SYMMETRY-AUGMENTED REPRESENTATION FOR TIME SERIES

Anonymous authors
Paper under double-blind review

ABSTRACT

We examine the hypothesis that the concept of symmetry augmentation is fundamentally linked to learning. Our focus in this study is on the augmentation of symmetry embedded in 1-dimensional time series (1D-TS). Motivated by the duality between 1D-TS and networks, we augment the symmetry by converting 1D-TS into three 2-dimensional representations: temporal correlation (GAF), transition dynamics (MTF), and recurrent events (RP). This conversion does not require *a priori* knowledge of the types of symmetries hidden in the 1D-TS. We then exploit the equivariance property of CNNs to learn the hidden symmetries in the augmented 2-dimensional data. We show that such conversion only increases the amount of symmetry, which may lead to more efficient learning. Specifically, we prove that a direct sum based augmentation will never decrease the amount of symmetry. We also attempt to measure the amount of symmetry in the original 1D-TS and augmented representations using the notion of persistent homology, which reveals symmetry increase after augmentation quantitatively. We present empirical studies to confirm our findings using two cases: reinforcement learning for financial portfolio management and classification with the CBF data set. Both cases demonstrate significant improvements in learning efficiency.

1 INTRODUCTION

One-dimensional time series (1D-TS) representations do not expose the co-occurent events and the latent states of the data (Prado, 1998) in a way that machine learning can easily recognize. For financial data, there are patterns at various scales that can be learned to improve performance. For ECG data, we want to identify the fundamental physical heart problems that are represented indirectly in the ECG data. Our approach is to find richer alternate representations to improve time series pattern identification for more efficient and effective machine learning methods by transforming 1D-TS data into 2-dimensional feature sets. This will provide a richer representation of the data and opens up the opportunity to apply a different set of algorithms adapted from computer vision such as Convolution Neural Networks (CNNs).

Physics defines a symmetry of a physical system as a physical or mathematical feature of the system (observed or intrinsic) that is preserved under some transformation. According to Noether's first theorem (Noether & Tavel, 2018; Byers, 1996), a symmetry preserving transformation results in the invariance of some quantity. According to the Nobel Prize-winning physicist *Philip Warren Anderson* "by symmetry, we mean the existence of different viewpoints from which the system appears the same. It is only slightly overstating the case to say that physics is the study of symmetry." (Anderson, 1972). It is important to understand that symmetries are a super-set of correlations and a more general concept. Correlation captures dependencies which are an extrinsic property of dynamical systems, while symmetry is related to transformations that leave a dynamical system invariant. Detection of an invariant property (e.g. translational) indicates that a related intrinsic property is conserved (e.g. momentum). Our paper aims to track the intrinsic (hidden) features of dynamical systems by leveraging the invariance property of CNNs and not the extrinsic features such as correlations which can simply be computed separately and taken as a feature into any ML system or by using recurrent neural networks (RNNs).

We identified financial portfolio management as a case study where we have multiple 1D-TS as input and we expect that more symmetries will emerge through augmentation. (Fischer, 2018; Aboussalah

& Lee, 2020; Aboussalah et al., 2020) summarize the state-of-the-art RL models in finance. Based on the work of Aboussalah et al. (2020), which indicated that policy-based RL works better than value-based RL, we studied two policy-based RL algorithms, namely deterministic policy gradient (DPG) and proximal policy optimization (PPO), to assess the benefits of our time series augmentation to RL. To explore the generality of our approach, we selected the CBF time series classification (Saito, 1994; Kadous, 1999) as another case study. We studied convolution neural networks with fully connected layers to assess the benefits of our symmetry augmentation on a traditional time series classification problem.

In summary, the contributions of our work are as follows:

- We provide a mathematical analysis of symmetry and show how augmentation can benefit learning. In particular, we prove that a direct sum based augmentation will never decrease the amount of symmetry.
- We leverage the translational invariance property of CNNs to learn hidden *spatio-temporal* symmetries created by our augmented representation that improves the learning.
- We develop an empirical measure of the amount of symmetry in our time series augmentations using computational *Homology*.
- We present empirical studies to confirm our findings.

