

# Lost in Time? A Meta-Learning Framework for Time-Shift-Tolerant Physiological Signal Transformation

Qian Hong<sup>1\*</sup>, Cheng Bian<sup>4\*</sup>, Xiao Zhou<sup>1,2,3†</sup>, Xiaoyu Li<sup>4</sup>, Yelei Li<sup>4</sup>, Zijing Zeng<sup>4</sup>

<sup>1</sup>Gaoling School of Artificial Intelligence, Renmin University of China, Beijing, China

<sup>2</sup>Beijing Key Laboratory of Research on Large Models and Intelligent Governance

<sup>3</sup>Engineering Research Center of Next-Generation Intelligent Search and Recommendation, MOE

<sup>4</sup>OPPO Health Lab, Shenzhen, China

{qianhong99, xiaozhou}@ruc.edu.cn, {biancheng, lixiaoyu5, liyelei1, zijing}@oppo.com

## Abstract

Translating non-invasive signals such as photoplethysmography (PPG) and ballistocardiography (BCG) into clinically meaningful signals like arterial blood pressure (ABP) is vital for continuous, low-cost healthcare monitoring. However, temporal misalignment in multimodal signal transformation impairs transformation accuracy, especially in capturing critical features like ABP peaks. Conventional synchronization methods often rely on strong similarity assumptions or manual tuning, while existing Learning with Noisy Labels (LNL) approaches are ineffective under time-shifted supervision, either discarding excessive data or failing to correct label shifts. To address this challenge, we propose **ShiftSyncNet**, a meta-learning-based bi-level optimization framework that automatically mitigates performance degradation due to time misalignment. It comprises a transformation network (*TransNet*) and a time-shift correction network (*SyncNet*), where *SyncNet* learns time offsets between training pairs and applies Fourier phase shifts to align supervision signals. Experiments on one real-world industrial dataset and two public datasets show that *ShiftSyncNet* outperforms strong baselines by 9.4%, 6.0%, and 12.8%, respectively. The results highlight its effectiveness in correcting time shifts, improving label quality, and enhancing transformation accuracy across diverse misalignment scenarios, pointing toward a unified direction for addressing temporal inconsistencies in multimodal physiological transformation.

**Code** — <https://github.com/HQ-LV/ShiftSyncNet>

## 1 Introduction

Invasive arterial blood pressure (ABP) measurement is widely regarded as the clinical gold standard for blood pressure monitoring. However, its measurement is generally based on invasive procedures (Ogedegbe and Pickering 2010), which carry risks of infection, discomfort, and the need for specialized medical staff, making them impractical for continuous monitoring. These limitations have motivated non-invasive alternatives such as photoplethysmography (PPG), which measures blood flow via light, and bal-

\*These authors contributed equally.

†Corresponding author.

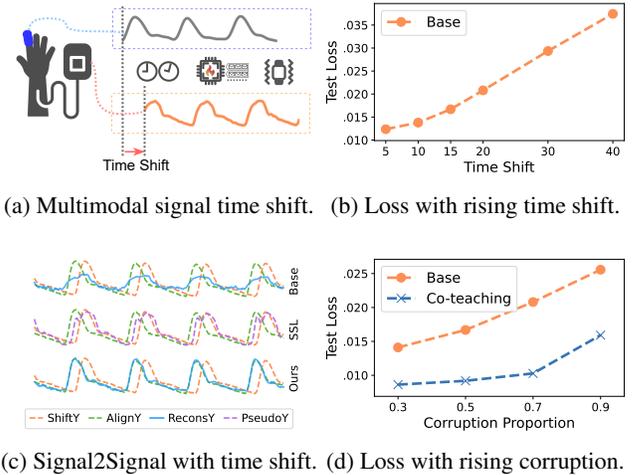


Figure 1: Time shift between multimodal signals degrades physiological signal transformation performance.

listocardiography (BCG), which captures cardiac mechanical activity (Fortino and Giampà 2010; Misawa, Suzuki, and Miura 2022; Lin et al. 2025; Li et al. 2024). Converting these signals into ABP-like waveforms enables continuous, cost-effective, and wearable-friendly blood pressure monitoring.

