Stressed GANs snag desserts a.k.a Spotting Symmetry Violation with Symmetric Functions

Anonymous authors
Paper under double-blind review

Abstract

Is parity violated at the LHC in some way that no-one has anticipated? This paper is intended to illustrate a simple way one could try to answer that question without relying on theoretical predictions from specific mechanisms of parity violation. Alternatively the method could be used as a calibration tool to remove parity-violating misalignments in imperfect detectors. A longer sister paper extends the concepts to general symmetries (beyond parity) and addresses issues connected to acceptance and reliability of real detectors. The work here deliberately ignores those issues and works in an idealised environment to allow greater simplicity of exposition.

1 Introduction

On the 18th June 2021, after more than a decade studying collision data from the Large Hadron Collider (LHC), the ATLAS Collaboration (The ATLAS Collaboration, 2008) submitted for publication its 1000th paper. Despite that huge body of work, not a single one of those papers (and not a single one of the papers of any other LHC collaboration then or since) has yet looked for new[1] mechanisms of parity violation.

The main reason for this omission is that there are few theoretical models which predict forms of parity violation that the LHC could see, and so without a steer from theory to guide the construction of specific analysis strategies, a view seems to have formed that progress in this area is either a waste of time or simply impossible to perform in an efficient manner.

The purpose of this paper is to dispel that myth. We aim to show that at the LHC (or at any similar machine) there is a simple testing framework which could be used to look for all new sources of parity violation within the collider's grasp. The use of the method would not require potential sources of parity violation to be known about in advance and simulated — the method processes data without additional simulation input. Not only that, we wish to make the point that the method is so general that we can illustrate its use by looking for parity violation in sources which have nothing to do with particle physics, such as human handwriting.

[1]The Standard Model of Particle Physics has been known to violate parity since the 1950s. The discovery by Wu et al. (1957) relied on the observation that beta particles emitted from certain nuclei were found to be preferentially emitted anti-parallel to the initial state nuclear spins of those nuclei – an observation which only admits explanations which violate parity. It was not necessary for the *mechanism* of parity violation to be known or understood for that discovery to have been made. Regardless of the cause, the evidence for parity violation was manifest in the results. It is signs such as those – signs which can *only* be explained away as parity violation (as opposed to being effects which may be caused by parity violation but which could, in principle, be explained away by alternative parity conserving explanations) – to which our proposal has sensitivity.

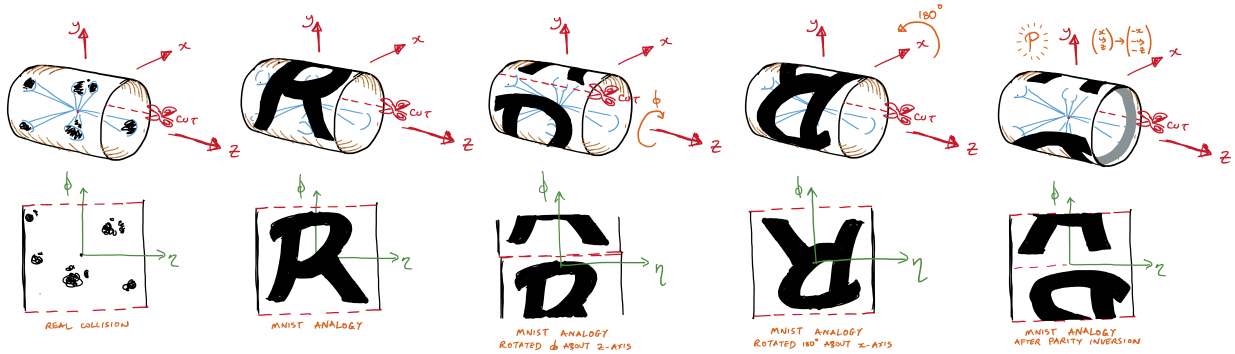


Figure 1: This figure aims to show how transformations of a 'real' three-dimensional collider calorimeter (top row) correspond to induced transformations in the unrolled two-dimensional (η, ϕ)-space in which those deposits are often represented. The first column shows deposits that would be typical of a five jet event. The second column shows an easier-to-visualise set of deposits resembling a letter '**R**'. The other three columns show how the deposits of the first '**R**' would move around if affected by (i) an arbitrary rotation about the z-axis, (ii) a rotation of 180 degrees about the x-axis and (iii) a parity inversion.

