

Constraint Latent Space Matters: An Anti-anomalous Waveform Transformation Solution from Photoplethysmography to Arterial Blood Pressure

Cheng Bian*, Xiaoyu Li, Qi Bi, Guangpu Zhu, Jiegeng Lyu, Weile Zhang, Yelei Li, Zijing Zeng

OPPO Health Lab, Shenzhen, China

{biancheng, lixiaoyu5, zhuguangpu, lvjiegeng, zhangweile, liyelei1, zijing}@oppo.com, 2009biqu@163.com

Abstract

Arterial blood pressure (ABP) holds substantial promise for proactive cardiovascular health management. Notwithstanding its potential, the invasive nature of ABP measurements confines their utility primarily to clinical environments, limiting their applicability for continuous monitoring beyond medical facilities. The conversion of photoplethysmography (PPG) signals into ABP equivalents has garnered significant attention due to its potential in revolutionizing cardiovascular disease management. Recent strides in PPG-to-ABP prediction encompass the integration of generative and discriminative models. Despite these advances, the efficacy of these models is curtailed by the latent space shift predicament, stemming from alterations in PPG data distribution across disparate hardware and individuals, potentially leading to distorted ABP waveforms. To tackle this problem, we present an innovative solution named the Latent Space Constraint Transformer (LSCT), leveraging a quantized codebook to yield robust latent spaces by employing multiple discretizing bases. To facilitate improved reconstruction, the Correlation-boosted Attention Module (CAM) is introduced to systematically query pertinent bases on a global scale. Furthermore, to enhance expressive capacity, we propose the Multi-Spectrum Enhancement Knowledge (MSEK), which fosters local information flow within the channels of latent code and provides additional embedding for reconstruction. Through comprehensive experimentation on both publicly available datasets and a private downstream task dataset, the proposed approach demonstrates noteworthy performance enhancements compared to existing methods. Extensive ablation studies further substantiate the effectiveness of each introduced module.

Introduction

Cardiovascular disease (CVD) is a common cause of morbidity and mortality especially among the elderly population. Hypertension is both a major cause and a key indicator of CVD. Thus, it is of great importance to do continuous and regular blood pressure monitoring. However, the gold standard of measuring blood pressure (BP) is invasive and requires expert intervention (Ogedegbe and Pickering 2010). Although cuff-based BP devices have been developed, they still give rise to not only physical and psychologi-

cal discomfort for patients, but also inconvenience to do longitudinal monitoring (Shaltis, Reisner, and Asada 2006b). Compared to conventional BP measurement, invasive arterial blood pressure (ABP) can deliver thorough insights into fluctuations in blood pressure, bearing significant essence in disease diagnosis and medical research analysis (Avolio, Butlin, and Walsh 2009). However, due to the high requirement for sensor positioning, it is rarely utilized in clinical practice (Chen 2022).

Thanks to the rapid development of intelligent algorithms and smart hardware, in recent years, photoplethysmography (PPG) has emerged as a promising alternative for measuring blood flow velocity (Lee et al. 2014). In fact, as a simple, user-friendly and affordable approach, it has been widely used in devices such as smartwatches and fitness trackers. The key principle of PPG is to measure changes in the blood flow velocity, which correlates with the artery pressure. It has been revealed through some research (Reisner et al. 2008; Millasseau et al. 2000) that PPG and ABP signals share a common physiological factor, which can help facilitate the potential for mutual conversion.

Owing to the rapid development of deep learning techniques (Pan et al. 2022; Bian et al. 2021, 2020; Bi et al. 2022), cutting-edge research in the field can be summarized into two categories, namely, estimation from blood pressure features (Miao et al. 2019; Wu, Pang, and Kwong 2014) and waveform transformation (Mehrabadi et al. 2022; Vardhan et al. 2021; Ibtehaz et al. 2022). Estimation from blood pressure features intends to predict systolic and diastolic blood pressure, but it inevitably discards the ABP waveform and results in the loss of cardiovascular disease details. It not only degrades the accuracy of blood pressure estimation (Miao et al. 2019), but also negatively impacts its generalization to downstream tasks such as measuring cardiac capacity (Mukkamala et al. 2006).

To comprehensively perceive the cardiac process (Mukkamala, Stergiou, and Avolio 2022) and measure the blood pressure from PPG, a precise mapping from the peripheral blood volume changes to the arterial blood pressure waveform is critical. Several studies attempted to address this bottleneck by applying deep learning-based techniques (e.g., convolutional networks (Brophy et al. 2021), variational autoencoders (VAE) (Shimazaki et al. 2018), Transformers (Chu et al. 2023), and generative adversarial net-

*Corresponding author: biancheng@oppo.com

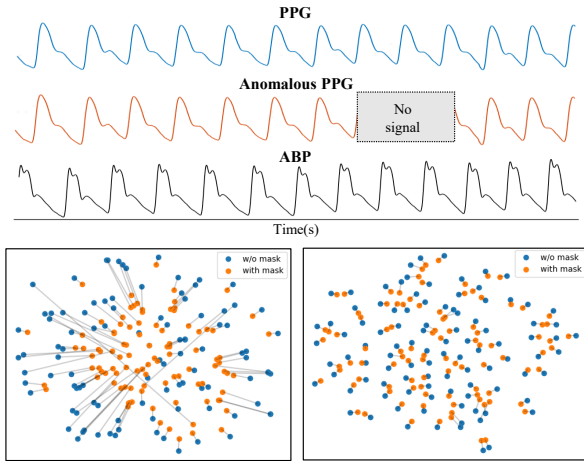


Figure 1: Top: Illustration of PPG, anomalous PPG and ABP signals. Bottom: Visualization of the latent space shift. Bottom Left: Visualization (t-SNE) results from naive discriminative model (a.k.a., Swin-Transformer (Liu et al. 2021)). Bottom Right: Results from ours. Note that blue points denote the original features from PPG signals in the latent space, while orange points are the features from anomalous PPG which is random masked under the ratio of 10%. Corresponding signals (w or w/o masks) are connected by gray lines, where the length of each line represents the distance.

works (Mehrabadi et al. 2022; Qin, Huang, and Zhang 2021)), as they have shown stronger representation learning capability than conventional analytical approaches (Shaltis, Reisner, and Asada 2006a). However, these methods highly rely on the consistency of waveform transformation. In practical scenarios such as wrist devices, anomalous PPG signals are captured commonly. Consequently, such pipelines inevitably trigger a rough alignment between the decoded waveform and the ground truth, resulting in an unstable shift within the latent space as illustrated in Fig. 1. Overall, how to learn a precise PPG to ABP mapping from anomalous signal transformation remains challenging.