Beyond population-level health monitoring (Wang et al. 2018), waveform transformation for personalized assessment is increasingly achieved using deep learning models, including CNNs (Ibtehaz et al. 2022; Cao et al. 2023; Chen et al. 2024), GANs (Golany and Radinsky 2019; Sarkar and Etemad 2021), and Transformers (Lan 2023; Yuan et al. 2024). However, temporal misalignment between source and target signals remains a critical unresolved challenge. Fig. 1a illustrates potential causes of time shifts between multimodal physiological signals. For instance, PPG signals from finger-clip sensors and ABP signals from invasive brachial arterial catheters are collected by different devices and subject to factors such as sensor-clock asynchrony, system scheduling delays, device placement variability, and firmware or driver faults, all of which can introduce temporal misalignment.

To illustrate the impact of time shifts on waveform transformation, Fig. 1b, 1c, 1d present PPG-to-ABP results the *Base model*, an InceptionTime (Ismail Fawaz et al. 2020) network trained on corrupted pairs with misaligned physiological signals. As the magnitude of the time shift increases (Fig. 1b) or the proportion of corrupted samples grows (Fig. 1d), the *Base model*'s performance declines markedly. To visualize this effect, Fig. 1c (top row) shows the predicted ABP waveforms, where "ShiftY", "AlignY", and "ReconsY" denote the misaligned labels, aligned labels, and model outputs. The results show degraded transformation performance, especially in capturing ABP peaks which is critical for hypertension diagnosis. Although existing biomedical signal synchronization methods (Xiao, Ding, and Hu 2022; Goodwin et al. 2023; Boljanić et al. 2023; Eleveld et al. 2024) can alleviate such issues, they generally assume similar waveform shapes or require manual tuning, which limits their broader applicability. This motivates our first research question (**RQ1**): *How can we automatically mitigate the performance degradation caused by sensor time misalignment in physiological signal transformations?*

A closely related direction to signal transformation under sensor time misalignment is *Learning with Noisy Labels* (LNL) (Song et al. 2022), where class label noise often stems non-expert annotations or automated labeling on crowdsourcing platforms. Existing LNL methods are typically categorized into **sample selection**, which filters out suspected noisy samples (Han et al. 2018), and **label correction**, which adjusts incorrect labels to improve supervision (Shu et al. 2019). As shown in Fig. 1d, the performance of Co-teaching (Han et al. 2018), a representative sample selection method, degrades sharply as the corruption rate increases. Early label correction methods, such as noise transition matrix estimation (Goldberger and Ben-Reuven 2017), mainly target discrete classification labels and therefore struggle with regression tasks or complex noise like time shifts. To reduce information loss, some studies employ semi-supervised learning (SSL) to generate hard or soft pseudo-labels (Zhang 2017; Menon et al. 2020; Arazo et al. 2019; Li, Socher, and Hoi 2020; Zheltonozhskii et al. 2022; Xiao et al. 2022; Zhang et al. 2024). However, when applied to noisy data with over-parameterized models, SSL often produces unreliable pseudo-labels. As shown in Fig. 1c, SSL-generated soft labels ("PseudoY") may blend peaks different time shifts, leading to misalignment with the true supervision signal. Although recent studies consider label corruption in time series (Ma et al. 2023; Nagaraj et al. 2024), they remain focused on classification and do not address time misalignment in sequence-to-sequence tasks. In waveform transformation, sensor time misalignment introduces label noise similar to LNL but with a key difference: misaligned labels still retain valuable waveform characteristics. Existing LNL methods often overlook this property, highlighting the need for approaches that handle high label corruption while retaining useful signal information. This motivates our second research question (**RQ2**): *How can we leverage time-shifted labels to infer the time offset and obtain a corrected, aligned supervision signal?*

In real-world scenarios, only a limited number of cor-

rectly labeled samples are typically available. This observation motivates the adoption of meta-learning (Hospedales et al. 2022) for label correction (Zheng, Awadallah, and Dumais 2021; Wu et al. 2021), where a separate meta-network is trained on a small, clean auxiliary dataset to correct corrupted supervision signals. Building on this approach, a meta-learning label correction scheme is expected to address these challenges by learning time offsets and fully utilizing the waveform features preserved in time-shifted labels.