2 Related Work

Learning Symmetry from CNN. Cohen & Welling (2016) presented Group Equivariant Convolutional Neural Networks (G-CNNs), a generalization of convolutional neural networks (CNNs) that decreases sample complexity by making the most of symmetries. G-CNNs make use of G-convolutions, a new kind of layer that enjoys a significantly higher degree of weight sharing compared to common convolution layers. G-convolutions typically increase the expressive capacity of the network, but the number of parameters is not increased. They used translation, rotation, and reflection symmetries to achieve exceptional results on rotated MNIST and CIFAR10. Cohen & Welling (2016) addresses the concept of groups of symmetries using CNNs in supervised learning for static data while our work uses dynamic data.

Learning Symmetry from Interactions. Higgins et al. (2018) proposed Symmetry-Based Disentangled Representation Learning (SBDRL), where they discuss how translation, rotation, color change symmetries can be used to disentangle features and therefore make them learnable. Caselles-Dupré et al. (2019) expanded on their work and proposed linear and nonlinear SB-disentangled representations learned through interactions with the environment. This work points to the applicability of RL as used in our work.

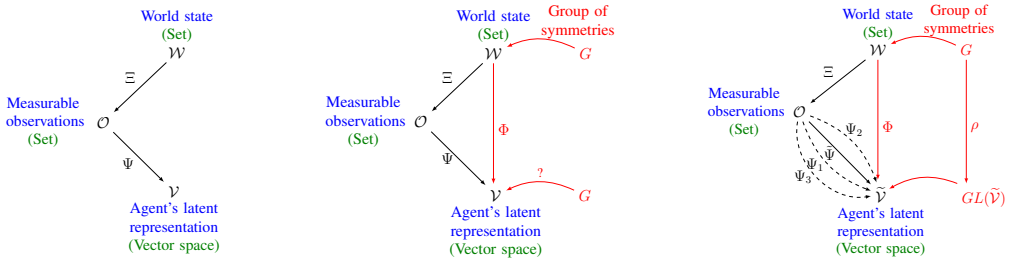
Augmentation for RL. Laskin et al. (2020) presented Reinforcement Learning with Augmented Data (RAD) module which can augment most RL algorithms. They have demonstrated that augmentations such as random translate, random convolutions, crop, patch cutout, amplitude scale, and color jitter can enable simple RL algorithms to outperform complex advanced methods on standard benchmarks. Kostrikov et al. (2020) presented a data augmentation method that can be applied to conventional model-free reinforcement learning (RL) algorithms, enabling learning directly from pixels minus the need for pre-training or auxiliary losses. The inclusion of this augmentation method, improves performance substantially, enabling a Soft Actor-Critic (SAC) agent to reach advanced functioning capability on the DeepMind control suite, outperforming model-based methods and contrastive learning. Laskin et al. (2020) and Kostrikov et al. (2020) show the benefit of RL-augmentation using CNNs for static data but do not connect to symmetries and do not handle dynamic data.

Our work uniquely connects time series, symmetries, CNNs, RL and augmentation, in realistic noisy dynamical environments using a rigorous mathematical foundation.

3 HIGH DIMENSIONAL STATE SPACE AUGMENTATION

3.1 MAPPING SYMMETRIES FROM THE WORLD STATE TO THE AGENT

Let's consider a generator process $\Xi : \mathcal{W} \rightarrow \mathcal{O}$ (Fig. 1a). $\mathcal{W}$, the world state, includes all the characteristics of the physical system and $\mathcal{O}$ is what we measure from the world state (*e.g.* 1D-TS), therefore depends on the measure being used. Let's also consider an embedding process $\Psi : \mathcal{O} \rightarrow \mathcal{V}$ from observations $\mathcal{O}$ to an agent's latent representation $\mathcal{V}$.



(a) Surjective mapping from the world state to the latent space of the agent.

(b) Can we transfer some of the world state symmetries into the latent space of the agent?

(c) Augmented representation $\tilde{\Psi}$ enables the exploitation of symmetries.

Figure 1: Mapping symmetries from the world state to the agent.

By construction, the mapping $\Psi \circ \Xi$ in (Fig. 1a) is surjective, which means that a potential loss of information such as symmetry of the system is possible when moving from $\mathcal{W}$ to $\mathcal{V}$. The recurrence Poincaré theorem guarantees the existence of such symmetries in dynamical systems (Katok & Hasselblatt, 1995).