2 Method and Principles

The core principles underlying the method which this paper proposes for making for general parity-violation detections are: (i) that a parity-odd function $f(x)$ can only have an asymmetric distribution if the dataset on which it has been evaluated contains some elements x_1, x_2, ... which differ in frequency from their parity-flipped partners Px_1, Px_2, ...; and that therefore (ii) the observation of any statistically significant asymmetry in any parity-odd variable when evaluated on a real dataset is bona fide evidence for parity violation in that dataset; and so (iii) the goal of the LHC collaborations [glossing over some important real-world considerations which are mentioned later] should be to create parity-odd[2] functions which, after optimisation for asymmetry on a 'training' part of a real dataset, are then tested for asymmetry on other independent portions of the same dataset. This process is general[3], avoids all first-order use of Monte Carlo, and (like the original discovery of parity violation by Wu et al. (1957)) requires no knowledge of the parity violation mechanism to reach its conclusion.

We later choose to *illustrate* the above principle in the simplest way we can imagine: asking a parity-odd neural-net[4] (which serves simply as a flexible function approximator) to try to *maximise the separation between zero and its mean* when evaluated on appropriate training and test datasets.[5] This illustration relies on the fact that a symmetric distribution with a mean will have a mean of zero, and so statistically significant non-zeroness in the mean is a measure of asymmetry.[6]

[2]Note that it is trivial to force a neural net (or any function) by-construction to be parity-odd. You only need to evaluate it twice, once on x and once on the mirror image of x, and then subtract one answer from the other. I.e. if $g(x)$ is a general net (or function) then the function $f(x)$ defined by $f(x) \equiv g(x) - g(Px)$ is automatically parity-odd.

[3]A lengthy proof of sufficiency regarding the use of parity-odd variables is presented in a sister paper (Lester, 2021a).

[4]Whether or not LHC collaborations could benefit from using Generative Adversarial Networks (GANs) has not been seriously considered by the authors. The inclusion of GANs in the title was, however, too good an opportunity to ignore given the symmetry which it would impart to those words. We beg forgiveness from any readers who might have expected to hear a lot about them!

[5]As will be seen later, this is a mean which has been suitably normalised so as to prevent the network gaining by (say) simply multiplying the net final output by a number greater than one.

[6]Note that our decision to use a neural-net rather than some other function approximator, or our decision to optimise for deviations of the mean of that parity-odd variable from zero rather than to optimise for some fancier loss function, are all unashamedly likely to be non-optimal. The sole aim of the paper is not to show which fancy loss function or network is best, but is rather to demonstrate that general searches for parity violation can be accomplished *at all* and with relative ease.

Unusual features of this proposal

It must be emphasised how different this pattern of net use is to that most frequently seen in particle physics. There are many ways one could illustrate this, but perhaps the most easy one to highlight is that the training step is (technically) optional. If the training step were entirely omitted, the net would be in a random state whose only property was its (unavoidable) characteristic of being parity-odd. Despite the absence of training, if such a net were able to generate a statistically non-zero mean for data in the test sample, then *that non-zero mean would be evidence for parity violation*! In practice, of course, the training step is very important. Without training there is little chance that the net will actually be attending to any useful features of the data. But the point we wish to stress is that the training process exists not to make the discovery of parity-violation *possible*, but only to increase the *probability* of being able to claim discovery.