To explore **RQ1**, we propose a meta-learning-based bi-level optimization framework, **ShiftSyncNet**, for waveform transformation tasks with tolerance to time-shifts. Specifically, our framework consists of two networks: the waveform transformation network (*TransNet*), which translates source signals into target signals, and the time-shift correction network (*SyncNet*), which generates potentially aligned supervision signals corrupted labels. Both networks are trained simultaneously via bi-level optimization: *TransNet* is updated using pseudo-labels *SyncNet*, while *SyncNet* is updated based on losses computed on a small, aligned meta-dataset. To address **RQ2**, *SyncNet* handles inherent time offsets in training pairs by exploiting the Fourier shift property. This allows frequency-domain phase shifts to align misaligned supervision with the correct targets, increasing effective training data and improving model performance.

The contributions of this paper are as follows:

- We tackle the time-shift challenge in waveform transformation using a meta-learning bi-level optimization framework, where *SyncNet* corrects misaligned labels.
- We propose the application of phase shifts in frequency domain to produce accurately aligned supervision signals, based on the time offsets learned by *SyncNet*.
- Experiments on real-world and public datasets show that our framework outperforms baselines in correcting time shifts, improving data usability, and enhancing accuracy.

## 2 Related Work

### 2.1 Physiological Signal Transformation

Recent work has focused on reconstructing ABP wearable signals using deep learning. LSTMs (Harfiya, Chang, and Li 2021), CNNs (Ibtehaz et al. 2022; Pan et al. 2024), and InceptionTime (Cao et al. 2023; Chen et al. 2024) are widely used for modeling temporal dependencies. GANs (Golany and Radinsky 2019; Sarkar and Etemad 2021) generate realistic waveforms but face stability issues. Transformer models (Lan 2023; Yuan et al. 2024) handle long-range dependencies but may lose temporal information in periodic time series tasks due to permutation-invariant attention and tend to memorize patterns rather than learn underlying periodicity (Zeng et al. 2023; Dong et al. 2024).

### 2.2 Multimodal Physiological Signal Alignment

Various methods have been proposed to correct time offsets in multimodal physiological signals. Dynamic Time Warping (DTW) (Hong, Park, and Baek 2017; Liu et al. 2019; Jiang et al. 2020) aligns time series of different lengths, while cross-correlation methods (Xiao, Ding, and Hu 2022)

align signals by finding maximum correlation. Shared physiological artifacts (e.g., heartbeat, motion, coughing) (Goodwin et al. 2023), can also support synchronization. However, these methods are effective mainly for signals with similar waveforms or inherent alignment (e.g., smartwatch ECG vs. patch ECG) (Goodwin et al. 2023), and perform poorly with large waveform differences or physiological delays (e.g., PPG and ABP). Empirical approaches (Eleveld et al. 2024) incrementally detect offsets but rely on manual tuning and domain knowledge, limiting scalability.

### 2.3 Learning with Noisy Labels

Many LNL methods (Song et al. 2022) begin by identifying and eliminating (Han et al. 2018) or down-weighting (Shu et al. 2019) noisy samples. However, their effectiveness drops as the proportion of corrupted labels increases because fewer samples remain usable. To overcome this, label-correction methods have been developed. Noise-transition matrix estimation (Goldberger and Ben-Reuven 2017) struggles under complex noise types. SSL-based approaches, such as regularization techniques (Zhang 2017; Menon et al. 2020) or dual-model frameworks (Arazo et al. 2019; Li, Socher, and Hoi 2020; Zheltonozhskii et al. 2022; Xiao et al. 2022; Zhang et al. 2024), rely on model predictions for soft or hard relabeling, but these pseudo-labels often become unreliable due to overfitting on noisy data. Meta-learning (Hospedales et al. 2022) instead corrects supervision using a separate meta-network trained on a small clean meta-dataset, and recent work reduces its computational cost with gradient approximation techniques (Zheng, Awadallah, and Dumais 2021; Wu et al. 2021).

