Time Series Data Augmentation for Deep Learning: A Survey

Qingsong Wen, **Liang Sun**, **Xiaomin Song**, **Jingkun Gao**, **Xue Wang** and **Huan Xu**

Machine Intelligence Technology, DAMO Academy, Alibaba Group

500 108th Ave, Bellevue 98004, WA, USA

{qingsong.wen, liang.sun, xiaomin.song, jingkun.g, xue.w, huan.xu}@alibaba-inc.com

Abstract

Deep learning performs remarkably well on many time series analysis tasks recently. The superior performance of deep neural networks relies heavily on a large number of training data to avoid overfitting. However, the labeled data of many real-world time series applications may be limited such as classification in medical time series and anomaly detection in AIOps. As an effective way to enhance the size and quality of the training data, data augmentation is crucial to the successful application of deep learning models on time series data. In this paper, we systematically review different data augmentation methods for time series. We propose a taxonomy for the reviewed methods, and then provide a structured review for these methods by highlighting their strengths and limitations. We also empirically compare different data augmentation methods for different tasks including time series anomaly detection, classification and forecasting. Finally, we discuss and highlight future research directions, including data augmentation in time-frequency domain, augmentation combination, and data augmentation and weighting for imbalanced class.

1 Introduction

Deep learning has achieved remarkable success in many fields, including computer vision (CV), natural language processing (NLP), and speech processing, etc. Recently, it is increasingly embraced for solving time series related tasks, including time series classification [Fawaz *et al.*, 2019], time series forecasting [Han *et al.*, 2019], and time series anomaly detection [Chalapathy and Chawla, 2019]. The success of deep learning relies heavily on a large number of training data to avoid overfitting. Unfortunately, many time series tasks do not have enough labeled data. As an effective tool to enhance the size and quality of the training data, data augmentation is crucial to successful application of deep learning models. The basic idea of data augmentation is to generate synthetic dataset covering unexplored input space while maintaining correct labels. Data augmentation has shown its effective-

ness in many applications, such as AlexNet [Krizhevsky *et al.*, 2012] for ImageNet classification.

However, less attention has been paid to finding better data augmentation methods specifically for time series data. Here we highlight some challenges arising from data augmentation methods for time series data. Firstly, the intrinsic properties of time series data are not fully utilized in current data augmentation methods. One unique property of time series data is the so-called temporal dependency. Unlike image data, the time series data can be transformed in the frequency and time-frequency domain and effective data augmentation methods can be designed and implemented in the transformed domain. This becomes more complicated when we model multivariate time series where we need to consider the potentially complex dynamics of these variables across time. Thus, simply applying those data augmentation methods from image and speech processing may not result in valid synthetic data. Secondly, the data augmentation methods are also task dependent. For example, the data augmentation methods applicable for time series classification may not be valid for time series anomaly detection. In addition, in many classification problems involving time series data, class imbalance is often observed. In this case, how to effective generate a large number of synthetic data with labels with less samples remains a challenge.

Unlike data augmentation for CV [Shorten and Khoshgoftaar, 2019] or speech [Cui *et al.*, 2015], data augmentation for time series has not yet been systematically reviewed to the best of our knowledge. In this paper, we summarize popular data augmentation methods for time series in different tasks, including time series forecasting, anomaly detection, classification, etc. We propose a taxonomy of data augmentation methods for time series, as illustrated in Fig. 1. Based on the proposed taxonomy, we review these data augmentation methods systematically. We start the discussion from the simple transformations in time domain first. And then we discuss more advanced transformations on time series in the transformed frequency and time-frequency domains. Besides the transformations in different domains for time series, we also introduce more advanced methods, including decomposition-based methods, model-based methods, and learning-based methods. In decomposition-based methods, the input time series is first decomposed into different components including trend, seasonality and residual, and then different manipulations can be performed on different components, leading to
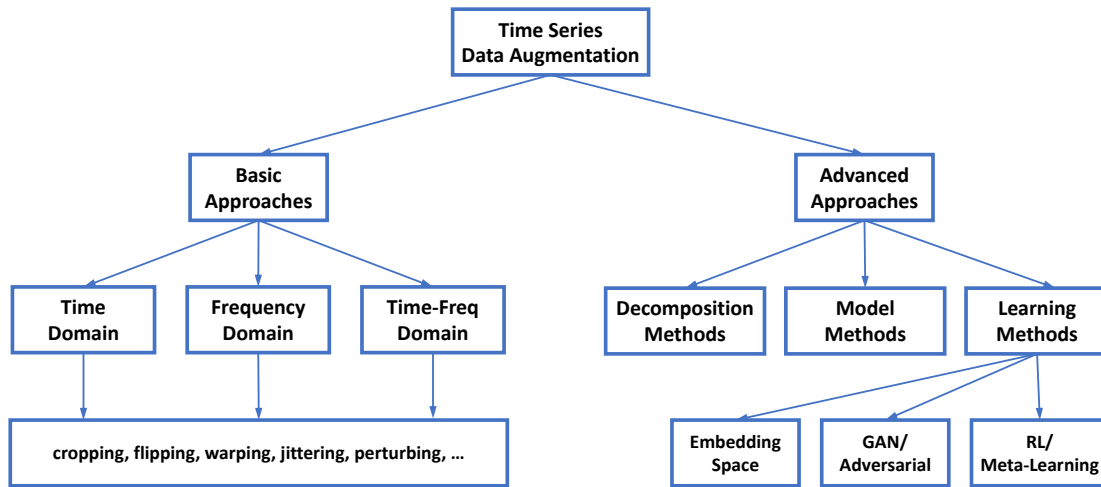
Figure 1: A taxonomy of time series data augmentation.