For related reasons 'overtraining' does not have the same negative overtones it would normally have.[7] The possibility of overtraining still exists but it cannot lead to an accidental mis-discovery of parity-violation as there simply is not a concept of 'right' or 'wrong' training, only of 'more efficient' or 'less efficient' training. For a discovery of parity violation all that matters is that the net manages to show a statistically significant non-zero mean on its test data. If it does, it's a good net, however it got there. The worst that can happen is that in training the net attends to entirely the wrong things — all unrelated to parity, so does not know how to generate an asymmetry in the test-set (or a mean away from zero in our illustrative example).

We note that the proposal is also unusual in that the test and training datasets are identically distributed, consist entirely of data (no Monte Carlo), and are unlabelled — and so are immediately usable in high luminosity and intense environments where simulation can be both complex and costly.

Benefits even if nature does not violate parity in a new way

No detector is perfectly aligned or calibrated or without regions where acceptance can vary. Consequently evidence of what we have called 'parity violation' cannot be pinned immediately at nature's door. On the contrary, it is quite possible that the biggest sources of parity violation could be mis-alignments and mis-calibrations. This is not really a disadvantage of the method, though. Instead it is a sign that if you can get a non-zero mean on the test-sample, then you know immediately that **either** you don't understand something about your detector, **or** you have an interesting discovery. Both are equally important, so a signal is a win-win outcome even if the source of the parity-violation is not immediately obvious.

Real world complications

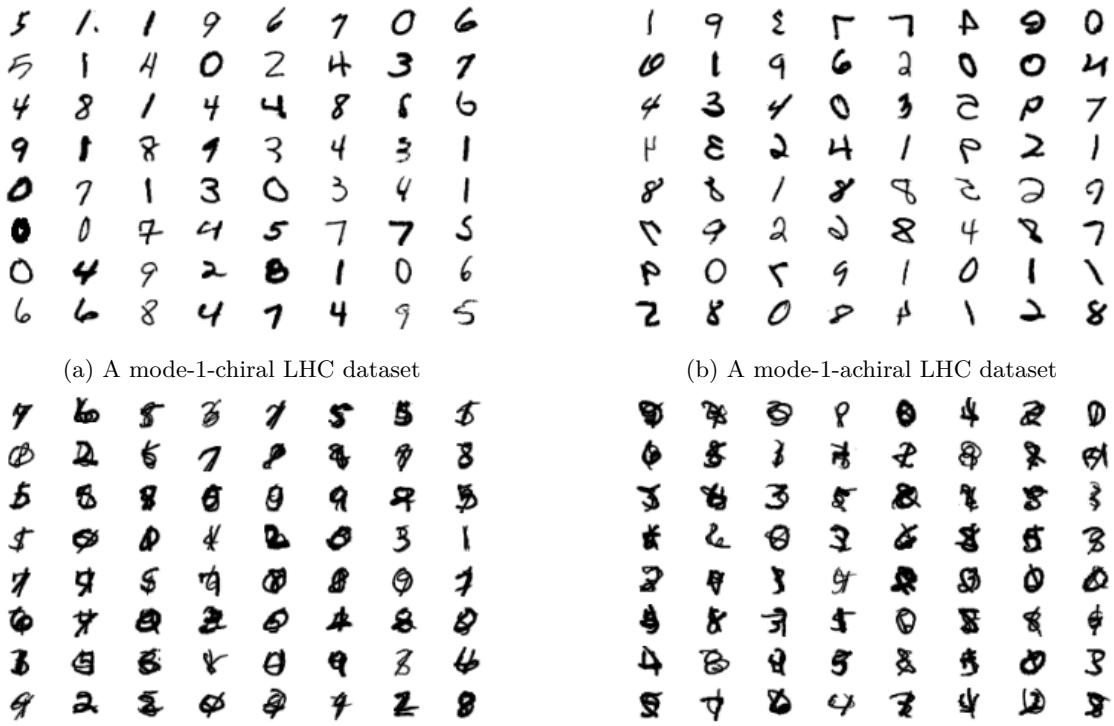
Real detectors and experiments are, of course, imperfect. Not all parts work with perfect efficiency, and acceptance does not extend uniformly into all directions. A sister paper (Tombs & Lester, 2021) addresses those absolutely critical questions, and also explains how to generalise the symmetry-violation discovery process to general symmetries (including continuous symmetries) rather than the narrow case of parity discussed herein. The current paper, however, deliberately eschews those details as the intention here is to provide a simple and pedagogically useful illustration of the key message, free from other distractions.

