MISTI: Multi-Style Transfer for Multivariate Time Series

Henri Hoyez*‡, Bruno Mirbach †, Jason Rambach†, Cedric Schockaert*, Didier Stricker†‡
*Paul Wurth S.A
henri.hoyez@isen.yncrea.fr, cedric.schockaert@sms-group.com
†German Research Center for Artificial Intelligence (DFKI)
{bruno.mirbach, jason.rambach, didier.stricker}@dfki.de
‡RPTU Kaiserslautern

***Abstract*—Time series data are generally easy to obtain but often suffer from issues such as incomplete labeling, missing values, and privacy constraints. Transferring data from one domain to another using Machine Learning offers a promising solution to these challenges. This paper introduces a novel feed-forward multi-style transfer algorithm for time series. The proposed approach utilizes dual encoders to disentangle content and style from input sequences, which are then recombined by a decoder to generate sequences with the specified characteristics. Additionally, we propose a new metric to evaluate domain shifts and quantify implicit differences between datasets. Our method demonstrates robust transfer performance across diverse datasets, ranging from synthetic datasets to multivariate human activity recognition time series.**

***Index Terms*—Machine Learning, Domain Adaptation, Multi-Style Style Transfer, Multivariate Time Series, Generative Models**

I. Introduction

Multivariate Time Series (MTS) measures multiple variables evolving over time. These measurements are able to characterize complex physical phenomena. For example, in the medical field, signals recorded such as heart rate or electroencephalograms, enabling the detection of abnormal heartbeats or epileptogenic zones [1].

However, MTS data can be complex and fundamentally multimodal in many cases because they measure physically different quantities, making the development of general algorithms difficult [2]. Moreover, an algorithm trained on one dataset may not necessarily perform well if the statistics of the data change. This can be caused by domain shift, which has gained interest in recent years not only in vision but also in time series [3], [4].

As an illustration, in the medical field, the sharing of patient data can be difficult due to privacy concerns [2]. To address this problem, research has sought solutions through data augmentation, by using generative models [5] or style transfer algorithms [6]. Style transfer is known to change the style of an image without altering its content. More precisely, Multiple-Style Transfer is able to learn from multiple styles in order to offer more variations of a given content [7]. This field has started to be explored in time series with iterative methods [6], [8]. However, these algorithms are only able to perform 1-to-1 translation. In other words, transfer a content time series to only one target style.

On the evaluation side, metrics such as Train on Synthetic Test on Real (TSTR) [9] or precision-recall [10] are commonly used. However, they require a pre-trained model to project the time series into a specific latent space. Some papers use more domain-based metrics such as correlation matrices [5]. However, these specific metrics need specific domain knowledge.

To address these challenges, in this paper, we propose Multi-Style Transfer for Time Series (MISTI). This method is based on two encoders, where the content encoder extracts the content while the style encoder learns from multiple styles, enabling 1-to-n style translation in contrast to other methods. Then, a generator is trained to generate a content sequence while satisfying the style constraint. To the best of our knowledge, this is the first feed-forward multi-style transfer algorithm designed for multivariate time series.

In addition, the recent literature introduced a more nuanced definition of domain shift either based on the out-of-distribution or a domain-based oriented definition. Based on the domain-based definition of domain shift [11], we propose a new metric named CorMet that evaluates the difference between datasets by extracting a signal signature through multiple correlations. This new metric aims to quantify the difference among sensor relationships, between two sets of time series. This metric is a step towards better time series generation evaluation across multiple styled datasets.

Our algorithm has been evaluated on a synthetic dataset for controlled domain adaptation [11] and a real multi-sensor Human Activity Recognition [12] dataset and is shown to provide robust performance over different settings. Our code is available[1].

II. Related Work

A. Time Series Disentanglement:

Disentanglement learning aims at a separation of factors of variation in order to learn a more optimized and meaningful representation from the data. Since its introduction in

This work is supported by Fonds National de la Recherche (FNR), Luxembourg grant AFR-PPP, number 15411817.