2 Basic Data Augmentation Methods

2.1 Time Domain

different data augmentation methods. In model-based methods, a statistical model is first learned from the data, and then perturbation is performed on the parameter space of the learned model to generate more synthetic data. We also review some recent works on data augmentation performed in the learned embedding space. In particular, as the generative adversarial networks (GAN) can generate synthetic data and increase training set effectively, we summarize some recent progress on the application of GAN on time series data. To compare different data augmentation methods, we empirically evaluate different augmentation methods in three typical time series tasks, including time series anomaly detection, time series classification, and time series forecasting. Finally, we discuss and highlight some future research directions such as data augmentation in time-frequency domain, augmentation combination of different augmentation methods, and data augmentation and weighting for imbalanced class.

The remaining of the paper is organized as follows. We introduce different transformations in time domain, frequency domain and time-frequency domain in Section 2. We discuss more advanced data augmentations in Section 3, including decomposition-based methods, model-based methods, and learning-based methods. The empirical studies are present in Section 4. The discussions for future work are described in Section 5, and the conclusion is summarized in Section 6.

2 Basic Data Augmentation Methods

2.1 Time Domain

The transforms in the time domain are the most straightforward data augmentation methods for time series data. Most of them manipulate the original input time series directly, like injecting Gaussian noise or more complicated noise patterns such as spike, step-like trend, and slope-like trend. Besides this straightforward methods, we will also discuss a particular data augmentation method for time series anomaly detection, i.e., label expansion in the time domain.

Window cropping or slicing has been mentioned in [Le Guennec *et al.*, 2016]. Introduced in [Cui *et al.*, 2016], window cropping is similar to cropping in CV area. It is a subsample method to randomly extract continuous slices from the original time series. The length of the slices is a tunable parameter. For classification problem, the labels of sliced samples are the same as the original time series. During test time, each slice from a test time series is classified using the learned classifier and use majority vote to generate final prediction label. For anomaly detection problem, the anomaly label will be sliced along with value series.

Window warping is a unique augmentation method for time series. Similar to dynamic time warping (DTW), this method selects a random time range, then compresses (down sample) or extends (up sample) it, while keeps other time range unchanged. Window warping would change the total length of the original time series, so it should be conducted along with window cropping for deep learning models. This method contains the normal down sampling which takes down sample through the whole length of the original time series.

Flipping is another method that generates the new sequence $x_1', \cdots, x_N'$ by flipping the sign of original time series $x_1, \cdots, x_N$, where $x_t' = -x_t$. The labels are still the same, for both anomaly detection and classification, assuming that we have symmetry between up and down directions.

Another interesting perturbation and also ensemble based method is introduced by [Fawaz *et al.*, 2018]. This method generates new time series with DTW and ensemble them by a weighted version of the DBA algorithm. It shows improvement of classification in some of the UCR dataset.

Noise injection is a method by injecting small amount of noise/outlier into time series without changing the corresponding labels. This includes injecting Gaussian noise, spike, step-like trend, and slope-like trend, etc. For spike, we can randomly pick index and direction, randomly assign magnitude but bounded by multiples of standard deviation of the original time series. For step-like trend, it is the cumulative summation of the spikes from left index to right index. The slope-like trend is adding a linear trend into the original time series. These schemes are mostly mentioned in [Wen

and Keyes, 2019]

In time series anomaly detection, the anomalies generally last long enough during a continuous span so that the start and end points are sometimes "blurry". As a result, a data point near to a labeled anomaly in terms of both time distance and value distance is very likely to be an anomaly. In this case, label expansion method is proposed to change those data points and their label as anomalies (by assign it an anomaly score or switch its label), which brings performance improvement for time series anomaly detection as shown in [Gao *et al.*, 2020].