3 Data for an illustrative example

Visualising deposits in the 'MNIST Calorimeter;

Calorimetric deposits in idealised versions of detectors like ATLAS or CMS are commonly represented as functions on a cylinder about the beam axis as depicted in the upper left corner of Figure 1. This cylinder may be opened out into a rectangle with pseudorapidity η on one axis and azimuthal angle ϕ on the other as depicted in the lower left corner of the same figure. In this flattened view the upper and lower edges are identified with each other.

[7] A common reaction the authors have experienced when describing their proposal to newcomers is 'My gosh! Are you not proposing a mechanism for maximising overtraining problems!?'



(a) A mode-1-chiral LHC dataset

(b) A mode-1-achiral LHC dataset

(c) A mode-2-chiral LHC dataset

(d) A mode-2-achiral LHC dataset

Figure 2: Four types of LHC dataset are shown. Further descriptions can be found in the body-text.

Real calorimetric deposits are blotchy and messy, and this makes it hard for humans to picture what they might look like when rotated or mirrored. Since the goals of this paper are pedagogical, we choose to imagine a strange world in which calorimetric deposits look just like hand-written letters of the alphabet. For example, the second column of Figure 1 imagines what would happen if calorimeter deposits happened to have landed in locations which (once unfolded) would resemble the capital letter '**R**'. Making use of this visual simplification, the other columns of Figure 1 illustrate the effect that different transformations would have on the first '**R**'.

The LHC datasets

A real analysis would use real data not Monte Carlo simulations. Unfortunately, as there is no real detector that produces such energy deposits that look like letters, we must generate some toy data of our own for our pedagogical purposes. There is a ready supply of hand-written numerals in the MNIST dataset (LeCun & Cortes, 2010) so we implement a generator that draws its calorimeter deposits from there. We call these ***L****arge datasets of* ***H****andwritten* ***C****alorimeter digits* – or 'LHC datasets' for short. We produce LHC datasets in four modes as shown in Figure 2 and defined as follows:

- **mode-1-chiral** : These are normal MNIST digits. These samples are chiral because humans write hand-written digits with handed-appendages called hands. The word chiral even come from the ancient Greek word for hand: $\chi\epsilon\iota\rho$;

- **mode-1-achiral** : These are MNIST digits which have been reversed left-to-right with probability 50%; The symmetrisation process makes these samples achiral.

- **mode-2-chiral** : These are random pairs of normal MNIST digits superimposed on each other. They inherit their chirality from their constituent digits.

- **mode-2-achiral** : These are random pairs of normal MNIST superimposed on each other with one member of the pair forwards and one reversed. The symmetrisation process makes these samples achiral.

Mode-1-achiral acts as a control for mode-1-chiral, and mode-2-achiral acts as a control for mode-2-chiral. The only purpose of mode-2 is to illustrate that the analysis is not dependent on there being only a small number of 'types' of deposit — in this case ten as there are ten sorts of digits. Both types of Mode-2 therefore act as a form of control against the possibility that mode-1 is too simple to be a good proxy for real and highly variable calorimeter data.

Removing remaining differences between real calorimeter data and MNIST derived calorimeter analogues

MNIST digits are 'oriented' because the humans who originally generated them were instructed to write them in boxes the conventional way up.