To address this issue, we present a novel approach named the Latent Space Constraint Transformer (LSCT) for anomalous PPG-to-ABP waveform conversion. In general, the codebook generated by Vector Quantized-Variational AutoEncoder (VQ-VAE) (Van Den Oord, Vinyals et al. 2017) can learn a relatively stable latent space by discretizing the waveform features into several codebook bases. This property is particularly valuable for scenarios dealing with information loss (Chen et al. 2020). However, the representation from the conventional VQ-VAE paradigm highly relies on several bases that have been discretized. It is less expressive for our anomalous PPG-to-ABP waveform conversion task, where a certain defective basis can significantly degrade the feature representation.

To overcome this limitation, we propose a novel Correlation-boosted Attention Module (CAM), which enables comprehensive querying of all relevant bases and allows the representation to be more stable even if some bases

are defective. It facilitates to seek a better substitution of defective basis from a global view. Moreover, we speculate that establishing a spectrum-wise information flow can improve the representation of dependent bases derived from the codebook. In light of this assumption, we propose an innovative Multi-Spectrum Enhancement Knowledge (MSEK) approach to incorporate a localized graph into the relevant bases. Consequently, the graph embedding is able to retrieve crucial details from the spectrum (a.k.a., channel) in the decoding stage while effectively reducing the distortion from ABP construction.

Our contribution can be summarized as follows:

- 1) Our work reveals the latent space shift in the defective waveform transformation, which degrades the conversion process. To address this challenge, we propose a novel Latent Space Constraint Transformer (LSCT) to quantize the latent space and reduce the distortion from anomalous PPG to ABP.
- 2) We propose a correlation-boosted attention module to query the highly relevant bases from a global view. It is capable of constructing comprehensive features via anomalous signals.
- 3) We propose a novel multi-spectrum enhancement knowledge, which treats each channel-wise feature as a node to construct a local knowledge graph flow. It boosts the expressiveness of each latent code and enhances the fidelity of the waveform conversion.
- 4) Through extensive experiments and analysis on MIMIC III and VitalDB, our framework can address the latent space shift well and achieves state-of-the-art performance.

Related Works

Discriminative Waveform Transformation

Discriminative waveform transformation approaches project the original waveform to the translating representation by leveraging either CNN, RNN, or Transformer architectures (Ji et al. 2021, 2023b,a). For CNN-based methods, the U-Net architecture has been reported as applicable to reconstructing one-dimensional signals. For example, Athaya *et al.* (Athaya and Choi 2021) utilized U-Net to reconstruct a non-invasive ABP signal from PPG signal. PPG2ABP (Ibtehaz et al. 2022) employed a two-stage cascaded U-Net with deep supervision to reconstruct and refine ABP waveform from PPG signal. However, the conventional convolution operations in the U-Net suffer from information loss and gradient vanishing, which undermines the representation from PPG signal (Zhang et al. 2017).

RNN-based methods have also been extensively investigated (Harfiya, Chang, and Li 2021). However, they rely too much on hidden memory states and thus negatively impact the temporal consistency (Bernard, Msigwa, and Yun 2022). More recently, Transformer has been reported effective for this task (Lan 2023). However, well-trained discriminative methods will establish a static latent space, necessitating a determined representation and thereby not readily extensible to anomalous signal processing.

Generative Waveform Transformation

Generative Adversarial Networks (GANs) (Goodfellow et al. 2014) have been widely used to translate ABP waveform from PPG signal (Mehrabadi et al. 2022). Recent studies step forward by using more advanced backbones for adversarial training. For instance, Esteban et al. (Esteban, Hyland, and Rättsch 2017) employed Recurrent GAN (RGAN) and Recurrent Conditional GAN (RCGAN) to produce realistic multi-dimensional time-series signals. TimeGAN (Yoon, Jarrett, and Van der Schaar 2019) integrated an unsupervised GAN framework into an autoregressive model to jointly optimize the learning of the embedding space. SigCWGAN (Ni et al. 2020) leveraged the path signature to capture the temporal dependencies. More recently, to effectively handle the longer sequences, TTS-CGAN (Li, Ngu, and Metsis 2022) introduced the Transformer model as a feature extractor. However, GAN models overemphasize the consistency of input and output, and thus tend to overlook the importance of the representation, which is detrimental to the task-specific waveform transformation. Moreover, these methods also struggle with the common challenges to train GANs, such as non-convergence, mode collapse, diminished gradient and etc.

In recent years, frontier generative models like vector quantization (VQ) models and diffusion models (Ho, Jain, and Abbeel 2020) have shown the potential (Van Den Oord, Vinyals et al. 2017). Time-VQVAE was the first to introduce VQ-VAE into time series generation, capturing global temporal consistency (Lee, Malacarne, and Aune 2023).

Method

Overview Architecture

The overall framework of the proposed LSCT model is shown in Fig. 2. A codebook is served as the discretized basis dictionary to support sparse representation learning for transforming PPG waveform to ABP.