2.2 Frequency Domain

While most of the existing data augmentation methods focus on time domain, only a few studies investigate data augmentation from frequency domain perspective for time series.

A recent work in [Gao *et al.*, 2020] proposes to utilize perturbations in both amplitude spectrum and phase spectrum in frequency domain for data augmentation in time series anomaly detection by convolutional neural network. Specifically, for the input time series $x_1, \cdots, x_N$, its frequency spectrum $F(\omega_k)$ through Fourier transform is calculated as:

$$F(\omega_k) = \frac{1}{N}\sum_{t=0}^{N-1} x_t e^{-j\omega_k t} = \Re[F(\omega_k)] + j\Im[F(\omega_k)]$$
$$= A(\omega_k)\exp[j\theta(\omega_k)] \qquad (1)$$

where $\Re[F(\omega_k)]$ and $\Im[F(\omega_k)]$ are the real and imaginary parts of the spectrum respectively, $\omega_k = \frac{2\pi k}{N}$ is the angular frequency, $A(\omega_k)$ is the amplitude spectrum, and $\theta(\omega_k)$ is the phase spectrum. For perturbations in amplitude spectrum $A(\omega_k)$, the amplitude values of randomly selected segments are replaced with Gaussian noise by considering the original mean and variance in the amplitude spectrum. While for perturbations in phase spectrum $\theta(\omega_k)$, the phase values of randomly selected segments are added by an extra zero-mean Gaussian noise in the phase spectrum. The amplitude and phase perturbations (APP) based data augmentation combined with aforementioned time-domain augmentation methods bring significant time series anomaly detection improvements as shown in the experiments of [Gao *et al.*, 2020].

Another recent work in [Lee *et al.*, 2019] proposes to utilize the surrogate data to improve the classification performance of rehabilitative time series in deep neural network. Two conventional types of surrogate time series are adopted in the work: the amplitude adjusted Fourier transform (AAFT) and the iterated AAFT (IAAFT) [Schreiber and Schmitz, 2000]. The main idea is to perform random phase shuffle in phase spectrum after Fourier transform and then perform rank-ordering of time series after inverse Fourier transform. The generated time series from AAFT and IAAFT can approximately preserve the temporal correlation, power spectra, as well as the amplitude distribution of the original time series. In the experiments of [Lee *et al.*, 2019], the authors conducted two types data augmentation by extending the data by 10 then 100 times through AAFT and IAAFT methods, and demonstrated promising classification accuracy improvements compared to the original time series without data augmentation.
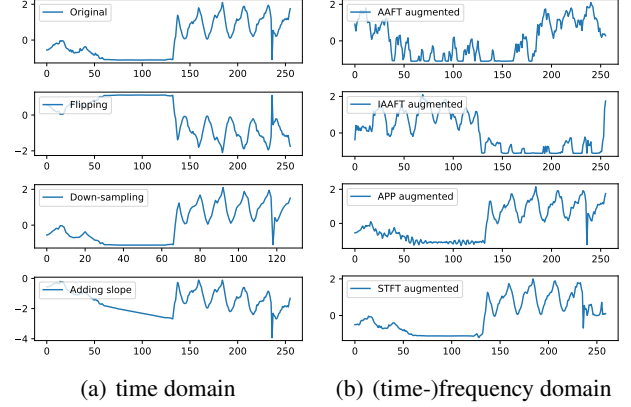

(a) time domain   (b) (time-)frequency domain

Figure 2: Illustration of several typical time series data augmentations in time, frequency, and time-frequency domains.

2.3 Time-Frequency Domain

Time-frequency analysis is a widely applied technique for time series analysis, which can be utilized as an appropriate input features in deep neural networks. However, similar to data augmentation in frequency domain, only a few studies considered data augmentation from time-frequency domain for time series.

The authors in [Steven Eyobu and Han, 2018] adopt short Fourier transform (STFT) to generate time-frequency features for sensor time series, and conduct data augmentation on the time-frequency features for human activity classification by a deep LSTM neural network. Specifically, two augmentation techniques are proposed. One is local averaging based on a defined criteria with the generated features appended at the tail end of the feature set. Another is the shuffling of feature vectors to create variation in the data. Similarly, in speech time series, recently SpecAugment [Park *et al.*, 2019] is proposed to make data augmentation in Mel-Frequency (a time-frequency representation based on STFT for speech time series), where the augmentation scheme consists of warping the features, masking blocks of frequency channels, and masking blocks of time steps. They demonstrate that SpecAugment can greatly improves the performance of speech recognition neural networks and obtain state-of-the-art results.

For illustration, we summarize several typical time series data augmentation methods in time, frequency, and time-frequency domain in Figure 2.