[1]https://github.com/Henri-Hoyez/MISTI-Multi-Style-Transfer-for-Multivariate-Time-Series

constrained latent space methods [13], it has been applied in a variety of fields [14]. In time series, disentanglement learning methods have been employed for Time Series Forecasting [15], Non-intrusive Load Monitoring [16], or Causal Discovery [17]. Inspired by the disentanglement field in the imaging community, Li et al. [15] proposed the first disentanglement learning algorithm for time series called DTS. This approach uses two levels of disentanglement to extract individual and group-level factors of variation. Further, Woo et al. [18] presented a contrastive learning-based approach for learning a seasonal-trend representation. The seasonal periodicity is captured by a learnable Fourier Layer, whereas the trend is captured with an auto-regressive model. Oublal et al. [16] introduced DIOSC where they release the statistical independence constraint by using weak contrastive learning and Attentive l-Variational Inference. All these methods show good performance in disentangling factors of variation in time series. These works build a domain-invariant space via disentanglement but do not separate specific features. Since our goal differs, we use a contrastive disentanglement loss from [7] to control feature extraction between content and style representations.

B. Time Series Generations:

Because time series data can sometimes be difficult to acquire, proper data augmentation methods have gained a lot of attention in recent years [19]. Yoon et al. [20], proposed an MTS generation method by applying an adversarial constraint on the generator's latent space. To solve the deterministic problem of previous methods, they proposed to train two encoders: one for real sequences, and another to project a random vector in the latent space of the generator. In the same spirit, Seyfi et al. [5] state that a good generative model for time series should not only make good generations for each time series but also capture the correlations between time series. They hence propose COSSCI-GAN, a generative model composed of local generators and discriminators to generate time series with a good level of quality and a central discriminator to ensure that the correlations are learned correctly. Yuan et al. [21] proposed Diffusion-TS, a diffusion model that learns a disentangled seasonal-trend representation. In this paper, we take advantage of the feed-forward architecture in order to make several generations in real-time by using the generation pipeline from Seyfi et al. [5].

C. Time Series Style Transfer:

In the field of a more constrained generation, Da Silva et al. [22] presented a style transfer algorithm applied to financial data. Using a pretrained denoising autoencoder, they extract the content and the style from pretrained layers and apply an iterative optimization algorithm. El-Laham et al. [8] propose to eliminate the use of pretrained architecture in optimizing predefined equations based on financial equations. In the medical field, Zanini et al. [6] state that measurements on patients affected by Parkinson's disease are challenging because the patient's movements can be restricted. To propose a dataset with a higher degree of movement, they propose to use DCGAN and a Style Transfer Algorithm for augmenting Parkinson's disease electromyography (EMG) signals. However, these algorithms are only able to make a 1-to-1 translation. In contrast, our method can perform 1-to-n translations.

D. Evaluation Metrics:

The unintuitive nature of time series makes the evaluation of generative models difficult. In reaction to this challenge, Yoon et al. [20] introduced Discriminative Score and Predictive Score. Esteban et al. [9] proposed Train on Synthetic and Test on Real to evaluate the alignment of generated features on the real data. Further, Sajjadi et al. [10] formulated an evaluation strategy based on precision and recall of distributions, where precision is associated with the quality of the sample and recall is associated with the overlap between the reference and the learned distribution. Going to more time series-specific metrics, Seyfi et al. [5] suggested evaluating the cross-correlation matrix as a part of their benchmark to check the time alignment of their generations. This is also related to Li et al. [23], who use Wavelet Coherence. Even though our method is similar to Li et al. in using feature extraction methods, our method is made to evaluate a set of generated sequences.

III. Task Formulation

Although style and content can be intuitive in images, the definition can be less natural for time series.