Real jet data, in contrast, is not oriented: if energy deposits were to appear in a configuration resembling a capital '**R**' then the deposits would be as likely to arrive in any of the configurations shown in columns two, three and four of Figure 1, provided that the collider has symmetric beams and no reason to choose any ϕ-orientation more than any other.

To avoid the MNIST-based dataset having an unfair handle (orientability) not possessed by real jet data, this discrepancy needs to be removed. One way of removing it would be to augment our dataset by adding y-translations of uniformly random magnitude to every image in an LHC dataset (being sure to wrap pixels lost from the top back onto the bottom) before then giving every image a 180 degree rotation with probability 50%.

Although one could augment the dataset in that way, it is more efficient to simply make the neural net invariant with respect to those transformations. Doing so is as easy as making the net parity-odd, the only difference being that one seeks symmetry rather than antisymmetry and so may combine sub-nets with symmetric operations (like maxpooling or summing) rather than antisymmetric operations (like subtractions).

4 An example parity-odd axisymmetric network

Recall that the necessary axisymmetry comes from establishing (i) an invariance with respect to y-translations of 28x28-pixel images corresponding to ϕ-invariance in the real world, and (ii) invariance with respect to 180 degree rotations of the 28x28 images corresponding to interchange of beams and isotropy.

Our final network $f(x)$ gains its parity-oddness and one-half of its axisymmetry (the part requiring invariance with respect to 180 degree rotations) by being defined in terms of a sub-network $g(x)$ as follows:

$$f(x) = g(x) + g(R_{180}x) - g(Px) - g(R_{180}Px). \tag{1}$$

The sub-network $g(x)$ (implemented in pytorch by Paszke et al. (2019)) therefore only needs to respect the ϕ-rotation (or y-translation) invariance. The $g(x)$ sub-network starts by having two 2D-convolutional network layers (one with kernel size 4×4 in $\eta \times \phi$, and the other with size 14×6 in $\eta \times \phi$, always stride 1) to provided basic image processing of the 28x28 pixel images fed to it. These two layers bring the number of features up to 32 and then to 64. They are then followed by a max-pooling over the ϕ direction to finally achieve the desired ϕ-invariance. $g(x)$ then concludes with two fully connected layers reducing 640 features to 128 and thence to a single feature so that $f(x)$ can then make used of $g(x)$ in the manner outlined in (1). To describe the structure of our network, we annotate its printout (as a pytorch Module)as follows:

```
Net(
  # phi-cyclic pad
  (conv1): Conv2d(1, 32, kernel_size=(4, 4), stride=(1, 1))
```

```
  # relu
  # phi-cyclic pad
  (conv2): Conv2d(32, 64, kernel_size=(14, 6), stride=(1, 1))
  # relu
  (max_pool): MaxPool2d(kernel_size=(28, 2), stride=(28, 2), ...)
  # flatten
  (fc1): Linear(in_features=640, out_features=128, bias=True)
  # relu
  (fc2): Linear(in_features=128, out_features=1, bias=True)
)
```

The network is trained on images provided in batches. Each batch

$$B = \{x_1, x_2, \ldots, x_{|B|}\}$$

contains $|B| = 64$ images. For each batch the mean net score μ_B is calculated:

$$\mu_B = \frac{1}{|B|} \sum_{x \in B} f(x).$$

The standard deviation of the scores of the images in the batch is also recorded as σ_B. The net is trained to maximise μ_B/σ_B. This process encourages the net to do its best to give every image a common constant positive value, and to eschew negative values if possible. The net is trained to maximise μ_B/σ_B rather than μ_B alone since the former (unlike the latter) cannot be trivially made larger by replacing $f(x)$ with $\lambda f(x)$ for some constant $\lambda > 1$. The complete code for the network may be found in Lester (2021b).

5 Results

The preponderance of green in plots (a) and (c) within Figure 3 shows that the networks trained on the chiral modes are able to label the majority of their inputs with positive scores, meaning that they have identified that these datasets are chiral (i.e. violate parity). In contrast the nets have not managed to achieve this for the achiral controls ((b) and (d) in the same figure), despite having been trained on achiral control data. It is notable that the networks are not flummoxed by the differences between mode-1 and mode-2.