3 Advanced Data Augmentation Methods

3.1 Decomposition-based Methods

Decomposition-based time series augmentation has also been adopted and shown success in many time series related tasks, such as forecasting, clustering and anomaly detection. In [Kegel *et al.*, 2018], authors discussed the recombination method to generate new time series. It first decomposes the time series x_t into trend, seasonality, and residual based on STL [Cleveland *et al.*, 1990]

$$x_t = \tau_t + s_t + r_t, \quad t = 1, 2, ...N \qquad (2)$$

where τ_t is the trend signal, s_t is the seasonal/periodic signal, and the r_t denotes the remainder signal. Then, it recombines new time series with a deterministic component and a stochastic component. The deterministic part is reconstructed by adjusting weights for base, trend, and seasonality. The stochastic part is generated by building a composite statistical model based on residual, such as an auto-regressive model. The summed generated time series is validated by examining whether a feature-based distance to its original signal is within certain range.

Meanwhile, authors in [Bergmeir *et al.*, 2016] proposed apply bootstrapping on the decomposed residuals to generate augmented signals, which are then added back with trend and seasonality to assemble a new time series. An ensemble of the forecasting models on the augmented time series has outperformed the original forecasting model consistently, demonstrating the effectiveness of decomposition-based time series augmentation approaches.

Recently, in [Gao *et al.*, 2020], authors showed that applying time-domain and frequency-domain augmentation on the decomposed residual that is generated using Robust-STL [Wen *et al.*, 2019b] and RobustTrend [Wen *et al.*, 2019a] can help increase the performance of anomaly detection, compared with the same method without augmentation.

To sum up, it is common to decompose time series into different components, such as trend, seasonality, and residual, where each component can be augmented using either bootstrapping or basic time domain and (time-)frequency domain augmentation methods.

3.2 Model-based Methods

Model-based time series augmentation approaches typically involve modelling the dynamics of the time series with statistical models. In [Cao *et al.*, 2014], authors proposed a parsimonious statistical model, known as mixture of Gaussian trees, for modeling multi-modal minority class time-series data to solve the problem of imbalanced classification, which shows advantages compared with existing oversampling approaches that do not exploit time series correlations between neighboring points. Authors in [Smyl and Kuber, 2016] use samples of parameters and forecast paths calculated by a statistical algorithm called LGT (Local and Global Trend). More recently, in [Kang *et al.*, 2019] researchers use of mixture autoregressive (MAR) models to simulate sets of time series and investigate the diversity and coverage of the generated time series in a time series feature space.

Essentially, these models describe the conditional distribution of the time series by assuming the value at time t depends on previous points. Once the initial value is perturbed, a new time series sequence could be generated following the conditional distribution.

3.3 Learning-based Methods

In [DeVries and Taylor, 2017], the data augmentation is proposed to perform in the learned space. It assumes that simple transforms applied to encoded inputs rather than the raw inputs would produce more plausible synthetic data due to the manifold unfolding in feature space. Note that the selection of the representation model in this framework is open and depends on the specific task and data type. When the time series data is addressed, a sequence autoencoder is selected in [DeVries and Taylor, 2017]. Specifically, the interpolation and extrapolation are applied to generate new samples. First k nearest labels in the transformed space with the same label are identified. Then for each pair of neighboring samples, a new sample is generated which is the linear combination of them. The difference of interpolation and extrapolation lies in the weight selection in sample generation. This technique is particular useful for time series classification.

As a generative model, GAN can generate synthetic samples and increase the training set effectively. Although the GAN framework has received significant attention in many fields, how to generate time series still remains an open problem. In this subsection, we briefly review several recent works on on GAN for time series.

Specifically, in GAN we need to build a good and general generative model for time series data. In [Nikolaidis *et al.*, 2019], a recurrent generative adversarial network is proposed to generate realistic synthetic data. As sever unbalance is faced in time series classification, it also enables the generation of balanced datasets. Furthermore, a multiple GAN architecture is proposed to generate more synthetic time series data that is potentially closer to the test data to enable personalized training. In [Esteban *et al.*, 2017], a Recurrent GAN and Recurrent Conditional GAN are proposed to produce realistic real-valued multi-dimensional time series data.

Recently, [Yoon *et al.*, 2019] proposed TimeGAN, a natural framework for generating realistic time-series data in various domains. TimeGAN is a generative time-series model, trained adversarially and jointly via a learned embedding space with both supervised and unsupervised losses. Specifically, a stepwise supervised loss is introduced to learn the stepwise conditional distributions in data. It also introduces an embedding network to provide a reversible mapping between features and latent representations to reduce the high-dimensionality of the adversarial learning space. Note that the supervised loss is minimized by jointly training both the embedding and generator networks.