Hoyez et al. [11] proposed a content-style definition based on a mathematical model where the content is defined as the global variation of the signals and the style as the relationship between signals, introduced by the model parameters. For example, Fig. 1 represents acceleration recordings of subjects performing different activities. In this example, the overall fluctuations in the acceleration data correspond to the type of activity being performed, whereas the specific patterns and interdependencies among the signals reveal the individual's unique manner of executing that activity. Then, given a sequence defined by specific content and style, our objective is to adjust its particular variations so they align with the

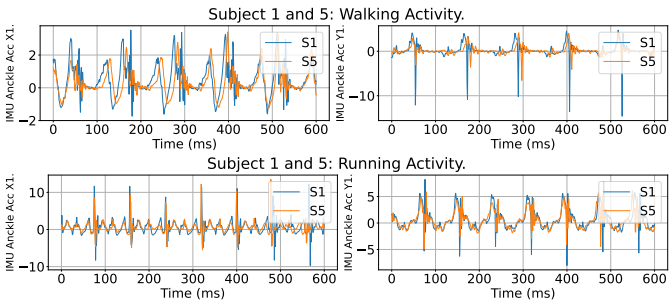

Fig. 1. Example of Content-Style in Human Activity Recognition settings. In this figure, the blue and the orange line represents a reference and a compared subject respectively. The left and right subplots represent x and y acceleration. The subjects perform different activities (i.e. walking on top, running on bottom). In this example, the content is defined by the activity and the style by the person performing it.

target style. In other words, we aim to transform one subject's walking pattern to resemble that of another subject.

Formally, let $X_{c_i,s_j} = \{x_t\}_{t=1}^{L}$ with $x_t \in \mathbb{R}^m$ be a multivariate time series of length L with m channels. Let $\mathbb{C}$ and $\mathbb{S}$ be the content and the style space respectively. In our framework, since an MTS has a component in the content and the style space, a given time series is written as a function of the content and the style $X_{c_i,s_j} = F(c_i, s_j)$ where $c_i \in \mathbb{C}$ and $s_j \in \mathbb{S}$ are points in the content space and style space respectively. Our goal is to provide a model that transfers the style of a given so-called content sequence X_{c_i,s_j} to the wanted style of a so-called style sequence Y_{c_k,s_l} without impacting the content. More formally, our goal is to learn a function $T(X_{c_i,s_j}, Y_{c_k,s_l}) : F(c_i, s_j) \rightarrow F(c_i, s_l)$.

IV. Correlation Metric (CorMet)

Due to the unintuitive side of time series, properly evaluating time series generative models can be difficult. Metrics such as Train on Synthetic and Test on Real [9] or precision-recall [10] provide a good basis. However, these are based on black box models, which may not easily highlight an evaluated model's weakness. To propose a more domain-specific evaluation, the literature proposed time series-specific metrics such as Wavelet Coherence [23] or Correlation matrices empirical comparison [5]. However, these specific metrics require domain knowledge to make the correct analysis.

To put another brick in time series evaluation, based on the content-style definition in [11], we propose a metric that evaluates the style difference between two MTS datasets based on temporal dependencies extraction.

To approximate the style, our metric must be sensitive to changes like noise and delays. Additionally, because the style may not be completely defined with one sequence, the metric has to summarize temporal features from several sequences.

To extract the temporal relationships, we use Pearson Cross correlations (PCC) across the combination of channels and on several sequences to get our signature. An example of these temporal signatures can be found in Fig 2. Then, the difference between these two signatures will result in our metric.

To extract these signatures, Let X_{c_i,s_j} be a multivariate time series sequence written $X_{i,j}$ for simplicity. From $X_{i,j}$, we take an input signal $u(\tau) = \{x_t^{(j)}\}_{t=\tau}^{\tau+l}$ of length l, where τ is a time varying index. We then take the first derivative of the output signal sequence as another channel of our MTS $o' = \{\frac{x_t^{(k)}}{dt}\}_{t=0}^{l}$ to enhance the temporal senssibility. We finally compute a signature as follows:

$$R_{u,o'}(\tau) = \frac{\text{Cov}(u(\tau), o')}{\sqrt{\sigma_u \times \sigma_{o'}}} \quad \tau = 0, \dots, L - l \tag{1}$$

Where $R_{u,o'}(\tau)$ between $u(\tau)$ and o', Cov is the covariance function, σ_u and $\sigma_{o'}$ are the standard deviation of $u(\tau)$ and o' respectively. Further, our signature is computed over all time index τ to reveal temporal dependencies fluctuations. This formula is computed over several sequences and several channels to get multiple signatures. These multiple signatures are stacked to form an area as Fig. 2, where the style difference can be measured by comparing the areas.

