MPPG: Pluggable Multi-Periodic Pattern-Guided Approach for Multivariate Time Series Anomaly Detection

Zhaobin Meng, Zongxia Xie ✉

School of Intelligence and Computing, Tianjin University 300354, China
zhbmeng@126.com,caddiexie@hotmail.com

Abstract. In recent years, reconstruction-based deep models have been widely applied in time series anomaly detection. However, these methods often face challenges in extracting sufficient features for complex time series, leading most models to also reconstruct anomalies well. Moreover these methods obtain patterns at a single scale and do not take into account the multi-periodicity of the time series. To address this issue, we propose **M**ulti-**P**eriodic **P**attern **G**uidance (**MPPG**), an approach that utilizes multi-periodic patterns to guide model reconstruction. MPPG is a pluggable method applicable to various reconstruction models based on the Encoder-Decoder Architecture. First, through an unsupervised loss function, it ensures that features within the same period are closely together, while features from different periods are distinctly separated. The process facilitates the extraction of pattern information unique to each period. Then it integrates this pattern information into the feature space of any reconstruction model to improve the reconstruction accuracy and anomaly detection capabilities. We evaluated MPPG on six real-world datasets and achieved good experimental results.

Keywords: Time series anomaly detection · Patterns extraction · Multi-period identification.

1 Introduction

In recent years, deep learning-based algorithms have focused on prediction and reconstruction methods. However, prediction-based models often fail to detect anomalies accurately due to rapid changes and unpredictability in time series[3]. Thus, we focus on reconstruction-based deep learning methods.

Existing reconstruction-based methods usually reconstruct the input sequence by learning its feature representation or a single pattern representation, and these methods suffer from the problem of inadequate feature representation of the input sequence. To address this issue, it is important to mine the multi-periodic patterns of time series data, which can more accurately describe the normal features of the sequence. We can integrate these features into the reconstruction model to enhance the reconstruction ability. Therefore, the normal and abnormal features become easy to distinguish, as shown in Fig.1.

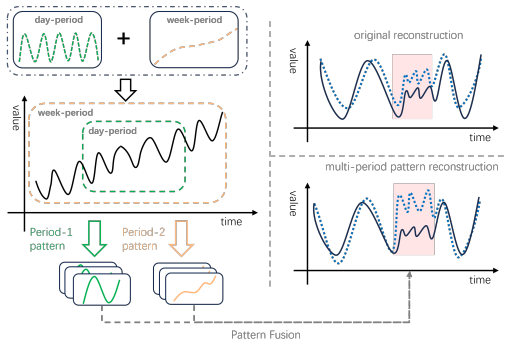

Fig. 1. Differences between original reconstruction and multi-periodic pattern reconstruction.

In this work, we analyze time series from the perspective of multi-periodic patterns to cope with complex temporal variations. We propose a new multi-periodic pattern extraction approach — Multi-Periodic Pattern Guidance (MPPG). The contributions are summarized as follows:

- We propose an efficient multi-periodic pattern extraction method MPPG. It can learn intra-periodic and inter-periodic distinguishable representations through an unsupervised approach and then obtain pattern information of the same and different periods.
- MPPG is a pluggable approach that can integrate the multi-periodic pattern information into the feature space of any reconstruction model based on Encoder-Decoder Architecture, which improves the reconstruction accuracy and anomaly detection capability.
- MPPG achieves state-of-the-art results on six real benchmark datasets, justified by extensive ablations and insightful case studies.

2 Related work

Reconstruction-based methods for time series anomaly detection have been evolving continuously. Early reconstruction-based methods rely on Variational Autoencoder or Autoencoder architectures. Recent methods such as TimesNet [9] and PACdetector [10] use more advanced architectures and incorporate periodic information. However, they still have limitations in feature representation. Some models utilize pattern information for time series tasks. In anomaly detection, MEMTO [6] and MACE [2] adopt different pattern-based approaches but overlook multi-periodic interactions.

To address these issues, we decompose complex temporal variations into multiple periods and extract unique patterns from each period to obtain richer information for guiding model reconstruction.