Reinforcement learning is also introduced in data augmentation [Cubuk *et al.*, 2019a] to automatically search for improved data augmentation policies. Specifically, it defines a search space where an augmentation policy consists of many sub-policies. Each sub-policy consists of two operations, and each operation is defined as an image processing function such as translation, rotation, or shearing. As we discussed early, data augmentation is closely related to the data and the task. Thus, how to introduce the reinforcement learning framework to time series data and different tasks still remains unclear.

4 Experiments

4.1 Time Series Anomaly Detection

Given the challenges of both *data scarcity* and *data imbalance* in time series anomaly detection, it is crucial to make use of time series data augmentation to generate more labelled data based on few existing labels. The work in [Gao *et al.*, 2020] has demonstrated the effectiveness of applying

data augmentations on decomposed signals, which increase the performance of anomaly detection on raw time series data.

We briefly describe the results by applying U-Net structure [Gao *et al.*, 2020] to public Yahoo Dataset [Laptev *et al.*, 2015] using different settings, including applying the model on the raw data (U-Net-Raw), applying the model on the decomposed residuals of signal with RobustSTL (U-Net-De), and applying the model on the residuals with decomposed-based data augmentation (U-Net-DeA). The applied data augmentation methods include flipping, cropping, label expansion, and APP based augmentation in frequency domain in the decomposed components. The performance is evaluated using the standard precision, recall, F1 score, as well as relax F1 score that allows a detection delay up to 3 points. Table 1 shows the performance comparison of different settings. It can be observed that the decomposition helps increasing the F1 score significantly and the decomposed-based data augmentation is able to further boost the performance with a 4.7% increase in F1 score and 5.4% increase in relax F1 score.

Table 1: Performance comparison of different settings. Acronyms in U-Net section: De - decomposition, A - augmentation.

	Precision	Recall	F1	Relax F1
U-Net-Raw	0.553	0.302	0.390	0.593
U-Net-De	0.738	0.551	0.631	0.747
U-Net-DeA	0.862	0.559	0.678	0.801

4.2 Time Series Classification

In this experiment, we compare the classification performance with and without data augmentation. Specifically, we collect 5000 time series of one-week-long and 5-min-interval samples with binary class labels (seasonal or non-seasonal) from a cloud monitoring system in a top cloud service provider. The data is randomly splitted into training and test sets where training contains 80% of total samples. We train a fully convolutional network [Wang *et al.*, 2017] to classify each time series in the training set. In our experiment, we inject different types of outliers, including spike, step, and slope, into the test set to evaluate the robustness of the trained classifier. The data augmentations methods applied include cropping, warping, and flipping. Table 2 summarizes the accuracies with and without data augmentation when different types of outliers are injected into the test set. It can be observed that data augmentation leads to $0.1\% \sim 1.9\%$ accuracy improvement.

Table 2: Accuracy improvement under outlier injection in time series classification

Outlier injection	w/o aug	w/ aug	Improvement
spike	96.26%	96.37%	0.11%
step	93.70%	95.62%	1.92%
slope	95.84%	96.16%	0.32%

4.3 Time Series Forecasting

The forecasting task is an important application in time series analysis. Time series forecasting plays a crucial role in many real-world applications, including labor/machine scheduling, inventory control, dynamic pricing, etc. In this subsection we demonstrate the practical effectiveness of data augmentation techniques in three models, MQRNN [Wen *et al.*, 2017], DeepAR [Salinas *et al.*, 2019] and Transformer [Vaswani *et al.*, 2017]. In Table 3, we report the relative performance improvement on mean absolute scaled error (MASE) on several public datasets: *electricity* and *traffic* from UCI Learning Repository[1] and 3 datasets from the M4 competition[2]. We consider the basic augmentation methods aforementioned in subsection 2.1-2.2, including cropping, warping, flipping, and APP based augmentation in frequency domain.

In Table 3, we can obversve that the data augmentation methods give very promising results for all three models in average sense. However, the negative results can still be observed for specific data/model pairs. As a future work, it motivates us to search for improved data augmentation policies that stabilize the influence of data augmentation specifically for time series forecasting.

Table 3: Relative forecasting improvement based on MASE

Dataset	MQRNN	DeepAR	Transformer
electricity	-9%	1.92%	-2%
traffic	25%	-12%	-16%
m4-hourly	115%	56%	38%
m4-daily	0.1%	10%	37%
m4-weekly	-8%	76%	23%
average	24.62%	26.38%	16.00%

5 Discussion and Future Work

5.1 Augmentation in Time-Frequency Domain

As discussed in Section 2.3, so far there are only limited studies of time series data augmentation methods based on STFT in the time-frequency domain. Besides STFT, wavelet transform and its variants including continuous wavelet transform (CWT) and discrete wavelet transform (DWT), are another family of adaptive time–frequency domain analysis methods to characterize time-varying properties of time series. Compared to STFT, they can handle non-stationary time series and non-Gaussian noises more effectively and robustly.