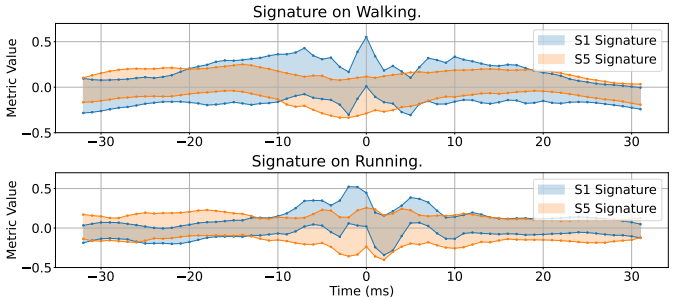

Fig. 2. Visualization of signature generated by CorMet between Subject 1 and Subject 5 on walking on top and running on bottom. The signatures are extracted from the x, y, and z acceleration from the ankle inertial sensor. We can see the difference in the signal relationships (style shift) influences the signatures and exhibit different patterns due to the different time relationships.

V. Proposed Method

In this section, we introduce our feed-forward multi-style transfer algorithm for time series. An overview of the approach can be found in Fig. 3.

Firstly, we build our transfer function T as a combination of 3 models: the content encoder E_c, the style encoder E_s, and the conditional generator G, similarly to [7]. More specifically, the E_c projects the input sequence to a Wiener-like content space to enable a time dependency inside the content space. E_s projects a given sequence into a point in the style space. Then, the generator G aims to generate realistic sequences given the content and the style projections. For readability, we will use $T(X, Y) = G(E_c(X), E_s(Y))$, where X and Y are a content and a style sequence coming from an unlabeled content dataset $\mathcal{U}_C$ and a labeled style dataset $\mathcal{U}_S$ respectively.

For the generator, we modified the architecture of [5] to adapt it for multi-style generations. G is composed of m channel conditional generators G^i , where $0 < i \leq m$, which aim to generate a particular channel of the MTS given content and style features. The quality of G^i's generations is controlled by a channel discriminator D^i. To ensure the correct time dependency between channels and the translation of the style, a global multitask discriminator [24] D_g is constructed.

A. Generation pipeline

Generating realistic samples is enforced by the following constraints:

Identity Reconstruction: Firstly, we want our model to keep content sequence information independently from the style. We hence employ a reconstruction loss for all style datasets In $\mathcal{U}_S$ with an L_2 distance:

$$\ell_r = \mathop{\mathbb{E}}_{(Y,y)\in\mathcal{U}_S}[||\hat{Y} - Y||_2] \tag{2}$$

Where $\hat{Y} = T(Y, Y)$ is the identity reconstruction, y is the associated style class of Y. To better guide local generators, this loss is also applied to each channel, providing a particular reconstruction loss for each local generator.

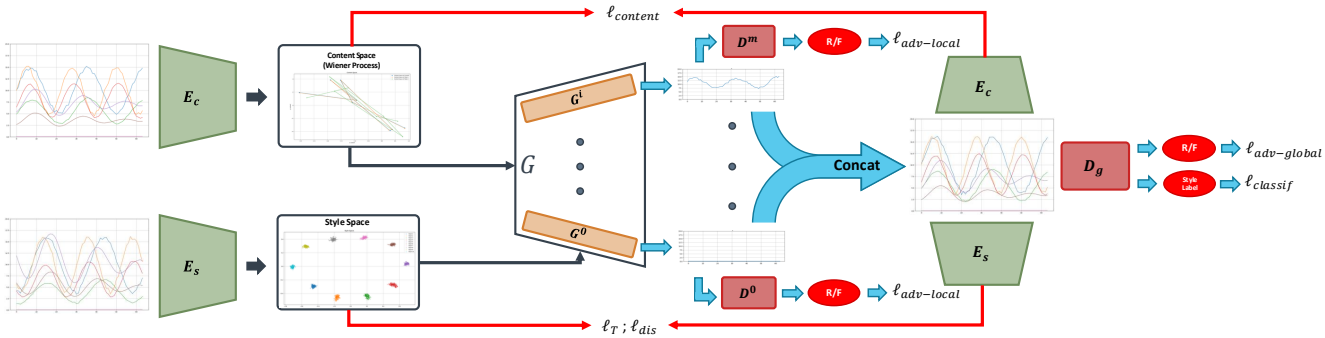

Fig. 3. Overview of our method. E_s and E_c represent our content and our style encoder. Our generator is composed of several channel generators named from G^0 to G^m. The adversarial part is composed of several channel discriminators denoted by D^0 to D^m. The global discriminator is denoted as D_g.