3 Method

3.1 Problem Definition and Overview

Given a multivariate time series $X \in R^{T \times N}$ where T is the length of each series and N is the number of variables (e.g., the number of sensors). The test set label is denoted by $y = \{y_1, ..., y_T\} \in \{0, 1\}$ where $y_t = 0$ denotes normal points and $y_t = 1$ denotes abnormal points. The difference between the reconstructed sequence and the input sequence $error_t$ is used as the anomaly score at each time point. For a decision threshold τ, if $error_t > \tau$, the time point is taken as an anomaly.

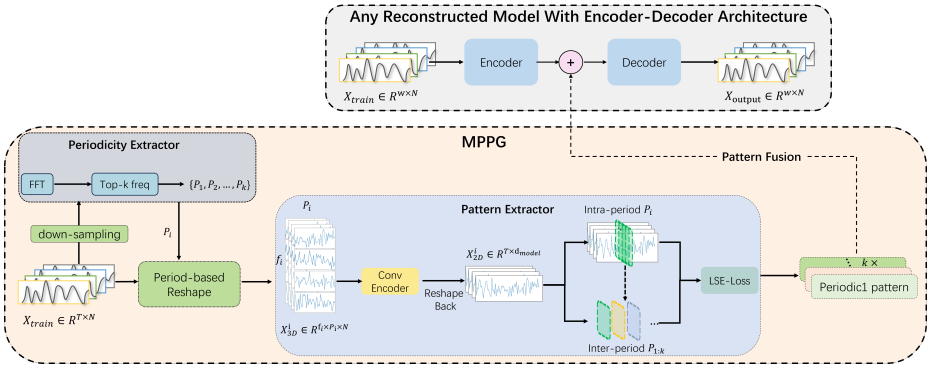

Fig. 2. The architecture of anomaly detection with plug-in MPPG.

Fig.2 illustrates the overall architecture of MPPG, which consists of two main modules: the periodicity extractor module and the pattern extractor module. More technical details about this framework will be elaborated in the subsequent sections.

3.2 Multi-period extraction module

The multi-period extraction module is designed to capture different periods of the input sequence. Inspired by [9], we use the Fast Fourier Transform (FFT) to compute its potential periods. However unlike [9], which uses window data as input, we choose to average downsampling on the entire dataset. Compared to window data, which can only capture local features, global downsampling can more comprehensively extract pattern information. In addition, this method can accelerate training and improve model performance:

$$X_{down}[w] = \frac{1}{r}\sum_{i=0}^{r-1} X[wr + i], w = 0, 1, ..., \lfloor \frac{T}{r} \rfloor - 1 \quad (1)$$

$$A = Avg_{i=1}^{N}(Amp(FFT(X_{down}))), \tag{2}$$

where $X_{down}[w]$ is the new sequence obtained after downsampling, and its length is $\lfloor \frac{T}{r} \rfloor$, r is the downsampling factor, $Amp(\cdot)$ is the calculation of amplitude, $FFT(\cdot)$ is the calculation of FFT, $Avg_{i=1}^{N}$ represents the calculation of the average of the N dimensions, and $A \in R^{L}, \{L = (T/2)+1\}$ represents the calculated amplitude of each frequency. Considering the sparsity in the frequency domain and avoiding noise from high frequencies, we select only the first k amplitude values and calculate the k main periods P_k:

$$\begin{aligned} \{f_1, f_2, ..., f_k\} &= arg(Top-k(A)), k \in \{1, ..., k\}, \\ P_k &= \lceil \frac{T}{f_k} \rceil, k \in \{1, ..., k\}, \end{aligned} \tag{3}$$

where $Top-k(\cdot)$ picks out k frequency values with the highest amplitude from A.

Based on the selected frequency and the corresponding period length, we can shape the input 2D time series into multiple 3D tensors by the following equation:

$$X_{3D}^{i} = Reshape_{p_i, f_i}(Padding(X_{2D})), i \in \{1, ..., k\}, \tag{4}$$

where $Padding(\cdot)$ is to extend the time series by zeros along temporal dimension to make it compatible for $Reshape_{p_i, f_i}$. p_i and f_i represents the number of rows and columns, respectively.