Among many wavelet transform variants, maximum overlap discrete wavelet transform (MODWT) is especially attractive for time series analysis [Percival and Walden, 2000; Wen *et al.*, 2020] due to the following advantages: 1) more computationally efficient compared to CWT; 2) ability to handle any time series length; 3) increased resolution at coarser scales compared with DWT. MODWT based surrogate time series have been proposed in [Keylock, 2006], where wavelet iterative amplitude adjusted Fourier transform (WIAAFT) is designed by combining the iterative amplitude adjusted Fourier transform (IAAFT) scheme to each level of MODWT coefficients. In contrast to IAAFT, WIAAFT does not assume sationarity and can roughly maintain the shape of the

---

[1]http://archive.ics.uci.edu/ml/datasets.php

[2]https://github.com/Mcompetitions/M4-methods/tree/master/Dataset

original data in terms of the temporal evolution. Besides WIAAFT, we can also consider to perturb both amplitude spectrum and phase spectrum as [Gao *et al.*, 2020] at each level of MODWT coefficients as a data augmentation scheme. It would be an interesting future direction to investigate how to exploit MODWT for an effective time-frequency domain based time series data augmentation in deep neural networks.

5.2 Augmentation Combination

Given different data augmentation methods summarized in Fig. 1, one strategy is to combine various augmentation methods together and apply them sequentially. The experiments in [Um *et al.*, 2017] show that the combination of three basic time-domain methods (permutation, rotation, and time warping) is better than that of a single method and achieves the best performance in time series classification. Also, the results in [Rashid and Louis, 2019] demonstrate substantial performance improvement for a time series classification task when using a deep neural network by combining four data augmentation methods (i.e, jittering, scaling, rotation and time-warping). However, considering various data augmentation methods, directly combining different augmentations may result in a huge amount of data, and may not be efficient and effective for performance improvement.

Recently, RandAugment [Cubuk *et al.*, 2019b] is proposed as an effective and practical way for augmentation combination in image classification and object detection. For each random generated dataset, RandAugment is based on only 2 interpretable hyperparameters N (number of augmentation methods to combine) and M (magnitude for all augmentation methods), where each augmentation is randomly selected from K=14 available augmentation methods. Furthermore, this randomly combined augmentation with simple grid search can be used in the reinforcement learning based data augmentation as [Cubuk *et al.*, 2019a] to significantly reduce the search space. An interesting future direction is how to extend similar idea as in [Cubuk *et al.*, 2019a] to augmentation combination in time series data for deep learning.

5.3 Augmentation for Imbalanced Class

In time series classification, the existence of imbalanced class in which one class may occupy the majority of the datasets is a common problem. One classical approach addressing imbalanced classification problem is to oversample the minority class as the synthetic minority oversampling technique (SMOTE) [Fernández *et al.*, 2018] to artificially mitigate the imbalance. However, this oversampling strategy may change the distribution of raw data and cause overfitting. Another approach is to design cost-sensitive model by using adjust loss function [Geng and Luo, 2018]. Furthermore, [Gao *et al.*, 2020] designed label-based weight and value-based weight in the loss function in convolution neural networks, which considers weight adjustment for class labels and each sample and its neighborhood. Thus, both class imbalance and temporal dependency are explicitly considered.

Performing data augmentation and weighting for imbalanced class together would be an interesting and effective direction. A recent study investigates this topic in the area of CV and NLP [Hu *et al.*, 2019], which significantly improves text and image classification in low data regime and imbalanced class problems. In future, it would be an interesting direction on how to design deep network by jointly considering data augmentation and weighting for imbalanced class for time series data.

6   Conclusion

As deep learning models are becoming more popular on time series data, the limited labeled data calls for effective data augmentation methods. In this paper, we give a comprehensive survey on time series data augmentation methods in various tasks. We organize the reviewed methods in a taxonomy consisting of basic and advanced approaches. We introduce representative methods in each category and compare them empirical in typical time series tasks, including time series anomaly detection, time series forecasting and time series classification. Finally, we discuss and highlight future research directions by considering the limitation of current approaches and inspirations from other deep learning domains.

References

[Bergmeir *et al.*, 2016] Christoph Bergmeir, Rob J. Hyndman, and José M. Benítez. Bagging exponential smoothing methods using STL decomposition and Box–Cox transformation. *International Journal of Forecasting*, 32(2):303–312, 2016.

[Cao *et al.*, 2014] Hong Cao, Vincent YF Tan, and John ZF Pang. A parsimonious mixture of gaussian trees model for oversampling in imbalanced and multimodal time-series classification. *IEEE TNNLS*, 25(12):2226–2239, 2014.