Channel-wise Realistic Generation: For the channel generation, we want our local generator to generate realistic enough samples to fool the local discriminator. Because one time series is not always sufficient to define the style, we first use the unconditional adversarial loss for the interactions between the local generator G^i and a local discriminator D^i.

$$\ell^i_{adv-local} = \frac{1}{2} \mathop{\mathbb{E}}_{\substack{X\sim\mathcal{U}_C \\ (Y,y)\sim\mathcal{U}_S}} [(D^i(G^i(E_c(X), E_s(Y))) - a)^2] + \frac{1}{2} \mathop{\mathbb{E}}_{(Y,y)\sim\mathcal{U}_S} [(D^i(Y) - b)^2] \tag{3}$$

To reduce the mode collapse inside the content space, we use the LS-GAN loss from [25]. In Eq. 3, a and b are constants set to values from the original paper (i.e. $a = 1$ and $b = 0$).

Global Generation & Style Preservation: To enhance the realism of our model generations in a specific style while ensuring correlation preservation, a Multitask Discriminator is employed [24]. The interactions between the local generators and the global discriminator are materialized by the conditional version of the adversarial loss for Eq. 3 to better guide G.

B. Feature Extraction

We want to generate multiple styles. To do so, we want to train E_s to project a given sequence into a smooth space where the same styles are clustered together and different styles are pushed away while being invariant to content changes. We therefore want to minimize these two constraints:

Style Separation: The first constraint is the separation between styles. To satisfy this constraint, we implement the triplet loss.

$$\ell_T = \mathop{\mathbb{E}}_{\substack{X\in\mathcal{U}_C \\ (Y_1,y_1),(Y_2,y_2)\in\mathcal{U}_S}} \max(0, \alpha + ||E_s(Y_1) - E_s(T(X, Y_1))||^2 - ||E_s(Y_1) - E_s(T(X, Y_2))||^2) \tag{4}$$

Where α is the margin between the positive and the negative samples.

Content-Style Disentanglement: To achieve a realistic style transfer, E_s has to correctly disentangle the content and the style. We implement this constraint by using a fixed point disentanglement loss.

$$\ell_{dis} = \mathop{\mathbb{E}}_{\substack{X_1,X_2\in\mathcal{U}_C \\ (Y,y)\in\mathcal{U}_C}} \max(0, ||E_s(T(X_1, Y) - E_s(T(X_2, Y))||^2 - ||E_s(T(X_1, Y)) - E_s(Y)||^2) \tag{5}$$

Content Extraction: We want the content of our sequence to be invariant during the style transfer, hence we train E_c to minimize its projection between a real sequence and the associated generated sequence:

$$\ell_{content} = \mathop{\mathbb{E}}_{\substack{X\in\mathcal{U}_C \\ (Y,y)\in\mathcal{U}_S}} [E_c(X) - E_c(T(X, Y))] \tag{6}$$

To achieve correct MTS generations, our model has to reach a saddle point characterised by this equation:

$$\min_G \max_D \mathcal{L}^* = \lambda_r \ell_r + \lambda_{adv} \ell_{adv-global} + \lambda_{adv} \ell_{adv-local} + \lambda_{style} \ell_T + \lambda_{dis} \ell_{dis} \tag{7}$$

Where λ_r, λ_{adv}, λ_{style}, λ_{dis} are parameters to tune the importance of each losses.

VI. Evaluation:

In this Section, we aim to evaluate our architecture against an appropriate baseline. This evaluation is twofold. Firstly, we evaluate methods on a multivariate synthetic dataset, providing different types of domain shifts. Further, to test our model on real data, we evaluate our model on the Human Activity Recognition dataset. This dataset provides real, visible and intuitive style shifts applied in MTS.