3.3 Pattern extraction module

The pattern extraction module is designed to leverage multi-period information to capture the feature pattern of each period. This module is divided into two steps: **feature representation** and **pattern extraction**. **In the first step**, we learn the features of the 3D sequence for each period through an Inception encoder, which consists of two consecutive inception blocks [8] connected by a GELU activation function to capture multi-scale temporal features. After encoder processing, the feature representation is transformed back to a 2D form, the process can be represented as follows:

$$\begin{aligned} \hat{X_{3D}^{i}} &= Inception(X_{3D}^{i}), i \in \{1, ..., k\}, \\ X_{2D}^{i} &= Reshape_{f_i \times p_i}(\hat{X_{3D}^{i}}), i \in \{1, ..., k\}. \end{aligned} \tag{5}$$

In the pattern extraction step, we adopt a two-stage sampling strategy to sample features within and between periods. This strategy is based on two key observations: 1) Intra-period sampling. M consecutive windows within the same period are likely to contain feature samples from the same pattern. We call these intra-periodic samples, and their representations in the embedding space should be close to each other. 2) Inter-periodic sampling. Windows from different periods are more likely to contain feature samples from different periodic patterns.

These patterns may be significantly different, which we call inter-periodic patterns. Based on these observations, we use the unsupervised LSE loss function:

$$LSE = L_{intra} + L_{inter}. \tag{6}$$

Intra-periodic feature loss L_{intra} aims to maximize the similarity between feature samples in the same period so that they are close to each other in the embedding space:

$$L_{intra} = \alpha_1 \sum_{i=1}^{k} \sum_{m=1}^{M} \sum_{n=1,n<m}^{M} -log(\sigma((o_m^i)^T o_n^i)), \tag{7}$$

where $\alpha_1 = \frac{2}{k*M*(M-1)}$ is used to average the similarity, σ is the sigmoid function, $o_m = X_{t_i+m:t_i+m+w}$ is the feature representation of the m-th window of the i-th period and w is the length of the window. $(o_m^i)^T o_n^i$ is the similarity. The inter-state component L_{inter} encourages minimizing the similarity of feature sample representations between periods and aims to separate them farther in the embedding space:

$$L_{inter} = \alpha_2 \sum_{i=1}^{k} \sum_{l=1,l<k}^{k} -log(\sigma(-c_i^T c_l)), \tag{8}$$

$$c_i = \frac{1}{M} \sum_{m=1}^{M} o_m^i, \tag{9}$$

where c_i is the similarity between the embedding centers of different groups. Note the negative sign before c_i , which implies that minimizing L_{inter} is equivalent to minimizing the embedding similarity of all period samples.

3.4 Pattern fusion with reconstruction model

The pattern fusion strategy integrates multi-periodic patterns into reconstruction models based on the Encoder-Decoder Architecture to enhance their performance. This approach can adaptively adjust the feature representation ability and enhance the models' capacity to reconstruct normal data while increasing the reconstruction error for abnormal data. The process can be represented as follows:

$$\begin{aligned} Weights &= Encoder(X_{in}) \times (Pattern)^T, \\ X_O &= Decoder(Weights \times Pattern + X_c), \end{aligned} \tag{10}$$

where *Encoder* and *Decoder* denote the encoder and decoder of any reconstructed model, X_{in} represents the input of the reconstructed model, X_c represents the representation of the input in the embedding space after the encoder, *Pattern* represents the information of the multi-period pattern extracted by the MPPG framework, $Weights$ represent the weights between each time point of X_c and each multi-periodic pattern, and X_O represents the output of the reconstructed model. It is worth noting that the reconstruction model and the MPPG model are optimized independently and do not affect each other.