Mode	ρ^+_{chiral}	ρ^+_{achiral}
mode–1	90±1 %	51±2 %
mode–2	89±4 %	42±6 %

Table 1: When tested on groups of 1024 images in the relevant test dataset, the above numbers show the typical fraction ρ^+ of images per batch which were assigned a positive value by trained networks. As desired, the chiral LHC datasets are flagged as chiral at a statistically significant level, while those LHC datasets which are achiral are not, with their fractions remaining consistent with the 50% expected by chance alone.

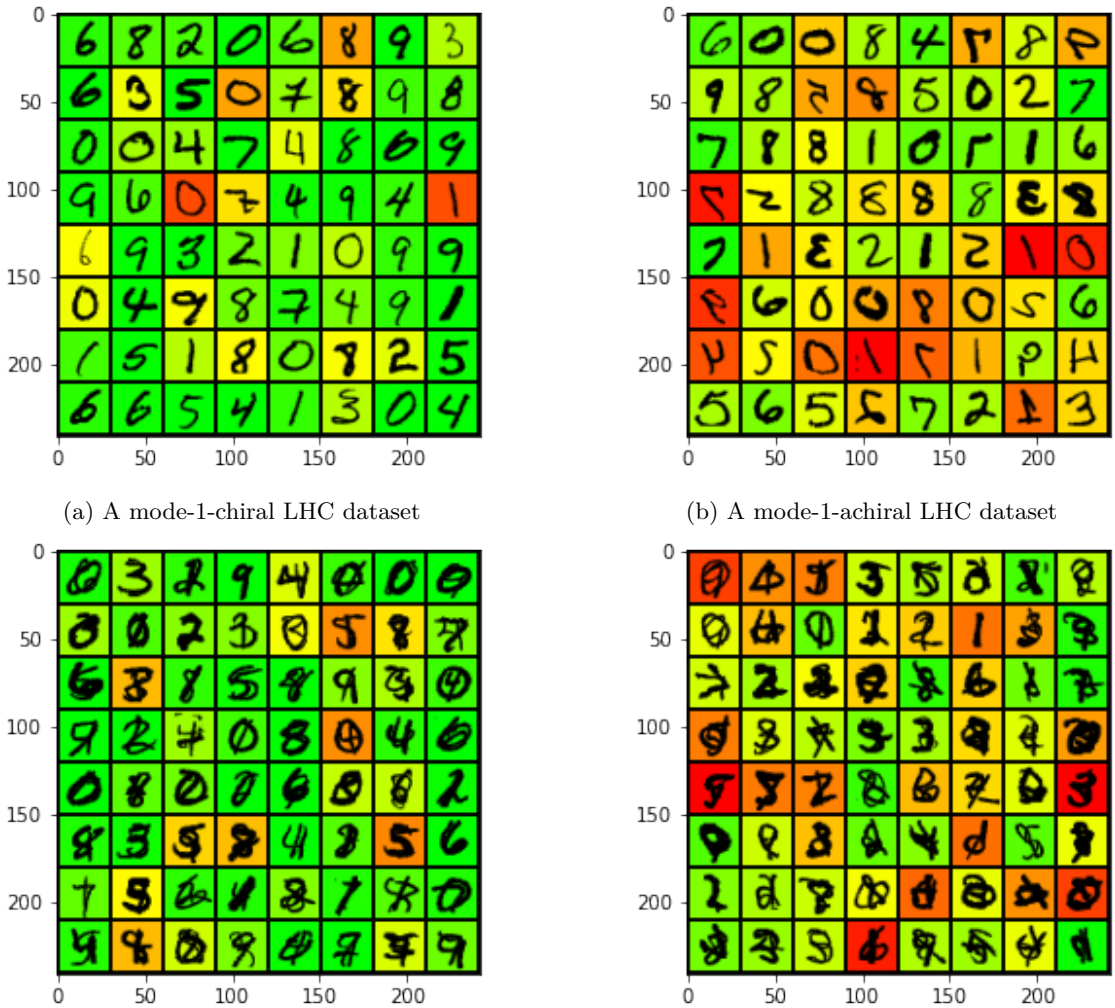
Quantitative results are presented in Table 1 and confirm that parity violation is discovered where it should be, and is not discovered where it should not be.