Baseline: We evaluate our method against StyleTime [8]. This iterative time series style transfer algorithm modifies the content sequence in order to minimize a proper style loss between the generation and the style sequence. Since no implementation of this method is available from the authors, we reimplemented it.

A. Datasets:

- **Synthetic Dataset of [11]:** Describes a simplified chemical reaction environment and provides controllable dynamics. Following the original paper, we evaluate on Input Noise, Output Noise, Time shift, and Causal shift.

- **Human Activity Recognition (PAMAP2) [12]:** Is a sensor recording of 3 inertial measurement units placed on the wrist, the chest, and the ankle. During the recording, the subject is asked to perform different activities. In this evaluation, we chose the inertial sensor placed on the wrist and ankle, with the x, y, and z acceleration.

B. Metrics:

- **Train on Synthetic Test on Real (TSTR) [9]:** This metric is the performance of the given model trained on generated data, tested on real data. By doing this, we can assess if the features of the generated time series are correctly aligned with the real dataset.
- **Correlation Metric (CorMet):** To complete our benchmark, we evaluate our model with our proposed metric described in Sec.IV.

TABLE I
COMPARISON OF THE PROPOSED METHOD AGAINST STYLETIME ON MULTIVARIATE SYNTHETIC AND REAL DATA. THE BOLDED VALUES REPRESENT THE BEST PERFORMANCE FOR THE METRIC WHEREAS THE UNDERLINED VALUES REPRESENT THE BEST PERFORMANCE FOR TSTR.

	Style Time [8]		MISTI [OURS]	
Dataset	CorMet ↓	TSTR ↑	CorMet ↓	TSTR ↑
Input Noise [11]	1.47	0.98	**1.09**	<u>0.99</u>
Output Noise [11]	1.90	<u>0.95</u>	**0.88**	0.94
Time Shift [11]	10.81	0.91	**1.97**	<u>0.95</u>
Causal Shift [11]	**1.15**	0.39	1.35	<u>0.40</u>
PAMAP2 [12]	3.51	0.65	**2.47**	<u>0.77</u>

C. Synthetic Dataset:

In Table I, we can see that our model outperforms the baseline for Input Noise and Time Shift independent of the applied metric, while for Output Noise and Causal Shift the result depends on the metric. For Output Noise, CorMet shows an advantage of MISTI over StyleTime because the performances of StyleTime started to drop on a higher noise value. The reverse applies to Causal shift where MISTI demonstrates worse results according to CorMet. In general, our novel proposed metric CorMet provides significantly larger relative differences between the two style transfer methods than TSTR.

D. Human Activity Recognition (PAMAP2):

For this evaluation, we select 5 activities (Walking, Running, Rope Jumping, Ascending, and descending stairs). We hence select the subjects that perform these activities the longest (e.g. subjects 1, 5, 6, and 8). As the Table. I shows, we consistently surpass the baseline in both metrics. Again, the relative difference measured in terms of CorMet is significant larger than in TSTR.

VII. CONCLUSION:

In this paper, we introduced MISTI, a novel multi-style transfer algorithm for time series. By leveraging dual encoders to disentangle content and style, our method enables flexible style transformations while preserving the original content. Additionally, we proposed CorMet, a new metric that extracts the temporal signature of a particular dataset. We evaluated our model on several multivariate time series datasets and showed that our method demonstrates state-of-the-art performances on Human activity translation while being robust across synthetic datasets. Future work will evaluate our method on other use cases such as domain adaptation and virtual sensor generation.

REFERENCES

[1] S. H. Wang, M. Marzulli, G. Arnulfo, L. Nobili, S. Palva, J. M. Palva, and P. Ciuciu, "Machine Learning Models Trained in a Low-Dimensional Latent Space for Epileptogenic Zone (EZ) Localization," in *2024 32nd of EUSIPCO*, Lyon, France, 2024.

[2] J. Ye, W. Zhang, K. Yi, Y. Yu, Z. Li, J. Li, and F. Tsung, "A Survey of Time Series Foundation Models: Generalizing Time Series Representation with Large Language Model," 2024.