[Chalapathy and Chawla, 2019] Raghavendra Chalapathy and Sanjay Chawla. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407*, 2019.

[Cleveland *et al.*, 1990] Robert B Cleveland, William S Cleveland, Jean E McRae, and Irma Terpenning. STL: A seasonal-trend decomposition procedure based on loess. *Journal of Official Statistics*, 6(1):3–73, 1990.

[Cubuk *et al.*, 2019a] Ekin D. Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V. Le. AutoAugment: Learning augmentation strategies from data. In *IEEE CVPR 2019*, pages 113–123, June 2019.

[Cubuk *et al.*, 2019b] Ekin D. Cubuk, Barret Zoph, Jonathon Shlens, and Quoc V. Le. RandAugment: Practical automated data augmentation with a reduced search space. *arXiv preprint arXiv:1909.13719*, 2019.

[Cui *et al.*, 2015] X Cui, V Goel, and B Kingsbury. Data augmentation for deep neural network acoustic modeling. *IEEE/ACM TASLP*, 23(9):1469–1477, 2015.

[Cui *et al.*, 2016] Zhicheng Cui, Wenlin Chen, et al. Multi-scale convolutional neural networks for time series classification. *arXiv preprint arXiv:1603.06995*, 2016.

[DeVries and Taylor, 2017] Terrance DeVries and Graham W. Taylor. Dataset augmentation in feature space. In *ICLR 2017*, pages 1–12, Toulon, 2017.

[Esteban *et al.*, 2017] Cristóbal Esteban, Stephanie L Hyland, and Gunnar Rätsch. Real-valued (medical) time series generation with recurrent conditional gans. *arXiv preprint arXiv:1706.02633*, 2017.

[Fawaz *et al.*, 2018] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, Lhassane Idoumghar, and Pierre-Alain Muller. Data augmentation using synthetic data for time series classification with deep residual networks. In *ECML/PKDD Workshop on AALTD*, 2018.

[Fawaz *et al.*, 2019] Hassan Ismail Fawaz, Germain Forestier, Jonathan Weber, et al. Deep learning for time series classification: a review. *Data Mining and Knowledge Discovery*, 33(4):917–963, 2019.

[Fernández *et al.*, 2018] Alberto Fernández, Salvador Garcia, Francisco Herrera, and Nitesh V Chawla. SMOTE for learning from imbalanced data: progress and challenges, marking the 15-year anniversary. *Journal of Artificial Intelligence Research*, 61:863–905, 2018.

[Gao *et al.*, 2020] Jingkun Gao, Xiaomin Song, Qingsong Wen, Pichao Wang, Liang Sun, and Huan Xu. RobustTAD: Robust time series anomaly detection via decomposition and convolutional neural networks. *arXiv preprint arXiv:2002.09535*, 2020.

[Geng and Luo, 2018] Yue Geng and Xinyu Luo. Cost-sensitive convolution based neural networks for imbalanced time-series classification. *arXiv preprint arXiv:1801.04396*, 2018.

[Han *et al.*, 2019] Z Han, J Zhao, H Leung, K F Ma, and W Wang. A review of deep learning models for time series prediction. *IEEE Sensors Journal*, page 1, 2019.

[Hu *et al.*, 2019] Zhiting Hu, Bowen Tan, Russ R Salakhutdinov, Tom M Mitchell, and Eric P Xing. Learning data manipulation for augmentation and weighting. In *NeurIPS 2019*, pages 15764–15775, 2019.

[Kang *et al.*, 2019] Yanfei Kang, Rob J Hyndman, and Feng Li. GRATIS: Generating time series with diverse and controllable characteristics. *arXiv preprint arXiv:1903.02787*, 2019.

[Kegel *et al.*, 2018] Lars Kegel, Martin Hahmann, and Wolfgang Lehner. Feature-based comparison and generation of time series. In *SSDBM 2018*, 2018.

[Keylock, 2006] C J Keylock. Constrained surrogate time series with preservation of the mean and variance structure. *Phys. Rev. E*, 73(3):36707, Mar 2006.

[Krizhevsky *et al.*, 2012] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NeurIPS 2012*, pages 1097–1105, 2012.

[Laptev *et al.*, 2015] Nikolay Laptev, Saeed Amizadeh, et al. Generic and scalable framework for automated time-series anomaly detection. *KDD*, pages 1939–1947, 2015.

[Le Guennec *et al.*, 2016] Arthur Le Guennec, Simon Malinowski, and Romain Tavenard. Data augmentation for time series classification using convolutional neural networks. In *ECML/PKDD Workshop on AALTD*, 2016.