Figure 4 demonstrates that our network function is indeed parity-odd and axisymmetric as desired.

6 Discussion.

The strategy has achieved its goals. What is perhaps more interesting to reflect on is what this buys in comparison to other methods for analysing multi-jet events for evidence of parity violation in colliders (such as the Large Hadron Collider) without polarised beams.

Only one work (Lester et al., 2022) has previously tried to tackle this task in a systematic way. To give a chastening example: when that work considered collisions of the form $S_{ab \to jjj+X}$ (three important jets,



(a) A mode-1-chiral LHC dataset

(b) A mode-1-achiral LHC dataset

(c) A mode-2-chiral LHC dataset

(d) A mode-2-achiral LHC dataset

Figure 3: Four types of LHC dataset from the test dataset are shown after having been coloured according to the output of the net. Bright green represents increasingly positive numbers. Shades of increasingly vibrant red represent increasingly negative numbers. Yellow represents zero.

and perhaps some other stuff X in the final state, resulting from the collision of an 'a' with a 'b') it was found that 19-dimensional continuous ideal vector parity (to use the nomenclature of Lester (2021a)) would, in principle, detect any parity-violating signature. However: (i) over 60 pages of algebra were necessary to generate that result, and (ii) there seems little prospect of generalising that result to four final-state particles (due to factorial increases in computation complexity) let alone to n final-state particles.

Why do we mention these limitations of Lester et al. (2022)? The reason is that the present work can be considered to be doing a very similar job to Lester et al. (2022) but for events with 'up to' 784 = 28x28 jets in the final state, yet in a way that scales much better.[8]

[8]The intensity of each pixel can, in principle, represent a very narrow jet, and so an image with (say) only three black pixels could represent a three-jet event. Moreover, the permutation symmetry modelled between jets in Lester et al. (2022) carries over to our images, since the pixels to not somehow list jets in a particular order. The one major difference between Lester et al. (2022) and the present work is, however, that Lester et al. (2022) provides guarantees of ideality (see definition of this term in Lester (2021a)) which the present work never can.

Figure 4: Here the same 64 letters from a mode-1-chiral LHC dataset are evaluated by the network in six different ways by the same trained network. In the first pair of rows they are evaluated without any changes at all, and most are coloured green as expected. In the next pair of rows they are re-evaluated but after a shift (or roll) down by 10 pixels. In the next two rows the transformation prior to evaluation is instead a 180 degree rotation. The key take home message from these first six rows is that all of these transformations do not affect the network score, and so the network is axisymmetric as desired. The lower six rows repeat the upper six row, but after an initial left-right parity flip. In this case we see that there is (as desired) an immediate flip in the output from positive to negative (or in rare cases vice versa) because the network is, by construction, parity-odd.

7 Conclusion

This paper has proposed a framework for conducting tests of symmetry violation at the LHC. A key feature of our proposed methodology is that it can be applied to recorded data: a discovery would be made by comparing collision events to themselves rather than to theoretical predictions (of which there are few). The proposal therefore plugs a gap in the literature by making discoveries possible which hitherto could easily have been missed on account of the lack of theoretical models resulting in a lack of dedicated searches. This is important: the LHC is exploring centre of mass energies which are orders of magnitude higher than those in which parity violation was first discovered, and there is therefore no knowing what mysterious signals could be hiding in LHC datasets created in the last decade. Important signals could be hiding in plain sight.