The set of all symmetries of any object forms a group. A representation of a group G on a finite dimensional vector space $\mathcal{V}$ is sometimes referred to as the homomorphism of groups from the group of symmetries of the world state G to $GL(\mathcal{V})$, which is the group of symmetries of the latent space $\mathcal{V}$. We aim to find a corresponding action "$\blacksquare$": $G \times \mathcal{V} \longrightarrow \mathcal{V}$ so that the symmetry structure of $\mathcal{W}$ is *transferred* into $\mathcal{V}$. However in practice, for a given $\Phi : \mathcal{W} \rightarrow \mathcal{V}$, there is no guarantee that we can find a compatible action "$\blacksquare$" $G \times \mathcal{V} \rightarrow \mathcal{V}$ (Fig. 1b) because we do not know a priori the hidden symmetries. The only remaining solution to transfer as many symmetries from the world state $\mathcal{W}$ to the latent space $\mathcal{V}$ would be to *augment* the symmetries of $\mathcal{O}$ through a new mapping $\tilde{\Psi}$ (Fig. 1c). In order to *augment* the amount of symmetries of $\mathcal{O}$, let's assume we are given some mappings that we call augmentations Ψ_i with $i \in \{1, \ldots, N\}$ and N the total number of augmentations. Whether or not these individual augmentations carry symmetries, our main focus is on the way to combine them into a single representation scheme so that the resulting representation will be only better or in the worst case the same as the original 1D-TS representation in terms of the amount of symmetry information. While others have utilized augmentations in ML, we are the first to provide a mathematical explanation that identifies symmetries as a significant factor behind the benefits of augmentations in ML.

Theorem 3.1 (Monotonicity of direct sum based representation) *Let G be the group of symmetries acting on the world state $\mathcal{W}$ and A a discrete subgroup of G. For any $i \in \{1, \ldots, n\}$, let's consider Ψ_i to be the map (augmentation) that takes $x \in \mathcal{O}$ and returns its image in its corresponding augmented vector space $\mathcal{V}_i$. Given a set of Ψ_i, the resulting mapping $\widetilde{\Psi}_n$:*

$$\widetilde{\Psi}_n \doteq \bigoplus_{i=1}^{n} \Psi_i : \mathcal{O} \longrightarrow \mathcal{O} \bigoplus_{i=1}^{n} \mathcal{V}_i \cong \widetilde{\mathcal{V}}_n, \tag{1}$$

associated with its corresponding representation

$$\tilde{\rho}_n : A \subset G \longrightarrow GL(\widetilde{\mathcal{V}}_n), \tag{2}$$

is non-decreasing in total symmetry as n increases independent of the quality of the Ψ_i augmentations being used.

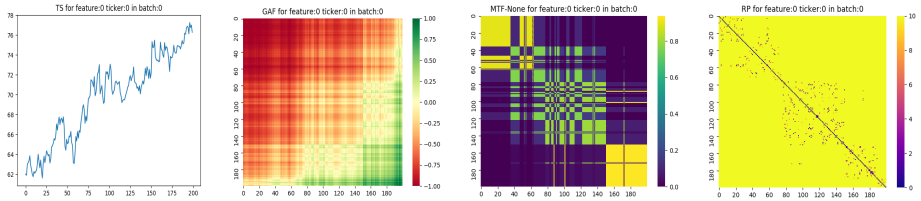
The theorem guarantees that we can create as many augmentations as we want and, even if some of them do not expose symmetries, we can safely combine them to achieve the benefits of those that do expose symmetries. The proof is provided in Appendix B.

Corollary 3.1 (Preservation of symmetries) *The group of symmetries of $\widetilde{\mathcal{V}}_n$ is larger than the group of symmetries of $\mathcal{O}$ and we write: $GL(\mathcal{O}) \subseteq GL(\widetilde{\mathcal{V}}_n)$.*

The implication of this corollary is that even if the augmentation does not expose more symmetries, ML performance would not be worse than with non-augmented 1D-TS.

3.2 THREE AUGMENTATIONS USING DUALITY BETWEEN TIME SERIES AND GRAPHS

To test our framework, we have developed an approach for transforming 1D-TS into multiple 2D feature sets using the duality between time series and graphs (Campanharo et al., 2011). Our literature search found three distinct TS-to-Network maps, namely the Recurrence Plots (RP) (Marwan et al., 2002), the Gramian Angular Field (GAF), and the Markov Transition Field (MTF) (Wang & Oates, 2015). We choose to use all three maps because they show different topological properties and different dynamic regimes. Therefore, we are testing our framework with $n = 3$ in Theorem 3.1. Fig. 2a shows a 1D-TS from our financial portfolio management case study and Fig. 2b, 2c, 2d show three augmentations: GAF, MTF, and RP.