Given a sample pair of PPG and ABP (x_p, \hat{x}_a) , a Short-time Fourier Transformation (STFT) is applied to transform the pair to time-frequency domain and obtain (u_p, \hat{u}_a) . The \hat{x}_a and \hat{u}_a serve as ground truth for the framework training, and u_p is sent to the encoder as our model input. The encoder \mathcal{F} is parameterized by δ , and the decoder \mathcal{H} is parameterized by φ . The encoder compresses the signal and generates a latent code \mathbf{z}_q :

$$\mathbf{z}_q = \mathcal{F}_\delta(u_p). \quad (1)$$

Then, in the CAM, the latent code \mathbf{z}_q is used to query bases from the codebook to form the attentional representation \mathbf{z}_v . In the MSEK, spectrum of \mathbf{z}_v are fused to produce the comprehensive representation \mathbf{z}_g . The summation of \mathbf{z}_q and \mathbf{z}_g is then sent to the decoder to acquire the transformed ABP \hat{u}_a in the time-frequency domain. The \hat{x}_a in the time domain can be computed by Inverse Short-time Fourier Transform:

$$\hat{u}_a = \mathcal{H}_\varphi(\underbrace{\text{MSEK}(\text{CAM}(\mathbf{z}_q))}_{\mathbf{z}_g} + \mathbf{z}_q). \quad (2)$$

The objective function is estimated to optimize all parameters by the mean square error in both the time and time-frequency domain:

$$\mathcal{L} = \|x_a - \hat{x}_a\|_2^2 + \|u_a - \hat{u}_a\|_2^2. \quad (3)$$

Correlation-boosted Attention Module

Compared to typical VQ-VAE alternatives applied in computer vision tasks, PPG-to-ABP waveform transformation task requires fine-grained feature modeling. Thus, the core concept of CAM is to learn an attentional representation by comprehensively querying and combining all relevant discretized bases. For this reason, we design a cross-attention mechanism for querying relative bases from the codebook. The codebook is a matrix $\mathbf{M} \in \mathcal{R}^{m \times d}$, where m denotes the codebook size (i.e., number of bases) and d denotes the codebook dimension (i.e., basis spectrum). The following experiment will discuss and analyze the influence of m and d .

Note the PPG signal latent code as $\mathbf{z}_q \in \mathcal{R}^{b \times n \times d}$, b denotes batch size and n denotes feature map length. Cross attention is designed to query and combine different bases from the codebook, where latent code \mathbf{z}_q serves as *Query*, \mathbf{M} serves as *Key* and *Value*:

$$\mathbf{z}_v = \text{softmax}\left(\frac{\mathbf{z}_q \mathbf{M}^T}{\sqrt{d}}\right) \mathbf{M}. \quad (4)$$

Multi-Spectrum Enhancement Knowledge

Based on the assumption that spectrum-wise information flow improves the representation of dependent bases from the codebook, we propose the multi-spectrum enhancement knowledge module to fuse channel-wise information by graph modeling. The graph consists of nodes and edges, $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The channels of \mathbf{z}_v is treated as the nodes $\mathcal{V} = (\mathbf{h}_1, \mathbf{h}_2, \dots, \mathbf{h}_d)$, $\mathbf{h}_i \in \mathcal{R}^{b \times n}$. For the edges, a fully connected adjacent matrix is initialized and half of the matrix values are randomly masked with zeros. An attention coefficient can be calculated for the edge e_{ij} :

$$e_{ij} = \mathbf{W} \mathbf{h}_i (\mathbf{W} \mathbf{h}_j)^T / \sqrt{n}, \quad (5)$$

where \mathbf{W} is a learnable weight shared by all nodes. The coefficient indicates the importance of \mathbf{h}_j to \mathbf{h}_i . To further consider the graph structure, the coefficients between node \mathbf{h}_i and its first-order neighbors \mathcal{N}_i are computed and normalized:

$$\alpha_{ij} = \frac{\exp(e_{ij})}{\sum_{k \in \mathcal{N}_i} \exp(e_{ik})}. \quad (6)$$

We can treat $\mathbf{W} \mathbf{h}_j$ as *Value* and update node feature \mathbf{h}_i with multi-head self-attention:

$$\hat{\mathbf{h}}_i = \frac{1}{K} \sum_{k=1}^K \sum_{j \in \mathcal{N}_i} \alpha_{ij}^k \mathbf{W}^k \mathbf{h}_j, \quad (7)$$

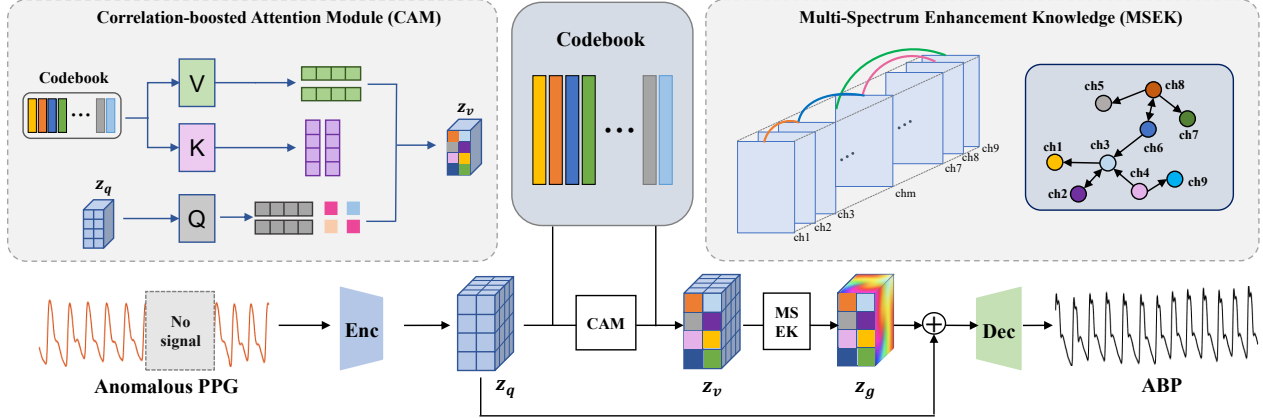


Figure 2: Overview of our proposed Latent Space Constraint Transformer (LSCT) framework. Firstly, anomalous PPG waveforms are used as inputs fed to the encoder. Then, the proposed Correlation-booster Attention Module (CAM) queries relative bases to form the attentional representation z_v . Afterward, Multi-Spectrum Enhancement Knowledge (MSEK) constructs spectrum-wise graph flow from the comprehensive representation z_g . Finally, the ABP waveforms are decoded by the summation of the latent code z_q and the comprehensive representation z_g .