4 Experiment

4.1 Experimental setup

Datasets. We evaluate MPPG on six real multivariate time series datasets. These six datasets are publicly available datasets widely used in the anomaly detection problem. (1) Server Machine Dataset (**SMD**) is a 5-week-long dataset collected from a large Internet company with 38 dimensions. (2) Pooled Server Metrics (**PSM**) is collected internally from multiple application server nodes at eBay with 26 dimensions. (3˜4) Mars Science Laboratory rover (**MSL**) and Soil Moisture Active Passive satellite (**SMAP**) are public datasets from NASA with 55 and 25 dimensions respectively. (5) Secure Water Treatment (**SWAT**) is obtained from 51 sensors of the critical infrastructure system under continuous operations for 11 days of continuous operation.(6) Water Distribution (**WADI**) is collected from the WADI testbed. It consists of 16 days of continuous operation, of which 14 days were collected under normal operation and 2 days with attack scenarios.

Metrics. The standard evaluation metrics for binary classification, including precision, recall, and F1-score, do not account for the sequential properties of time series data, making them insufficient for assessing contextual and collective anomalies. Therefore, we use adjusted versions of these metrics, which are widely used in time series anomaly detection [4].

Baselines. We compare the fusion of our approach MPPG with four representative reconstructed models.

The **MLP** model is based on a **simple linear architecture**. A recent study [5] showed complex deep learning models can be simulated by a simple MLP. We reconstruct the data with encoders and decoders consisting of two linear layers and the RELU activation function.

The **USAD** [1] model is based on **autoencoder architecture**. Its autoencoder architecture makes it capable of learning in an unsupervised way.

The **OmniAnomaly** [7] model is based on **the VAE architecture**. Its core idea is to learn data robust representations with key techniques such as stochastic variable connection and planar normalizing flow.

The **A.T** [11] model is based on **Transformer architecture**. It proposed a new computerized mechanism for computing association differences.

4.2 Main results

Results. We conducted two sets of experiments on each model: the first using the original model and the second using MPPG on the original model. Table 1 presents the evaluation results of four models across six datasets. The results show that the average F1 score for anomaly detection improved by 1.34%, 1.88%, 0.5%, and 1.56% in the four benchmark models, respectively. These findings demonstrate that MPPG effectively enhances anomaly detection performance. However, slight performance degradation was observed in a subset of

Table 1. Results (as %) on six real-world datasets. P, R, F1 denotes the results without the multi-periodic pattern. P (+MPPG), R (+MPPG), F1 (+MPPG) denotes the results with the multi-periodic pattern.

Dataset	Metrics	MLP	USAD	Omni	A.T
SMD	P	76.48	81.44	98.77	79.64
	R	91.25	99.73	99.85	91.77
	F1	83.22	89.67	99.31	85.27
	P (+MPPG)	77.66	96.34	98.59	79.73
	R (+MPPG)	94.17	99.73	99.85	92.47
	F1 (+MPPG)	85.12 +1.90	98.01 +8.34	99.22 −0.09	85.63 +0.36
MSL	P	87.22	79.90	78.09	95.18
	R	85.42	99.99	99.99	78.61
	F1	86.31	88.82	87.69	86.11
	P (+MPPG)	88.38	83.33	81.58	95.95
	R (+MPPG)	85.76	99.99	99.99	89.97
	F1 (+MPPG)	87.05 +0.74	90.91 +2.09	89.85 +2.16	92.86 +6.75
SWAT	P	99.31	94.39	88.46	89.34
	R	77.66	77.17	81.37	82.88
	F1	87.16	84.91	84.77	85.99
	P (+MPPG)	99.25	88.57	88.54	91.34
	R (+MPPG)	79.23	82.25	82.25	85.10
	F1 (+MPPG)	88.12 +0.96	85.29 +0.38	85.28 +0.51	88.11 +2.12
WADI	P	65.70	72.71	76.60	82.29
	R	36.90	40.10	48.61	99.99
	F1	47.26	51.69	59.48	90.28
	P (+MPPG)	67.28	72.97	76.60	81.70
	R (+MPPG)	42.24	40.10	48.61	99.99
	F1 (+MPPG)	51.90 +4.64	51.76 +0.07	59.48 +0.00	89.93 −0.35
SMAP	P	88.69	75.02	74.73	93.61
	R	58.05	99.99	99.99	99.13
	F1	70.17	85.73	85.53	96.29
	P (+MPPG)	89.53	75.56	75.33	93.61
	R (+MPPG)	57.81	99.99	99.99	99.41
	F1 (+MPPG)	70.25 +0.08	86.08 +0.34	85.93 +0.40	96.42 +0.13
PSM	P	91.44	99.86	99.96	98.26
	R	95.13	77.95	77.77	96.85
	F1	93.25	87.56	87.48	97.55
	P (+MPPG)	90.55	99.80	99.98	98.10
	R (+MPPG)	95.56	78.12	77.79	97.66
	F1 (+MPPG)	92.99 −0.26	87.64 +0.08	87.49 +0.01	97.87 +0.32
Avg F1 Improvement		+1.34	+1.88	+0.50	+1.56