(a) 1D-TS (AMT) (b) GAF (AMT) (c) MTF (AMT) (d) RP (AMT)

Figure 2: Three augmentations corresponding to the American Tower Corp. (AMT) 1D-TS.

GAF is the graph obtained from a 1D-TS in which the vertices correspond to the value of the TS at each time point, and the arcs between different vertices represent the temporal correlation between them (Wang & Oates, 2015). It uses the Gram matrix to compute the linear dependence of a set of vectors in the inner product space. Since the inner product in the 1D-TS space does not allow us to separate the signal from Gaussian noise, we add a new dimension in order to take advantage of *geometric* relationships. Wang & Oates (2015) define a polar coordinate system where the value of the TS is the angle and time is the radius. The advantages of this mapping are that the encoding is bijective, and it preserves temporal dependency through the radius coordinate. The transformation to polar coordinates allows us to compute the Gram matrix (GAF image). The GAF image (Fig. 2b) represents temporal correlations that are not exposed in the 1D-TS representation. The pixels in the main diagonal from top left to bottom right represent the values of the 1D-TS, while the non-diagonal elements represent the temporal correlation between each state in polar coordinates. The color of the pixels of the main diagonal indicates a general upward trend like the 1D-TS representation. The off-diagonal red areas represent low correlation, the green areas represent high correlation, and the yellow areas represent uncorrelated states.

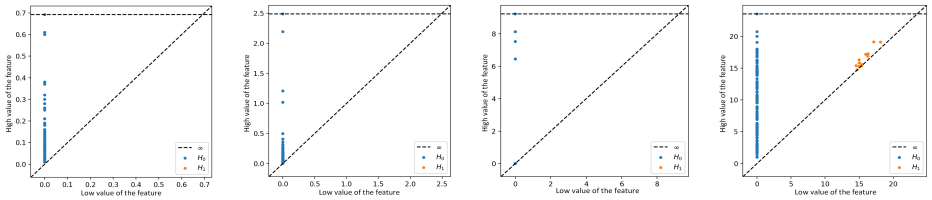
MTF extends the framework proposed by (Campanharo et al., 2011) for encoding dynamical transition statistics by depicting the Markov transition probabilities sequentially to conserve information in the time domain. The chief idea of MTF is to construct the Markov matrix using quantile bins after the discretization of the 1D-TS and encoding the dynamic transition probabilities in a quasi-Gram adjacency matrix (Wang & Oates, 2015). This matrix (MTF image) encodes the dynamic transition probability by counting transitions between adjoining quantile bins in the same way as a first-order Markov chain along the time axis. The MTF image (Fig. 2c) represents the transition probabilities for a discretized time series. Yellow and green areas indicate a high probability of transitions between two different states of the 1D-TS. Blue represents low transition probability. The yellow and green areas are mainly concentrated along the diagonal which corresponds to the self-transition probabilities of the 1D-TS. Having a non-zero probability transition of landing in a state that is not

exactly on the main diagonal allows us to explore, in a natural way, other dynamic transitions which could help improve generalization.

RP (Fig. 2d) is a plot showing, for each moment i in time, the times at which a phase space trajectory visits roughly the same area in the phase space. We track the recurrence of similar events by recording the states that remain approximately ϵ-close to each other. Once a state enters a ball of ϵ-radius centred on a given state, we say that we have detected a recurrence and we mark it with a point as shown in Fig. 2d. The image is mostly yellow because there are only a few sparsely located recurring events represented by the non yellow dots due to the non stationary character of the 1D-TS.

3.3 Quantifying Symmetry with Persistent Homology

Homology is the study of topological invariant properties and symmetry is associated with invariant properties of dynamical systems. Therefore, we looked to homology for tools to help characterize symmetries. Persistent homology (PH) (Afra & Gunnar, 2004; Gunnar, 2009; Edelsbrunner & Harer, 2010) characterized by persistence diagrams (PDs) computes topological features that are more likely to represent true intrinsic characteristics of the relationships between the data points. PD captures the persistence of these homology-based features across multiple scales. We generated PDs of the 2D augmentations and compared them to the PD generated from the 1D-TS (Fig. 3a, 3b, 3c, 3d). The x-axis represents the low value of a feature and the y-axis its high value. The range of each axis is bounded by the range of the 1D-TS values. Each dot represents a feature and higher values identify more topologically invariant features (persistent symmetries). Features that are less persistent occur close to the diagonal and are typically considered to be noise, while the most persistent features occur at a greater distance from the diagonal.