Item	MIMIC III	VitalDB	OML
Number of subjects	942	144	20
Number of segments	127,260	523,080	24,234
Average SBP (mmHg)	134.19	120.38	120.84
SD of SBP (mmHg)	22.93	19.63	12.19
Average DBP (mmHg)	66.14	60.21	79.10
SD of DBP (mmHg)	11.45	12.60	8.52

Table 1: Summary of statistical information of the proposed datasets.

where K is the head number. Then the node features evolve as the comprehensive representation z_g after the graph information aggregation.

Experiment

Datasets and Evaluation Metrics

In the experiment, we evaluate our method on two public datasets and one private dataset. Particularly, public datasets including MIMIC-III and VitalDB are curated for the wave transformation task. Meanwhile, the private OML dataset is developed for the BP values regression, which is the sub-task of PPG-to-ABP wave transformation used to validate the generalization of our method. Statistical details on three datasets can be referred to Table 1.

MIMIC-III. As one of the critical datasets, MIMIC-III is conducted by using a processed subset of the Medical Information Mart for Intensive Care (MIMIC-III) (Johnson et al. 2016; Johnson, Pollard, and Mark 2016; Goldberger et al. 2000). It encompasses a collection of 12,000 PPG,

ECG, and ABP signals, which are synchronously sampled from 942 patients in practical clinical environments. All signals are sampled at 125 Hz via filtering and preprocessing. Specifically, we segment ABP signals with a length of 8.192 seconds, each slice containing 1,024 points. Likewise, we employ the same operation on the PPG signal and align it to the sampled ABP signals. Overall, 127,260 ABP and PPG segment pairs are selected for evaluation.

VitalDB. The VitalDB dataset includes synchronized ECG, PPG, and ABP recordings (Lee et al. 2022). It comprises 6,388 records from 6,090 ICU patients who underwent surgeries at the Seoul National University Hospital. This dataset provides a substantial collection of over 16,000 PPG and ECG signals, randomly selected from 550 subjects, with a duration of 8.192 seconds (i.e., 1,024 points) and sampled at a frequency of 125 Hz.

OML. Our private dataset, denoted as OML, consists of 24,234 systolic blood pressure (SBP) and diastolic blood pressure (DBP) value pairs over 20 subjects, where SBP and DBP substitute the ABP waveform records in this dataset for longitudinal tracking efficiency. All SBP and DBP values are sampled from the high-precision blood pressure monitor, whereas PPG signals are captured by our professional wearable device. The collection period spans a duration of at least 30 days for validity purposes. We manually align an SBP-DBP value pair for one PPG record. To ensure consistency in the sampling rate and segment length across the PPG dataset, we downsample the PPG signals to 125Hz and divide them into multiple segments of equal length with 1,024 for analysis.

Masked Strategy. The random masked strategy will be implemented on the PPG signal in both of the mentioned datasets. For example, a 50% masked ratio means that a zero

Methods	MIMIC III			VitalDB		
	RMSE ↓	PRD ↓	FD ↓	RMSE ↓	PRD ↓	FD ↓
<i>Discriminative Waveform Transformation</i>						
LSTM (Harfiya, Chang, and Li 2021)	0.156	3.937	3.711	0.033	0.504	0.919
Unet (Athaya and Choi 2021)	0.124	3.618	3.427	0.020	0.311	0.745
PPG2ABP (Ibtehaz et al. 2022)	0.101	3.743	3.428	0.019	0.281	0.729
<i>Generative Waveform Transformation</i>						
CycleGAN (Mehrabadi et al. 2022)	0.163	4.825	4.809	0.048	0.706	1.520
TTS-CGAN (Li, Ngu, and Metsis 2022)	0.150	4.062	3.869	0.032	0.475	1.036
VQ-VAE (Van Den Oord, Vinyals et al. 2017)	0.108	3.873	3.693	0.019	0.285	0.641
Time-VQVAE (Lee, Malacarne, and Aune 2023)	0.109	3.877	3.701	0.020	0.298	0.659
Ours	0.097	3.402	3.107	0.016	0.258	0.566

Table 2: Quantitative comparisons with the state-of-the-art methods on MIMIC III dataset and VitalDB dataset. “↓” indicates that metrics with lower values correspond to better performance.

mask, with a size equal to half the length of the original series, will be applied to modify the PPG series.

Evaluation Metrics. Root Mean Squared Error (RMSE), Percentage Root Mean Squared Difference (PRD) and Fréchet Distance (FD) (Eiter and Mannila 1994) are commonly utilized to measure the waveform transformation performance (Zhu et al. 2019; Sarkar and Etemad 2021). We use the three metrics to compare different models and variants. RMSE measures the stability between two signals S and \hat{S} :

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (S_i - \hat{S}_i)^2}. \quad (8)$$

PRD measures the distortion between the signals:

$$PRD = \sqrt{100 \times \frac{\sum_{i=1}^N (S_i - \hat{S}_i)^2}{\sum_{i=1}^N (S_i)^2}}. \quad (9)$$

FD distance measures the similarity between the two signals, considering the location and the order of data points:

$$FD = \min_M \max_{(s, \hat{s}) \in M} \|s - \hat{s}\|_2^2, \quad (10)$$

where s and \hat{s} are the data points in the signal S and \hat{S} respectively, and M is the matrix $\{(s, \hat{s}) | s \in S, \hat{s} \in \hat{S}\}$. Besides, for the downstream regression task, we utilize mean average error (MAE), mean error (ME) and standard deviation (SD) as metrics.

Implementation Details

Our algorithm is implemented using the PyTorch framework (Paszke et al. 2019). The encoder \mathcal{F} and decoder \mathcal{H} utilized in our study are implemented with four identical

blocks of Swin Transformer (Liu et al. 2021). The encoder blocks had channel sizes of 8, 16, 32, and 64 respectively, while the decoder blocks had channel sizes of 64, 32, 16, and 8 correspondingly. Downsample block in the encoder and upsample block in the decoder are implemented with PatchMerging (Liu et al. 2021) and transpose convolution layer (Zeiler et al. 2010). For the CAM module, the size of the codebook m is set to 128, and the codebook dimension d is configured as 64. Random initialization is applied. For the MSEK module, n is set to 40 and K is set to 8. We utilize the vanilla Adam optimizer (Kingma and Ba 2015) to update all parameters, with the learning rate of 1e-3, weight decay of 1e-5, and batch size b of 512. The cosine annealing schedule (Loshchilov and Hutter 2016) is employed.