Compared to manual signal time synchronization and classical LNL methods, *ShiftSyncNet* directly learns time offsets and automatically generates aligned supervision, improving data usability and enabling time-shift-tolerant signal transformation.

## 3 Problem Statement

We formulate the problem of waveform transformation affected by sensor time misalignment. Given a source signal time series  $x$ , the objective is to learn a waveform transformation network  $f_\theta$  parameterized by  $\theta$ , which maps  $x$  to the target signal series  $y : y = f_\theta(x)$ . In scenarios where there is a unknown temporal shift  $s$  between the source and target signals, two types of datasets are utilized: temporally misaligned training dataset:  $D' = \{(x, y')\}^N$ , and temporally aligned metaset:  $D = \{(x_m, y_m)\}^M$ , where  $M \ll N$ . In the training dataset  $D'$ , the sample pairs  $(x, y')$  exhibit unknown temporal shifts, leading to disrupted physiological correspondences between the source and target signals.

## 4 Methodology

As shown in Fig. 2, we propose a time-shift-tolerant meta-learning framework to address the challenge of physiological waveform transformation under time-shift interference. *TransNet*  $f_\theta$  is optimized as the primary objective to perform the waveform transformation task. To obtain corrected supervision signals, we introduce a meta-network, time-shift

correction network (*SyncNet*)  $h_\alpha$ , whose optimization serves as the meta-objective. *SyncNet* learns the time-shift  $s$  misaligned training samples  $(x, y')$  in  $D'$  and applies phase shifting in frequency domain to generate aligned supervision signals  $y_c$ . These corrected signals provide more accurate supervision for the *TransNet*. Within this meta-learning framework, we jointly optimize both  $f_\theta$  and  $h_\alpha$ .

This section begins by formulating the bi-level optimization problem, followed by detailed introductions to the meta-gradient approximation method, the *SyncNet* architecture, and the *sample-selection-based training strategy*.

### 4.1 bi-level Optimization Problem Formulation

Intuitively, if *SyncNet*  $h_\alpha$  provides high-quality corrected and aligned supervision for  $D'$ , *TransNet*  $f_\theta$  should obtain low loss on the aligned metaset  $D$ . This leads to the following bi-level optimization problem:

$$\min_{\alpha} \mathcal{L}_D(\theta_\alpha^*) \quad \text{s.t.} \quad \theta_\alpha^* = \arg \min_{\theta} \mathcal{L}_{D'}(\alpha, \theta). \quad (1)$$

We specify the upper- and lower-level objectives as:

$$\mathcal{L}_D(\theta_\alpha^*) \triangleq \mathbb{E}_{(x_m, y_m) \in D} \ell(y_m, f_{\theta_\alpha^*}(x_m)), \quad (2)$$

$$\mathcal{L}_{D'}(\alpha, \theta) \triangleq \mathbb{E}_{(x, y') \in D'} \ell(h_\alpha(f_\theta(x), y'), f_\theta(x)), \quad (3)$$

where the upper-level optimization objective  $\mathcal{L}_D(\theta_\alpha^*)$  is to adjust the meta-network parameters  $\alpha$  to minimize the loss of  $f_\theta$  on the metaset  $D$ ; the lower-level optimization objective  $\mathcal{L}_{D'}(\alpha, \theta)$  is to adjust the parameters  $\theta$  to minimize the training loss of  $f_\theta$  on  $D'$ , where the labels in  $D'$  are corrected by  $h_\alpha$ .  $\ell$  represents loss function for waveform transformation, such as Mean Squared Error (MSE) loss.