(a) PD 1D-TS (AMT) (b) PD GAF (AMT) (c) PD MTF (AMT) (d) PD RP (AMT)

Figure 3: Persistence diagrams (PDs) of the 1D-TS and the three augmentations corresponding to the American Tower Corp. (AMT).

Table 1: Feature invariance

Stock	Persistence of features			
	1D	GAF	MTF	RP
AMT	0.610	2.192	8.135	20.736
AXP	0.239	1.877	6.387	20.248
BA	0.639	3.334	7.528	18.708
CVX	0.314	2.125	5.599	22.248
JNJ	0.659	4.423	7.754	18.220
KO	1.080	1.567	9.167	22.226
MCD	1.049	1.330	7.768	17.691
MSFT	1.439	2.117	4.284	19.416
T	0.659	3.216	4.237	19.026
WMT	0.599	1.72	8.915	18.248
Average	**0.7287**	**2.3901**	**6.9774**	**19.6767**

All four of the PDs show the 0-dimensional (0D) homology based features ($H0$) as blue dots. In our setting, the $H0$ topological features are the connected components that correspond to connections between local minima and maxima (critical points) of the 1D-TS. The x-coordinate of the blue dots is the global minimum of the 1D-TS. Fig. 3d includes 1D homology based features ($H1$) as orange dots. The RP augmentation is a richer representation of the 1D-TS and therefore gives rise to a PD with more persistent $H0$ (blue) and $H1$ (orange) features. $H0$ is sufficient to capture the linear nature of the off-diagonal components in GAF and MTF (temporal correlation and first-order dynamical transitions) but it is not suitable for non linear topological features such as in RP (recurrent states). Therefore, $H1$ is required to represent these higher dimensional topological features built upon the $H0$ connected components. These $H1$ features capture recurrent events and their x-axis values are not constrained to the global minimum like the $H0$ features. In table 1, the first column shows the stocks from our financial portfolio management case study. The other columns show the maximum values for the PDs. The maximum PD values for the augmentations are much higher than the maximum PDs for the original 1D-TS indicating features with more persistence and therefore with more symmetries.

4 EXPERIMENTS

4.1 SYMMETRY-BASED AUGMENTATION FOR REINFORCEMENT LEARNING

In this section, we experimentally demonstrate the benefits of our methodology which can be used in conjunction with any TS-RL module. We are considering portfolio management as a case study to demonstrate the benefit of augmenting TS-RL. Our RL models are trained and tested on a portfolio consisting of ten selected stocks. To promote diversification of the portfolio, these stocks were selected from different sectors of S&P 500 so that they are uncorrelated as much as possible as shown in (Fig. 10) in appendix D.

In our experiments, the investment decisions are made daily and each element in the input signal χ_t represents the daily return. The training section covers 200 trading days ($\sim$ 11 months) and the testing section covers 10 ($\sim$ 2 weeks). We performed 5 iterations, sliding the windows two weeks each time. The plots in Fig. 5 and Fig. 6 show the cumulative return from the last sliding window iteration for training and testing. We use the translational invariance property of CNNs in order to better learn *spatio-temporal* symmetries of the underlying assets in our portfolio. We built two CNN-based RL architecture designs: CNN with and without dense layers Fig. 4.

$$
\begin{aligned}
&\max_{\theta} \sum_{t=1}^{T} \gamma^t \mathcal{R}_t \quad \textit{\#Loss} \\
&\text{s.t. } \mathcal{R}_t = \ln\Big(\langle \mathbf{a}_{t-1}, \mathbf{r}_t \rangle - c \sum_{i=1}^{m} \mid a_{i,t} - a_{i,t-1} \mid \Big) \quad \textit{\#Reward} \\
&\quad a_t = softmax(W^{(4)} h_3 + b^{(4)}) \quad \textit{\#Portfolio vector weights @ time-step t} \\
&\quad h_3 = ReLU(W^{(3)} h_2 + b^{(3)}) \quad \textit{\#Dense layer} \\
&\quad h_2 = ReLU(W^{(2)} \Phi_t + b^{(2)}) \quad \textit{\#Dense layer} \\
&\quad \Phi_t = Conv(K^{(1)} * \chi_t + b^{(1)}) \quad \textit{\#Symmetry extraction}
\end{aligned}
\tag{3}
$$

where $\{(K^{(1)}, b^{(1)}), W^{(2)}, b^{(2)}) \ldots, (W^{(4)}, b^{(4)})\}$ represent the parameters of the policy network and χ_t the input financial time series signal. In the reward function, we used the typical transaction costs $c = 0.25\%$ for portfolio trades.