Ablation Studies

To evaluate the effectiveness of the proposed modules individually, we conduct comprehensive ablation studies on two datasets, as listed in Table 3, in which four types of module ablations, as well as the baseline denoted as the Swin-Transformer, will be included. Note that baseline is the vanilla VQ-VAE using K-nearest neighbors (Kramer and Kramer 2013) as the basis selection. Other attempts include our proposed CAM and MSEK modules. As we can see, 12.5%, 6.6%, 7.1% decreases have been achieved in RMSE, PRD and FD respectively when involving the CAM module, demonstrating that our method can facilitate codebook selection and query valuable basis for the waveform transformation. After that, notable overall performance gains have been observed by introducing the MSEK module, verifying that constructing graph flow in the channel can contribute to enhancing the expressiveness of the basis. Furthermore, data ablations are being performed as shown in Table 3. As the increasing masked ratio MR , the performance of baseline in RMSE, PRD and FD decreases in the average of 111.8%, 97.7% and 128.7%, respectively. On the contrary, our method surpasses the baseline under to the same MR by a large margin with average improvements of 101.6%, 89.3% and 88.3% in RMSE, PRD and FD, suggesting that

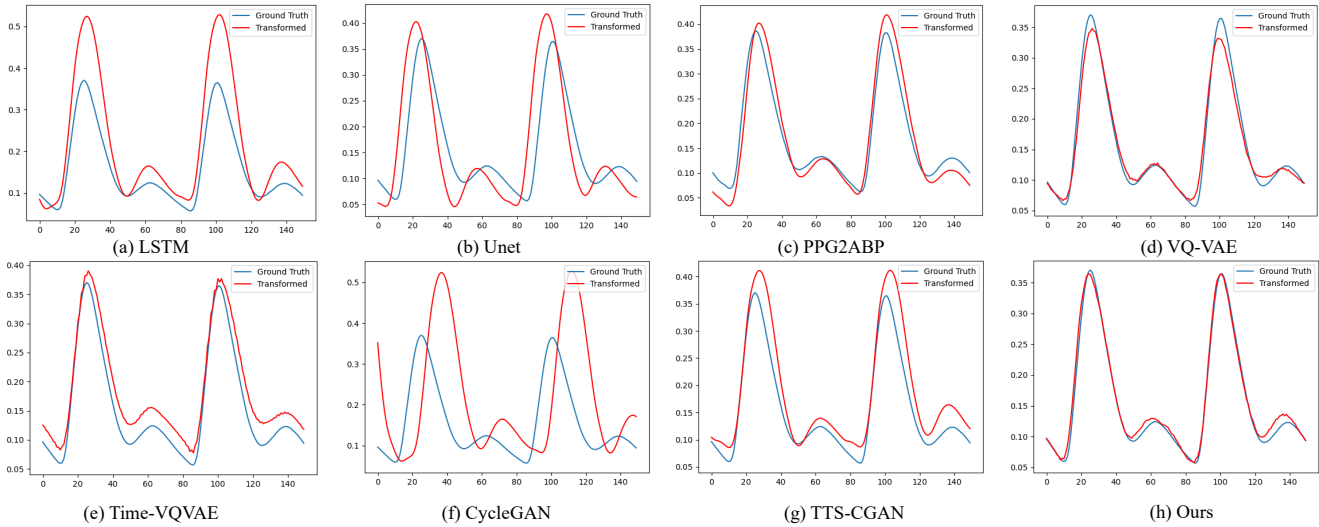


Figure 3: Qualitative comparisons of ours with the state-of-the-art methods on the MIMIC-III dataset, using the mask ratio of 10%.

Methods	CAM	MSEK	MR	RMSE	PRD	FD
Baseline			10%	0.019	0.285	0.641
Module ablations		✓	10%	0.018	0.275	0.606
	✓	✓	10%	0.016	0.258	0.566
Data ablations			30%	0.025	0.342	0.841
		✓	30%	0.022	0.322	0.798
		✓	50%	0.030	0.428	0.882
		✓	50%	0.026	0.379	0.833
		✓	70%	0.037	0.557	1.635
		✓	70%	0.034	0.505	1.112
		✓	90%	0.052	0.745	1.903
	✓	✓	90%	0.047	0.692	1.520

Table 3: Ablation studies on the VitalDB dataset. Note that the mask ratio on the data ablations ranging from 10% to 90%.

our method has the ability to deal with the anomalous signals transformation effectively and robustly. From another perspective, Fig. 4 illustrates a fair comparison of RMSE from typical state-of-the-art generative model, discriminative model and ours given by different *MRs*. It can also be noticed that our proposed method is anti-anomalous, surpassing frontier robust methods.

Investigation on MSEK

To investigate which information flow can promote expressiveness of basis, we conduct an experiment on MIMIC III and analyze its impacts on different variants. Quantitative re-

Graph Flow Variants	RMSE	PRD	FD
Self-attention (SA)	0.127	4.457	4.129
Graph-based Codebook (GBC)	0.138	4.562	4.351
Ours	0.097	3.402	3.107

Table 4: Investigation of graph flow variants of the MIMIC III datasets.

sults are shown in Table 4. One of the compared methods is self-attention denoted as SA, which directly utilizes the attentional representation as query, key and value to compute the spectrum-wise weights on each basis. The other compared method is the graph-based codebook denoted as GBC, which constructs a codebook graph on relative bases. Compared to our method, it is noticed that the methods of SA encounter performance drops of 30.9%, 31.0% and 32.9%, meanwhile GBC has the performance drops of 42.2%, 34.1% and 40.0% in RMSE, PRD and FD, respectively. The underlying reason is the way of self-attention employing the global graph for the computation, where not every channel of the basis is critical for the expression of the waveform transformation. Besides, GBC builds a computational graph between basis, which does harm to the orthogonality of the basis from the discretized codebook. In contrast, our methods can not only maintain the information flow within spectrum space, but also ensure the orthogonality between bases.