This paper is intended to serve a primarily pedagogical and awareness-raising purposes. It is also intended to approachable to readers outside of particle physics. Accordingly it was decided to demonstrate the viability of proposed methodology by applying it to a toy problem wherein the hidden parity violating signature was the handedness of human handwriting. This is a simple target symmetry (parity) in an idealised detector possessing equally simple 'unimportant' intrinsic symmetries (axisymmetry and beam interchange symmetry) which are reminiscent of symmetries found in detectors at particle physics colliders.

A more complex sister paper (Tombs & Lester, 2021) describes how the framework presented here could easily be extended to consider continuous symmetries and shows how to deal with important issues (notably acceptance and inefficiency) which were ignored here for reasons of accessibility.

8 Acknowledgements

The authors greatly acknowledge discussions with Ben Nachman related to the present work and to Collins et al. (2018), and with Shikma Bressler related to the present work and to Volkovich et al. (2022). Useful discussions with Radha Mastandrea, Daniel Noel and Vidhi Lalchand are all very much appreciated, as has been the patience of Tom Gillam in assisting with numerous questions about pytorch.

References

Jack Collins, Kiel Howe, and Benjamin Nachman. Anomaly detection for resonant new physics with machine learning. *Phys. Rev. Lett.*, 121:241803, 12 2018. doi: 10.1103/PhysRevLett.121.241803. URL https://link.aps.org/doi/10.1103/PhysRevLett.121.241803.

Yann LeCun and Corinna Cortes. MNIST handwritten digit database. 2010. URL http://yann.lecun.com/exdb/mnist/.

Christopher G. Lester. Chiral measurements. 11 2021a. doi: 10.48550/arXiv.2111.00623. URL https://doi.org/10.48550/arXiv.2111.00623.

Christopher G. Lester. https://github.com/kesterlester/ColabForParityPaper3, 11 2021b. URL https://github.com/kesterlester/ColabForParityPaper3.

Christopher G. Lester, Ward Haddadin, and Ben Gripaios. Lorentz and permutation invariants of particles III: constraining non-standard sources of parity violation. *International Journal of Modern Physics*, 37, 8 2022. doi: 10.1142/S0217751X22500932. URL https://doi.org/10.1142/S0217751X22500932.

Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, Alban Desmaison, Andreas Kopf, Edward Yang, Zachary DeVito, Martin Raison, Alykhan Tejani, Sasank Chilamkurthy, Benoit Steiner, Lu Fang, Junjie Bai, and Soumith Chintala. Pytorch: An imperative style, high-performance deep learning library. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett (eds.), *Advances in Neural Information Processing Systems 32*, pp. 8024–8035. Curran Associates, Inc., 2019. URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.

The ATLAS Collaboration. The ATLAS Experiment at the CERN Large Hadron Collider. *JINST*, 3:S08003, 2008. doi: 10.1088/1748-0221/3/08/S08003. URL https://doi.org/10.1088/1748-0221/3/08/S08003.

Rupert Tombs and Christopher G. Lester. A method to challenge symmetries in data with self-supervised learning. 2021. doi: 10.48550/arXiv.2111.05442. URL https://doi.org/10.48550/arXiv.2111.05442.

Sergey Volkovich, Federico De Vito Halevy, and Shikma Bressler. The Data-Directed Paradigm for BSM searches. *Eur. Phys. J. C*, 82(3):265, 2022. doi: 10.1140/epjc/s10052-022-10215-1. URL https://doi.org/10.1140/epjc/s10052-022-10215-1.

C. S. Wu, E. Ambler, R. W. Hayward, D. D. Hoppes, and R. P. Hudson. Experimental test of parity conservation in beta decay. *Phys. Rev.*, 105:1413–1415, Feb 1957. doi: 10.1103/PhysRev.105.1413. URL https://link.aps.org/doi/10.1103/PhysRev.105.1413.