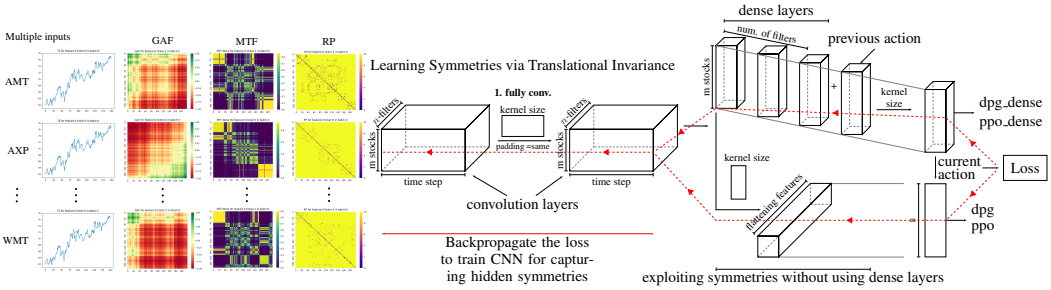


Figure 4: Deep RL architecture design handles multiple inputs and produces two RL policies (with and without dense layers).

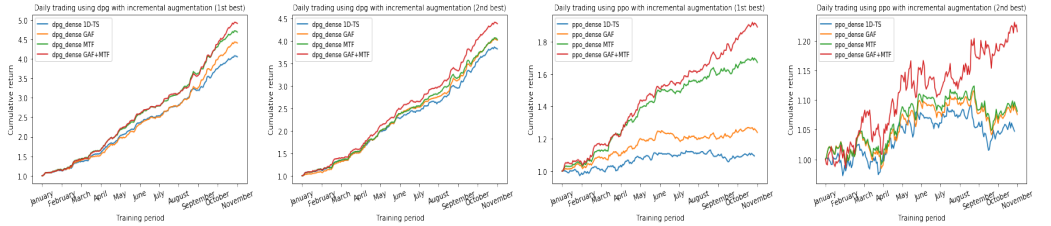


Figure 5: Sample efficiency gains due to symmetry augmentations.

Symmetry and sample efficiency. To illustrate the sample efficiency of the combination of augmentations, we show in Fig. 5 the training results using the original 1D-TS data and a set of 2D augmentations. The numerical results for the best DPG model are shown in table 2 and complete results are summarized in tables 6, 7, 8 and 9 (appendix D.5). The plots show that each augmentation

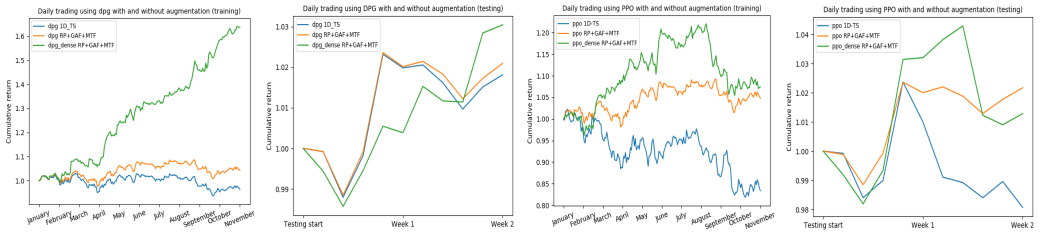


Figure 6: Testing demonstrates generalization with potential overfitting.

increases the cumulative return and the combination increases it further. Since the average value corresponding to the MTF augmentation is quite close to that of GAF+MTF, this leads us to believe that the symmetry information from MTF allowed the most symmetry transfer to the agent when the two augmentations are combined together. The maximum values reached in a single trading day are positively correlated with the cumulative returns and the average values as well. The minimum values show that it is possible for the RL agents to lose money below the initial capital of 1 unit allocated at the beginning of the training period. All augmentations demonstrate better convergence and hence improved sample efficiency.

Table 2: Summary of results of the best DPG model with various augmentations.