Impact of the Size of Codebook and Dimension of Bases

To understand whether the size of a codebook will affect transformation performance, we employed various codebook sizes (64, 128, 256 and 512) to ascertain its effect on the transformation performance. The result is plotted in

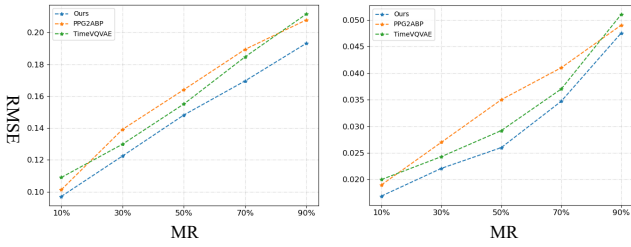


Figure 4: RMSE comparison of typical state-of-the-art methods and ours given by different MRs.

Fig. 5. It can be seen that variation in size settings has a negligible impact on the performance of our methodology. The dimension of bases will influence the construction of the knowledge graph flow. Therefore, another experiment is performed to track the performance manifestations under various dimensions of bases. As shown in Fig. 5, we achieve the best performance by the optimal dimension d of 64.

Comparison with State-of-the-arts

To validate the efficacy of the proposed method, a comparison is made between the proposed framework and competitive methods. For a fair comparison, three representative discriminative waveform transformation methods and four generative waveform transformation methods are chosen in our experiment under the same experimental configuration and masked ratio (10%). Table 2 summarizes the experimental results of the proposed method. The proposed method yields the best against all compared methods. In particular, our method has shown a performance improvement of at least 4.1% in RMSE compared to discriminative methods, indicating that our method exhibits the capability to mitigate and overcome latent space shift ensuring robust and reliable performance. Similarly, our model’s performance has surpassed that of generative methods in both criteria. This observation further corroborates that a stable latent space can effectively eliminate distortions during wave transformation. From Fig. 3(a), (b), (c) and (h), we note that the performance of PPG2ABP surpasses that of Unet and LSTM. The reason is PPG2ABP contains a cascaded Unet model, which allows a coarse-to-fine refinement procedure so that it alleviates anomalous PPG signals to some extent. As to Fig. 3(d), (e), (f) and (g), we found that the overall performance of VQ-VAE alternatives is better than GANs. It substantiates our view that representation is critical to waveform transformation. Contrasting to these approaches, CAM and MSEK modules offer attentional representation and spectrum-wise information flow, thereby effectively mitigating the latent space shift and yield optimal performance.

Generalization of Downstream Task

As presented in Table 5, an additional OML dataset has been used to investigate the SBP and DBP predictions in longitudinal tracking. Additionally, typical frontier discriminative methods including InceptionTime (Ismail Fawaz et al. 2020), ResNeXt-101 (Xie et al. 2017) and CoAtNet (Dai

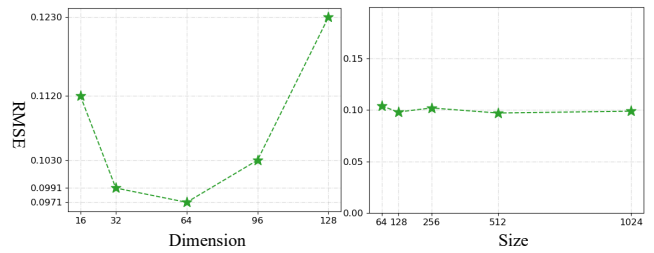


Figure 5: Performance with varied size of codebook and dimension of bases.

Methods	SBP		DBP	
	MAE	ME \pm STD	MAE	ME \pm STD
InceptionTime	15.51	-4.70 \pm 17.48	9.10	-3.80 \pm 10.56
ResNeXt-101	11.36	-1.72 \pm 12.48	8.31	-0.46 \pm 9.61
CoAtNet	10.79	-1.61 \pm 11.26	7.02	0.76 \pm 8.05
SwinTransformer	11.17	-1.08 \pm 11.83	7.35	0.79 \pm 9.03
PPG2ABP	9.18	-1.74 \pm 9.50	7.29	0.36 \pm 7.80
Time-VQVAE	8.92	-0.89 \pm 8.32	6.87	-0.48 \pm 7.02
VQ-VAE	8.63	-1.37 \pm 8.35	6.73	-0.82 \pm 6.93
Ours	8.04	-1.19\pm8.18	6.49	-0.72\pm6.75

Table 5: Generalization of downstream task on the private OML dataset.

et al. 2021), will be involved for the downstream task comparison. As we can see, our framework improves upon the state-of-the-art models remarkably. It is obvious that our proposed modules not only effectively work on the short-term waveform transformation task but also shed light on their potential to address the BP longitudinal tracking.

Conclusion

This paper first speculates the latent space shift in anomalous PPG-to-ABP waveform transformation. Afterward, we propose a Latent Space Constraint Transformer (LSCT) to address this problem. Firstly, we utilize discretized codebook to quantize the latent space. Secondly, the Correlation-boosted Attention Module (CAM) has been proposed to correct the selection bias of codebook, which can substantially diminish the latent space shift from the anomalous inputs. Lastly, a Multi-Spectrum Enhancement Knowledge (MSEK) is introduced to present graph embedding in the channel view, which promotes the expressiveness of the latent code. Our method achieved competitive performance compared with the existing state-of-the-art waveform transformation methods. It also possesses substantial possibilities for the generalization of downstream tasks like BP longitudinal tracking. In the future, more endeavors will be made to explore its efficacy across broader waveform transformation tasks.