[Lee *et al.*, 2019] Tracey Kah-Mein Lee, YL Kuah, Kee-Hao Leo, Saeid Sanei, Effie Chew, and Ling Zhao. Surrogate rehabilitative time series data for image-based deep learning. In *EUSIPCO 2019*, pages 1–5, 2019.

[Nikolaidis *et al.*, 2019] Konstantinos Nikolaidis, Stein Kristiansen, Vera Goebel, Thomas Plagemann, Knut Liestøl, and Mohan Kankanhalli. Augmenting physiological time series data: A case study for sleep apnea detection. In *ECML/PKDD 2019*, pages 1–17, 2019.

[Park *et al.*, 2019] Daniel S Park, William Chan, Yu Zhang, Chung-Cheng Chiu, Barret Zoph, Ekin D Cubuk, et al. SpecAugment: A simple data augmentation method for automatic speech recognition. In *INTERSPEECH 2019*, pages 2613–2617, 2019.

[Percival and Walden, 2000] Donald B Percival and Andrew T Walden. *Wavelet methods for time series analysis*, volume 4. Cambridge university press, New York, 2000.

[Rashid and Louis, 2019] Khandakar M Rashid and Joseph Louis. Times-series data augmentation and deep learning for construction equipment activity recognition. *Advanced Engineering Informatics*, 42:100944, 2019.

[Salinas *et al.*, 2019] David Salinas, Valentin Flunkert, Jan Gasthaus, and Tim Januschowski. DeepAR: Probabilistic forecasting with autoregressive recurrent networks. *International Journal of Forecasting*, 2019.

[Schreiber and Schmitz, 2000] Thomas Schreiber and Andreas Schmitz. Surrogate time series. *Physica D: Nonlinear Phenomena*, 142(3):346–382, 2000.

[Shorten and Khoshgoftaar, 2019] Connor Shorten and Taghi M Khoshgoftaar. A survey on image data augmentation for deep learning. *Journal of Big Data*, 6(1):60, 2019.

[Smyl and Kuber, 2016] Slawek Smyl and Karthik Kuber. Data preprocessing and augmentation for multiple short time series forecasting with recurrent neural networks. In *36th International Symposium on Forecasting*, June 2016.

[Steven Eyobu and Han, 2018] Odongo Steven Eyobu and Dong Seog Han. Feature representation and data augmentation for human activity classification based on wearable IMU sensor data using a deep LSTM neural network. *Sensors*, 18(9):2892, 2018.

[Um *et al.*, 2017] Terry T Um, Franz M J Pfister, Daniel Pichler, Satoshi Endo, Muriel Lang, Sandra Hirche, Urban Fietzek, and Dana Kulić. Data augmentation of wearable sensor data for Parkinson's disease monitoring using convolutional neural networks. In *ACM ICMI 2017*, pages 216–220, 2017.

[Vaswani *et al.*, 2017] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, et al. Attention is all you need. In *NeurIPS 2017*, pages 5998–6008, 2017.

[Wang *et al.*, 2017] Zhiguang Wang, Weizhong Yan, and Tim Oates. Time series classification from scratch with deep neural networks: A strong baseline. In *IJCNN*, pages 1578–1585, 2017.

[Wen and Keyes, 2019] Tailai Wen and Roy Keyes. Time series anomaly detection using convolutional neural networks and transfer learning. In *IJCAI Workshop on AI4IoT*, 2019.

[Wen *et al.*, 2017] Ruofeng Wen, Kari Torkkola, Balakrishnan Narayanaswamy, and Dhruv Madeka. A multi-horizon quantile recurrent forecaster. *arXiv preprint arXiv:1711.11053*, 2017.

[Wen *et al.*, 2019a] Qingsong Wen, Jingkun Gao, Xiaomin Song, Liang Sun, and Jian Tan. RobustTrend: A Huber loss with a combined first and second order difference regularization for time series trend filtering. In *IJCAI*, pages 3856–3862, 2019.

[Wen *et al.*, 2019b] Qingsong Wen, Jingkun Gao, Xiaomin Song, Liang Sun, Huan Xu, and Shenghuo Zhu. RobustSTL: A robust seasonal-trend decomposition algorithm for long time series. In *AAAI*, pages 1501–1509, 2019.

[Wen *et al.*, 2020] Qingsong Wen, Kai He, Liang Sun, Yingying Zhang, Min Ke, and Huan Xu. RobustPeriod: Time-frequency mining for robust multiple periodicities detection. *arXiv preprint arXiv:2002.09535*, 2020.

[Yoon *et al.*, 2019] Jinsung Yoon, Daniel Jarrett, and Mihaela van der Schaar. Time-series generative adversarial networks. In *NeurIPS 2019*, pages 5508–5518, 2019.