experiments, likely due to dataset-specific characteristics and variations in compatibility between the MPPG framework and different base models. We plan to investigate this phenomenon further in future work.

4.3 Ablation study

To verify the effectiveness of MPPG, we designed two sets of experiments to compare with the original model. First, we compared the performance of the original model with that of the model using MPPG. Additionally, we extracted the original patterns for experimentation. Unlike the period decomposition using FFT, the extraction of original patterns was performed directly on the original sequence. We randomly selected M discontinuous windows in the input time series, with each window containing a specific type of pattern information. For each type of pattern information, we obtained N consecutive windows and extracted the patterns on the original sequence using the unsupervised loss function LSE. The results are shown in Table 2.

Table 2. F1-score on 6 datasets with different pattern fusion methods

Methods	Patterns		Datasets						Avg.
	Mul-peri	Original	SMD	MSL	SWAT	WADI	SMAP	PSM	
backbone	×	×	89.37	87.23	85.71	62.18	84.43	91.46	83.39
backbone+pattern	×	✓	89.55	87.88	86.35	60.11	84.42	91.48	83.30
backbone+MPPG	✓	✓	**91.99**	**90.17**	**86.70**	**63.27**	**84.67**	**91.50**	**84.72**

4.4 Discussion and analysis

Multi-periodic patterns can provide richer information. In our experiments, we performed T-SNE dimensionality reduction on each multi-periodic pattern. Subsequently, we plotted the minimum convex polygons of these points using the convex hull algorithm. By comparing the areas of these polygons, we can quantitatively assess the amount of information in the patterns extracted by the different methods. As shown in Fig.3, this result clearly demonstrates that MPPG not only contains richer pattern information but also characterizes the data more comprehensively.

In addition, for the time series without clear periodicity, MPPG can still obtain accurate pattern information. If the period is the entire sequence, the intra-period variation is just the original variation of raw series. As shown in Fig.4, this is a visualization of the patterns extracted from the MSL dataset. We can find that the patterns extracted by MPPG from data with insignificant periodicity are very similar.

Multi-periodic patterns can amplify the difference between normal and abnormal values. To verify this, we used the Wasserstein distance as a

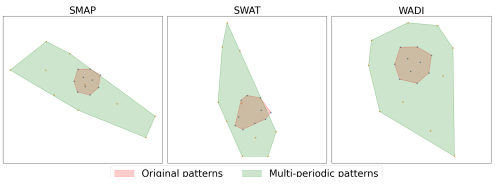

Fig. 3. The green and red regions represent the information amounts of multi-periodic patterns and original patterns in different methods respectively.

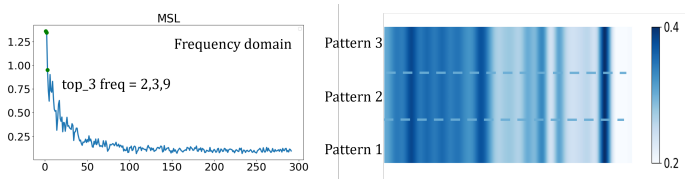

Fig. 4. For the time series without clear periodicity, their pattern has little difference.

measure to quantify the difference between normal and abnormal values. We calculated the Wasserstein distance between normal and abnormal data for each variable and averaged the distance values for all variables. As shown in Fig.5 (left), the experimental results show that the difference between the normal and abnormal values of the reconstructed data increases significantly after the inclusion of the multi-periodic pattern. This finding suggests that our proposed MPPG can significantly improve the accuracy of the reconstruction and the detection ability of the model by amplifying the difference between normal and abnormal values.