References

- Athaya, T.; and Choi, S. 2021. An estimation method of continuous non-invasive arterial blood pressure waveform using photoplethysmography: A U-Net architecture-based approach. *Sensors*, 21(5): 1867.
- Avolio, A. P.; Butlin, M.; and Walsh, A. 2009. Arterial blood pressure measurement and pulse wave analysis—their role in enhancing cardiovascular assessment. *Physiological measurement*, 31(1): R1.
- Bernard, D.; Msigwa, C.; and Yun, J. 2022. Toward IoT-Based Medical Edge Devices: PPG-Based Blood Pressure Estimation Application. *IEEE Internet of Things Journal*, 10(6): 5240–5255.
- Bi, Q.; Zhou, B.; Qin, K.; Ye, Q.; and Xia, G.-S. 2022. All grains, one scheme (AGOS): Learning multigrain instance representation for aerial scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 60: 1–17.
- Bian, C.; Yuan, C.; Ma, K.; Yu, S.; Wei, D.; and Zheng, Y. 2021. Domain adaptation meets zero-shot learning: an annotation-efficient approach to multi-modality medical image segmentation. *IEEE Transactions on Medical Imaging*, 41(5): 1043–1056.
- Bian, C.; Yuan, C.; Wang, J.; Li, M.; Yang, X.; Yu, S.; Ma, K.; Yuan, J.; and Zheng, Y. 2020. Uncertainty-aware domain alignment for anatomical structure segmentation. *Medical Image Analysis*, 64: 101732.
- Brophy, E.; De Vos, M.; Boylan, G.; and Ward, T. 2021. Estimation of continuous blood pressure from PPG via a federated learning approach. *Sensors*, 21(18): 6311.
- Chen, M.; Radford, A.; Child, R.; Wu, J.; Jun, H.; Luan, D.; and Sutskever, I. 2020. Generative pretraining from pixels. In *ICML*, 1691–1703. PMLR.
- Chen, Y. 2022. Estimating Cuff-less Continuous Blood Pressure from Fingertip Photoplethysmogram Signals with Deep Neural Network Model. *medRxiv*, 2022–03.
- Chu, Y.; Tang, K.; Hsu, Y.-C.; Huang, T.; Wang, D.; Li, W.; Savitz, S. I.; Jiang, X.; and Shams, S. 2023. Non-invasive arterial blood pressure measurement and SpO₂ estimation using PPG signal: a deep learning framework. *BMC Medical Informatics and Decision Making*, 23(1): 131.
- Dai, Z.; Liu, H.; Le, Q. V.; and Tan, M. 2021. Coatnet: Marrying convolution and attention for all data sizes. *Advances in neural information processing systems*, 34: 3965–3977.
- Eiter, T.; and Mannila, H. 1994. Computing discrete Fréchet distance.
- Esteban, C.; Hyland, S. L.; and Rätsch, G. 2017. Real-valued (medical) time series generation with recurrent conditional gans. *arXiv preprint arXiv:1706.02633*.
- Goldberger, A. L.; Amaral, L. A.; Glass, L.; et al. 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation*, 101(23): e215–e220.
- Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; and Bengio, Y. 2014. Generative adversarial nets. *Advances in neural information processing systems*, 27.
- Harfiya, L. N.; Chang, C.-C.; and Li, Y.-H. 2021. Continuous blood pressure estimation using exclusively photoplethysmography by LSTM-based signal-to-signal translation. *Sensors*, 21(9): 2952.
- Ho, J.; Jain, A.; and Abbeel, P. 2020. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33: 6840–6851.
- Ibtihaz, N.; Mahmud, S.; Chowdhury, M. E.; Khandakar, A.; Salman Khan, M.; Ayari, M. A.; Tahir, A. M.; and Rahman, M. S. 2022. PPG2ABP: Translating photoplethysmogram (PPG) signals to arterial blood pressure (ABP) waveforms. *Bioengineering*, 9(11): 692.
- Ismail Fawaz, H.; Lucas, B.; Forestier, G.; Pelletier, C.; Schmidt, D. F.; Weber, J.; Webb, G. I.; Idoumghar, L.; Muller, P.-A.; and Petitjean, F. 2020. Inceptiontime: Finding alexnet for time series classification. *Data Mining and Knowledge Discovery*, 34(6): 1936–1962.
- Ji, W.; Li, J.; Bian, C.; Zhang, Z.; and Cheng, L. 2023a. SemanticRT: A Large-Scale Dataset and Method for Robust Semantic Segmentation in Multispectral Images. In *Proceedings of the 31st ACM International Conference on Multimedia*, 3307–3316.
- Ji, W.; Li, J.; Bian, C.; Zhou, Z.; Zhao, J.; Yuille, A. L.; and Cheng, L. 2023b. Multispectral Video Semantic Segmentation: A Benchmark Dataset and Baseline. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 1094–1104.
- Ji, W.; Yu, S.; Wu, J.; Ma, K.; Bian, C.; Bi, Q.; Li, J.; Liu, H.; Cheng, L.; and Zheng, Y. 2021. Learning calibrated medical image segmentation via multi-rater agreement modeling. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 12341–12351.
- Johnson, A.; Pollard, T.; and Mark, R. 2016. MIMIC-III clinical database (version 1.4). *PhysioNet*, 10(C2XW26): 2.
- Johnson, A. E.; Pollard, T. J.; Shen, L.; Lehman, L.-w. H.; Feng, M.; Ghassemi, M.; Moody, B.; Szolovits, P.; Anthony Celi, L.; and Mark, R. G. 2016. MIMIC-III, a freely accessible critical care database. *Scientific data*, 3(1): 1–9.
- Kingma, D.; and Ba, J. 2015. Adam: A method for stochastic optimization in: Proceedings of the 3rd international conference for learning representations (iclr’15). *San Diego*, 500.
- Kramer, O.; and Kramer, O. 2013. K-nearest neighbors. *Dimensionality reduction with unsupervised nearest neighbors*, 13–23.
- Lan, E. 2023. Performer: A Novel PPG-to-ECG Reconstruction Transformer for a Digital Biomarker of Cardiovascular Disease Detection. In *WACV*, 1991–1999.
- Lee, D.; Malacarne, S.; and Aune, E. 2023. Vector Quantized Time Series Generation with a Bidirectional Prior Model. *arXiv preprint arXiv:2303.04743*.
- Lee, H.-C.; Park, Y.; Yoon, S. B.; Yang, S. M.; Park, D.; and Jung, C.-W. 2022. VitalDB, a high-fidelity multi-parameter vital signs database in surgical patients. *Scientific Data*, 9(1): 279.
- Lee, Y.-J.; Lee, C.-K.; Kang, M.; Kang, S.-J.; et al. 2014. Magneto-plethysmographic sensor for peripheral blood flow velocity. *IEEE Sensors journal*, 14(5): 1341–1342.