Agent performance	**Cumulative return**	**Average value**	**STD value**	**Max value**	**Min value**
dpg dense 1D-TS	3.0473	2.3418	0.9230	4.0716	1.0
dpg dense GAF	3.4016	2.3762	1.0052	4.4262	1.0
dpg dense MTF	3.6808	2.5853	1.1141	4.7137	1.0
dpg dense GAF+MTF	3.9046	2.6046	1.1462	4.9305	1.0

Generalization. To illustrate the generalization capacity of the augmented RL models, we show in Fig. 6 the testing results for the original data (1D-TS) and the performance of two different RL architectures (with and without dense layers) when using all three augmentations. The numerical results are summarized in tables 10, 11, 12 and 13 (appendix D.6). The plots show that when the three augmentations are used in combination, they are accompanied by an increase in the cumulative return for both training and testing. The use of dense layers results in better cumulative returns and mostly better average values. The maximum values reached in a single trading day are positively correlated with the cumulative returns and the use of dense layers. The minimum value is usually the starting value of 1 but in some cases dips lower indicating a loss of portfolio value. Both RL methods with symmetry augmentation demonstrate generalization but the dense layers method also shows potential overfitting.

4.2 Symmetry-based augmentation for supervised learning

The CBF dataset (Saito, 1994; Kadous, 1999) is a synthetic 1D-TS dataset with three classes of patterns having Cylinder, Bell, and Funnel shapes plus added noise. Each sample has 128 points and the size and the offset of the shapes vary. Example data and their augmentations are shown in Fig. 7. There are 30 training samples and 900 testing samples. The limited training data makes it challenging to generalize learning but allows us to demonstrate the benefit of symmetry augmentation. Fig. 7 shows the CNN architecture with convolution and dense layers used for training. The details of the architecture are provided in table 5 (Appendix D.4).

We conducted ten trials. Table 3 shows the average test accuracy and the persistence of H_0 features (PDs). Complete results are in table 14 (Appendix E). For this dataset, RP has the highest accuracy and the highest H_0 persistence feature score. GAF was ranked second in accuracy and in H_0 persistence feature score. The combinations significantly improved the accuracy compared to the individual augmentations. Fig.8 a-b shows the training and testing loss for 100 epochs for one trial. Fig.8 c-d shows the training and testing accuracy. The top 3 methods all include RP augmentation.

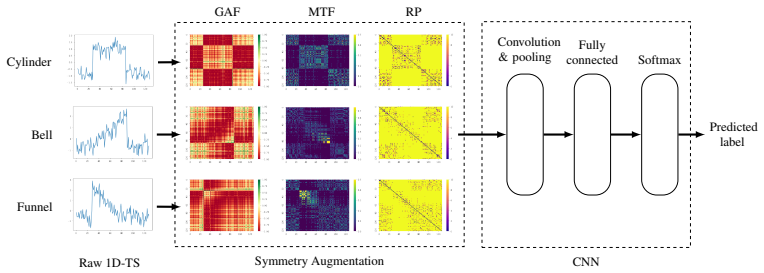
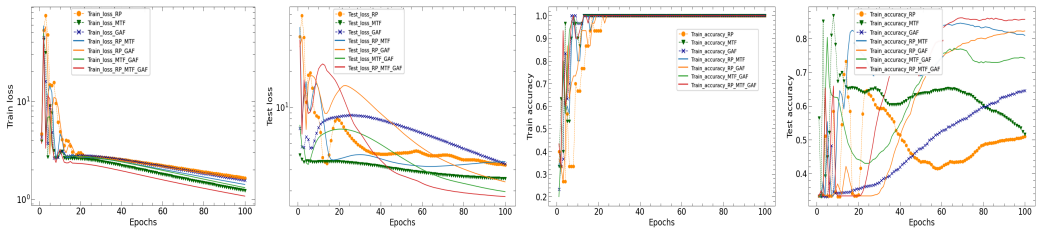


Figure 7: CNN architecture design handles CBF 1D-TS and their augmentations.

Table 3: Test accuracy with CBF symmetry incremental augmentation

Test accuracy	**RP**	**MTF**	**GAF**	**RP+MTF**	**RP+GAF**	**MTF+GAF**	**RP+MTF+GAF**
Average	0.7681	0.4686	0.5891	0.8313	0.8574	0.6282	0.9112
Persistence of H_0 features			**1D**	**GAF**	**MTF**	**RP**	
Average			0.2561	4.1304	1.5723	20.2854	



(a) Train loss (b) Test loss (c) Train accuracy (d) Test accurary

Figure 8: CBF training and testing performance with symmetry augmentation.