[3] Q. Liu and H. Xue, "Adversarial spectral kernel matching for unsupervised time series domain adaptation," in *Proceedings of the Thirtieth IJCAI*, 2021.

[4] X. Jin, Y. Park, D. Maddix, H. Wang, and Y. Wang, "Domain adaptation for time series forecasting via attention sharing," in *ICML*, 2022.

[5] A. Seyfi, J.-F. Rajotte, and R. T. Ng, "Generating multivariate time series with COmmon Source CoordInated GAN (COSCI-GAN)," in *Advances in NeurIPS*, 2022.

[6] R. Anicet Zanini and E. Luna Colombini, "Parkinson's Disease EMG Data Augmentation and Simulation with DCGANs and Style Transfer," 2020.

[7] D. Kotovenko, A. Sanakoyeu, S. Lang, and B. Ommer, "Content and Style Disentanglement for Artistic Style Transfer," 2019.

[8] Y. El-Laham and S. Vyetrenko, "StyleTime: Style Transfer for Synthetic Time Series Generation," in *Proceedings of the Third ACM International Conference on AI in Finance*, 2022.

[9] C. Esteban, S. L. Hyland, and G. Rätsch, "Real-valued (Medical) Time Series Generation with Recurrent Conditional GANs," 2017.

[10] M. S. M. Sajjadi, O. Bachem, M. Lucic, O. Bousquet, and S. Gelly, "Assessing Generative Models via Precision and Recall," 2020.

[11] H. Hoyez, B. Mirbach, J. Rambach, C. Schockaert, and D. Stricker, "Which Time Series Domain Shifts can Neural Networks Adapt to?" in *2024 32nd of EUSIPCO*, Lyon, France, 2024.

[12] A. Reiss and D. Stricker, "Introducing a New Benchmarked Dataset for Activity Monitoring," in *2012 16th International Symposium on Wearable Computers*, 2012.

[13] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," 2013.

[14] X. Wang, H. Chen, S. Tang, Z. Wu, and W. Zhu, "Disentangled Representation Learning," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2024.

[15] Y. Li, Z. Chen, D. Zha, M. Du, D. Zhang, H. Chen, and X. Hu, "Learning Disentangled Representations for Time Series," 2021.

[16] K. Oublal, S. Ladjal, D. Benhaiem, E. Le-borgne, and F. Roueff, "DISENTANGLING TIME SERIES REPRESENTATIONS VIA CONTRASTIVE INDEPENDENCE-OF-SUPPORT ON l-VARIATIONAL INFERENCE," 2024.

[17] S. Gao, R. Addanki, T. Yu, R. A. Rossi, and M. Kocaoglu, "Causal Discovery in Semi-Stationary Time Series," in *Advances in NeurIPS*, 2023.

[18] G. Woo, C. Liu, D. Sahoo, A. Kumar, and S. Hoi, "CoST: Contrastive Learning of Disentangled Seasonal-Trend Representations for Time Series Forecasting | OpenReview," 2022.

[19] C. Hu, Z. Sun, C. Li, Y. Zhang, and C. Xing, "Survey of time series data generation in iot," *MDPI - Sensors*, 2023.

[20] J. Yoon, D. Jarrett, and M. van der Schaar, "Time-series Generative Adversarial Networks," in *Advances in NeurIPS*, 2019.

[21] X. Yuan and Y. Qiao, "Diffusion-TS: Interpretable diffusion for general time series generation," in *ICLR*, 2024.

[22] B. Da Silva and S. S. Shi, "Style Transfer with Time Series: Generating Synthetic Financial Data," 2019.

[23] X. Li, M. Sakevych, G. Atkinson, and V. Metsis, "BioDiffusion: A Versatile Diffusion Model for Biomedical Signal Synthesis," 2024.

[24] A. Odena, C. Olah, and J. Shlens, "Conditional Image Synthesis with Auxiliary Classifier GANs," in *Proceedings of the 34th ICML*, 2017.

[25] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE ICCV*, 2017, pp. 2794–2802.