Parameter sensitivity. We also performed an analysis of the parameter sensitivity and training time of MPPG and show the results in Fig.5 (right). First, we investigate the effect of the selection of the *top-k* parameter on the performance of the model. As depicted in Fig.5, we can find that our proposed MPPG can present a stable performance under different choices of k in all datasets and the model achieves optimal results when k is set to 5. Then, since this method is inspired by TimesNet, we evaluated the time consumed per epoch of training for both MPPG and TimesNet. Because the input of MPPG is global data processed by downsampling, while TimesNet's input is windowed data intercepted by a sliding window, MPPG shows a significant performance improvement in terms of training time.

5 Conclusion and future work

We propose MPPG, a pluggable multi-period pattern extraction approach that enhances time series anomaly detection accuracy through periodic pattern fusion. Its effectiveness is validated by conducting experiments on real-world multivariate time series benchmarks. In future work, we plan to systematically investi-

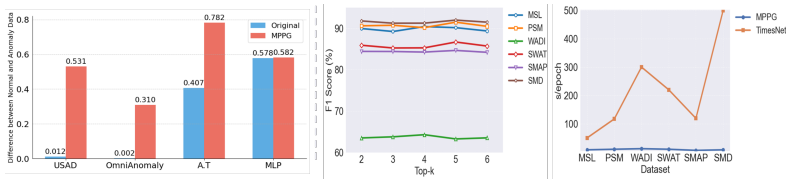

Fig. 5. Left:Difference between normal and abnormal values after reconstruction of SWAT.Right:Parameter sensitivity and training time analysis

gate the framework's compatibility with diverse model architectures, identifying optimal integrations and diagnosing underlying causes of performance degradation in suboptimal combinations.

References

1. Audibert, J., Michiardi, P., Guyard, F., Marti, S., Zuluaga, M.A.: Usad: Unsupervised anomaly detection on multivariate time series. In: Proceedings of the 26th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 3395–3404 (2020)
2. Chen, F., Zhang, Y., Qin, Z., Fan, L., Jiang, R., Liang, Y., Wen, Q., Deng, S.: Learning multi-pattern normalities in the frequency domain for efficient time series anomaly detection. In: 2024 IEEE 40th International Conference on Data Engineering (ICDE). pp. 747–760. IEEE (2024)
3. Darban, Z.Z., Webb, G.I., Pan, S., Aggarwal, C.C., Salehi, M.: Deep learning for time series anomaly detection: A survey. arXiv preprint arXiv:2211.05244 (2022)
4. Goswami, M., Challu, C., Callot, L., Minorics, L., Kan, A.: Unsupervised model selection for time-series anomaly detection. arXiv preprint arXiv:2210.01078 (2022)
5. Saquib Sarfraz, M., Chen, M.Y., Layer, L., Peng, K., Koulakis, M.: Position paper: Quo vadis, unsupervised time series anomaly detection? arXiv e-prints pp. arXiv–2405 (2024)
6. Song, J., Kim, K., Oh, J., Cho, S.: Memto: Memory-guided transformer for multivariate time series anomaly detection. Advances in Neural Information Processing Systems **36** (2024)
7. Su, Y., Zhao, Y., Niu, C., Liu, R., Sun, W., Pei, D.: Robust anomaly detection for multivariate time series through stochastic recurrent neural network. In: Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining. pp. 2828–2837 (2019)
8. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1–9 (2015)
9. Wu, H., Hu, T., Liu, Y., Zhou, H., Wang, J., Long, M.: Timesnet: Temporal 2d-variation modeling for general time series analysis. arXiv preprint arXiv:2210.02186 (2022)
10. Xu, J., Zhu, Z., Ye, W., Gui, N.: Periodicity association based contrastive learning for time series anomaly detection. In: 2024 International Joint Conference on Neural Networks (IJCNN) (2024)
11. Xu, J., Wu, H., Wang, J., Long, M.: Anomaly transformer: Time series anomaly detection with association discrepancy. arXiv preprint arXiv:2110.02642 (2021)