Note that RP was the best augmentation for both datasets but GAF was second for the CBF and MTF was second for financial portfolio management. Intuitively, the state transition probability encoded in MTF is better for capturing the dynamics of financial markets while the angular temporal correlation in GAF is better for shape classification.

5 Conclusion and Future Work

We hypothesize that the concept of symmetry augmentation is fundamentally linked to learning. We augmented 1D-TS data using three methods designed to expose symmetries. We exploited the invariance property of CNNs to learn hidden symmetries in the augmented data. We proved that by using a direct sum based representation we can maintain at least the same amount of symmetry information regardless of the quality of the augmentations being used. Our results with RL and supervised learning show the benefit of symmetry augmentation. Because this approach operates on the input data, it is invariant to the choice of any ML algorithm. As future work, we would like to investigate tensor product representations that could be useful in the multidimensional-TS cases where richer and mixed symmetries could potentially emerge. We also would like to explore approaches for designing augmentations to maximize symmetry exposure. Finally, we would like to study the utility of symmetry augmentation for transfer learning.

REFERENCES

A. M. Aboussalah and C.-G. Lee. Continuous control with stacked deep dynamic recurrent reinforcement learning for portfolio optimization. *Expert Systems with Applications*, 2020.

A. M. Aboussalah, C.-G. Lee, and Ziyun Xu. What is the value of the cross-sectional approach to deep reinforcement learning? *arXiv*, 2020.

Zomorodian Afra and Carlsson Gunnar. Computing persistent homology. *Discrete & Computational Geometry*, 33(2):249–274, 2004.

P. W. Anderson. More is different. *Science*, 177:393–396, 1972.

N. Byers. E. noether's discovery of the deep connection between symmetries and conservation laws. *In Proceedings of a Symposium on the Heritage of Emmy Noether*, 1996.

A. S. L. O. Campanharo, M. I. Sirer, R. D. Malmgren, F. M. Ramos, and L. A. N. Amaral. Duality between time series and networks. *PLoS ONE*, 6(8):e23378, 2011.

Hugo Caselles-Dupré, Michael Garcia-Ortiz, and David Filliat. Symmetry-based disentangled representation learning requires interaction with environments. *33rd Conference on Neural Information Processing Systems (NeurIPS)*, 2019.

T. Cohen and M. Welling. Group equivariant convolutional networks. *Proceedings of The 33rd International Conference on Machine Learning*, 48:2990–2999, 2016.

H. Edelsbrunner and J. Harer. Computational topology: An introduction. *American Mathematical Society*, 2010.

T. G. Fischer. Reinforcement learning in financial markets - a survey. *Institute for Economics*, 2018.

Carlsson Gunnar. Topology and data. *American Mathematical Society Bulletin*, 46(2):255–308, 2009.

Irina Higgins, David Amos, David Pfau, Sebastien Racaniere, Loic Matthey, Danilo Rezende, and Alexander Lerchne. Towards a definition of disentangled representations. *arXiv preprint arXiv:1812.02230*, 2018.

M. W. Kadous. Learning comprehensible descriptions of multivariate time series. *In Proceedings of the Sixteenth International Conference on Machine Learning, ICML'99*, pp. 454–463, 1999.

A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*. Cambridge University Press, Cambridge, 1995.

Ilya Kostrikov, Denis Yarats, and Rob Fergus. Image augmentation is all you need: Regularizing deep reinforcement learning from pixels. *arXiv:2004.13649*, 2020.

M. Laskin, K. Lee, A. Stooke, L. Pinto, P. Abbeel, and A. Srinivas. Reinforcement learning with augmented data. *arXiv:2004.14990*, 2020.

N. Marwan, N. Wessel, U. Meyerfeldt, A. Schirdewan, and J. Kurths. Recurrence-plot-based measures of complexity and their application to heart-rate-variability data. *Physical Review E*, 26 (026702), 2002.

E. Noether and M. A. Tavel. Invariant variation problems (english translation of noether's theorems (1918)). *arXiv:physics/0503066*, 2018.

R. Prado. *Latent structure in non-stationary time series*. PhD thesis, Duke University, 1998.

N. Saito. Local feature extraction and its application using a library of bases. *PhD thesis, Department of Mathematics, Yale University*, 1994.

Z. Wang and T. Oates. Imaging time-series to improve classification and imputation. *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*, 2015.