- Li, X.; Ngu, A. H. H.; and Metsis, V. 2022. Tts-cgan: A transformer time-series conditional gan for biosignal data augmentation. *arXiv preprint arXiv:2206.13676*.
- Liu, Z.; Lin, Y.; Cao, Y.; et al. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, 10012–10022.
- Loshchilov, I.; and Hutter, F. 2016. Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.
- Mehrabadi, M. A.; Aqajari, S. A. H.; Zargari, A. H. A.; Dutt, N.; and Rahmani, A. M. 2022. Novel blood pressure waveform reconstruction from photoplethysmography using cycle generative adversarial networks. In *EMBC*, 1906–1909. IEEE.
- Miao, F.; Liu, Z.-D.; Liu, J.-K.; et al. 2019. Multi-sensor fusion approach for cuff-less blood pressure measurement. *IEEE journal of biomedical and health informatics*, 24(1): 79–91.
- Millasseau, S. C.; Guigui, F. G.; Kelly, R. P.; Prasad, K.; Cockcroft, J. R.; Ritter, J. M.; and Chowienczyk, P. J. 2000. Noninvasive assessment of the digital volume pulse: comparison with the peripheral pressure pulse. *Hypertension*, 36(6): 952–956.
- Mukkamala, R.; Reisner, A. T.; Hojman, H. M.; et al. 2006. Continuous cardiac output monitoring by peripheral blood pressure waveform analysis. *IEEE Transactions on Biomedical Engineering*, 53(3): 459–467.
- Mukkamala, R.; Stergiou, G. S.; and Avolio, A. P. 2022. Cuffless blood pressure measurement. *Annual Review of Biomedical Engineering*, 24: 203–230.
- Ni, H.; Szpruch, L.; Wiese, M.; Liao, S.; and Xiao, B. 2020. Conditional sig-wasserstein gans for time series generation. *arXiv preprint arXiv:2006.05421*.
- Ogedegbe, G.; and Pickering, T. 2010. Principles and techniques of blood pressure measurement. *Cardiology clinics*, 28(4): 571–586.
- Pan, J.; Bi, Q.; Yang, Y.; Zhu, P.; and Bian, C. 2022. Label-efficient hybrid-supervised learning for medical image segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, 2026–2034.
- Paszke, A.; Gross, S.; Massa, F.; et al. 2019. Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32.
- Qin, K.; Huang, W.; and Zhang, T. 2021. Deep generative model with domain adversarial training for predicting arterial blood pressure waveform from photoplethysmogram signal. *Biomedical Signal Processing and Control*, 70: 102972.
- Reisner, A.; Shaltis, P. A.; McCombie, D.; Asada, H. H.; Warner, D. S.; and Warner, M. A. 2008. Utility of the photoplethysmogram in circulatory monitoring. *The Journal of the American Society of Anesthesiologists*, 108(5): 950–958.
- Sarkar, P.; and Etemad, A. 2021. Cardiogan: Attentive generative adversarial network with dual discriminators for synthesis of ecg from ppg. In *AAAI*, volume 35, 488–496.
- Shaltis, P.; Reisner, A.; and Asada, H. 2006a. Calibration of the photoplethysmogram to arterial blood pressure: capabilities and limitations for continuous pressure monitoring. In *2005 IEEE Engineering in Medicine and Biology 27th Annual Conference*, 3970–3973. IEEE.
- Shaltis, P. A.; Reisner, A.; and Asada, H. H. 2006b. Wearable, cuff-less PPG-based blood pressure monitor with novel height sensor. In *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 908–911. IEEE.
- Shimazaki, S.; Bhuiyan, S.; Kawanaka, H.; and Oguri, K. 2018. Features extraction for cuffless blood pressure estimation by autoencoder from photoplethysmography. In *2018 40th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2857–2860. IEEE.
- Van Den Oord, A.; Vinyals, O.; et al. 2017. Neural discrete representation learning. *Advances in neural information processing systems*, 30.
- Vardhan, K. R.; Vedanth, S.; Poojah, G.; Abhishek, K.; Kumar, M. N.; and Vijayaraghavan, V. 2021. BP-Net: Efficient deep learning for continuous arterial blood pressure estimation using photoplethysmogram. In *2021 20th IEEE International Conference on Machine Learning and Applications (ICMLA)*, 1495–1500. IEEE.
- Wu, T. H.; Pang, G. K.-H.; and Kwong, E. W.-Y. 2014. Predicting systolic blood pressure using machine learning. In *7th international conference on information and automation for sustainability*, 1–6. IEEE.
- Xie, S.; Girshick, R.; Dollár, P.; Tu, Z.; and He, K. 2017. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1492–1500.
- Yoon, J.; Jarrett, D.; and Van der Schaar, M. 2019. Time-series generative adversarial networks. *Advances in neural information processing systems*, 32.
- Zeiler, M. D.; Krishnan, D.; Taylor, G. W.; and Fergus, R. 2010. Deconvolutional networks. In *2010 IEEE Computer Society Conference on computer vision and pattern recognition*, 2528–2535. IEEE.
- Zhang, L.; Ji, Y.; Lin, X.; and Liu, C. 2017. Style transfer for anime sketches with enhanced residual u-net and auxiliary classifier gan. In *ACPR*, 506–511. IEEE.
- Zhu, F.; Ye, F.; Fu, Y.; Liu, Q.; and Shen, B. 2019. Electrocardiogram generation with a bidirectional LSTM-CNN generative adversarial network. *Scientific reports*, 9(1): 6734.