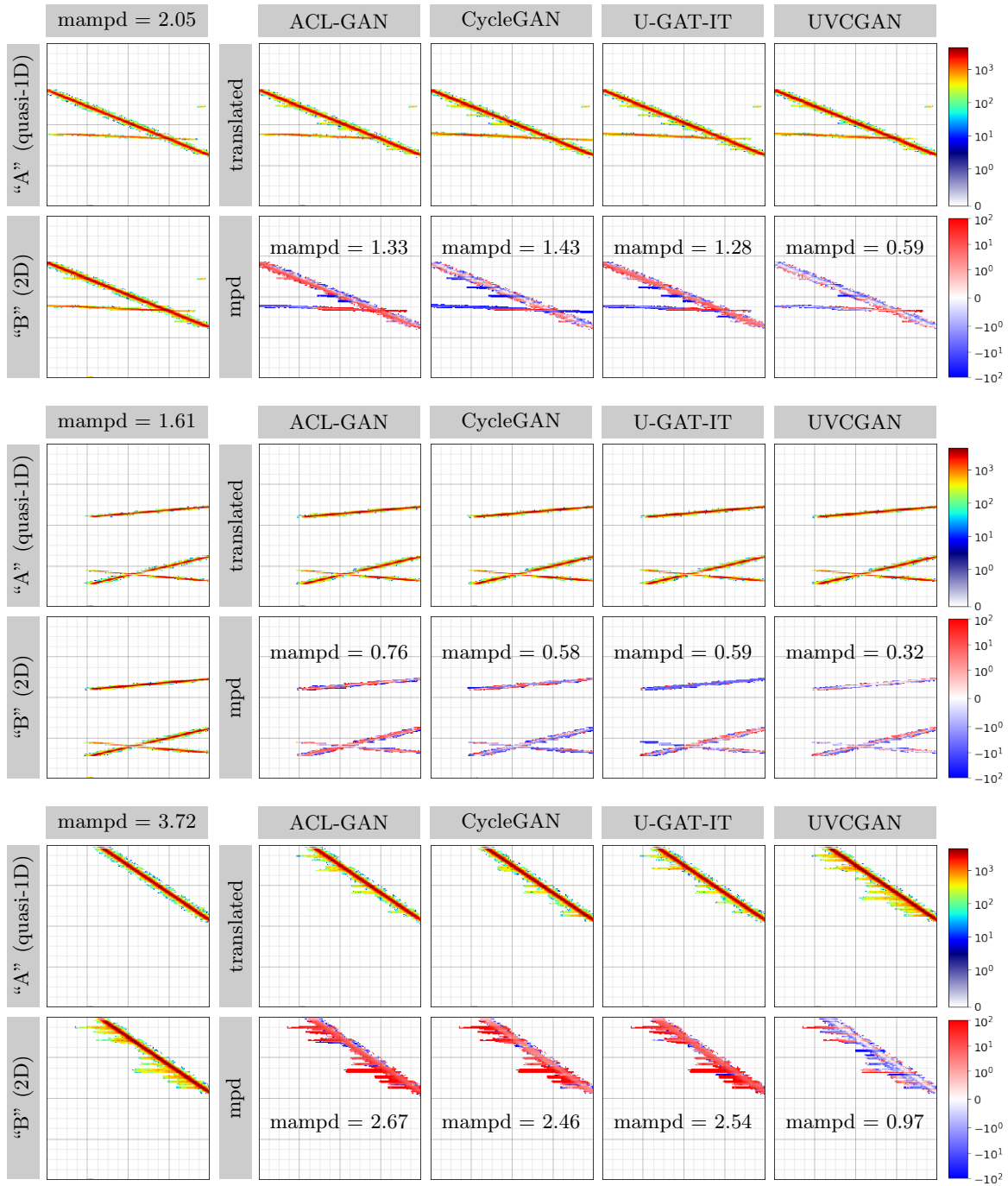


Figure B1: **Signal processing results for three events with pixel-wise modified percent difference.** For each event, on the first column, we show the signal processing results for a tile in domain “A” and its counterpart in domain “B”. In the remaining four columns, we show the signal processing results of the translations produced by the four algorithms (top) and pixel-wise modified percentage difference (bottom).