If obtaining optimal parameters  $\theta^*$  for every meta-parameter  $\alpha$  were required, iterative gradient-based optimization would become computationally prohibitive. Therefore, instead of precisely solving for  $\theta^*$  at every  $\alpha$ , we adopt a widely used alternative: approximating  $\theta^*$  after performing  $k$ -step gradient descent (GD) updates on  $\theta$ .

### 4.2 K-step GD Lookahead Meta-Gradient

**One-step GD Meta-Gradient Approximation.** We start by discussing the one-step approximation for meta-parameter, i.e., when  $k = 1$ . Given meta-parameter  $\alpha$ , the optimal  $\theta^*$  can be approximated by performing one-step GD update:

$$\theta^* = \theta^{t+1} \approx \theta^t - \eta \nabla_{\theta} \mathcal{L}_{D'}(\alpha, \theta^t), \quad (4)$$

where  $\theta^t$  denotes the current parameters,  $\theta^{t+1}$  the updated parameters, and  $\eta$  the learning rate for updating  $\theta$ .

The gradient of  $\alpha$  is calculated via the chain rule:

$$\frac{\partial \mathcal{L}_D(\theta^*)}{\partial \alpha} = \frac{\partial \mathcal{L}_D(\theta^{t+1})}{\partial \theta^{t+1}} \frac{\partial \theta^{t+1}}{\partial \alpha} = -\eta g_{\theta^{t+1}} H_{\theta, \alpha}^t, \quad (5)$$

where  $g_{\theta^{t+1}}$  represents the gradient of the training loss on the metaset  $D$ , and  $\frac{\partial \theta^{t+1}}{\partial \alpha}$  is computed as follows:

$$\frac{\partial \theta^{t+1}}{\partial \alpha} = -\eta \frac{\partial}{\partial \alpha} \nabla_{\theta} \mathcal{L}_{D'}(\alpha, \theta^t) = -\eta H_{\theta, \alpha}^t, \quad (6)$$

where  $H_{\theta, \alpha}^t = \nabla_{\theta, \alpha} \mathcal{L}_{D'}(\alpha, \theta^t)$ .

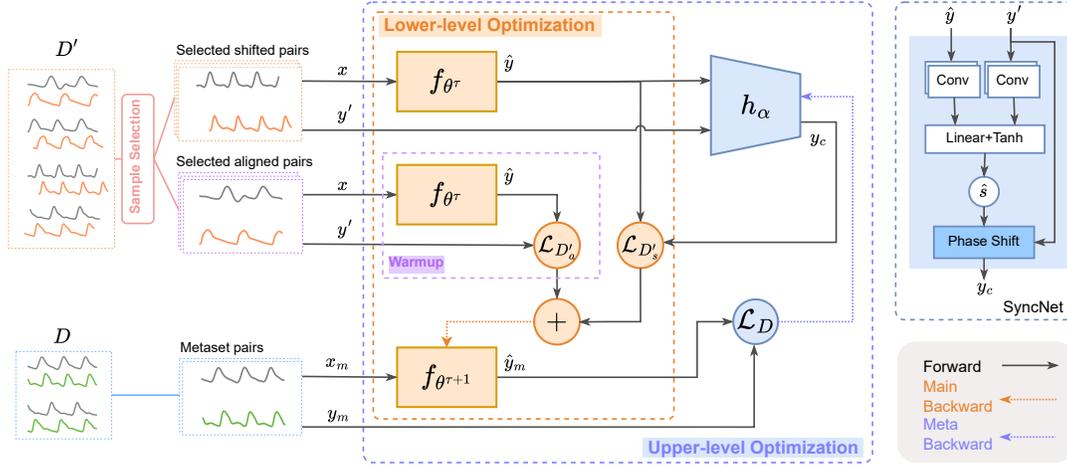


Figure 2: Overview of *ShiftSyncNet*. It follows a bi-level optimization structure: the lower level updates *TransNet*  $f_\theta$  to minimize the training loss on the misaligned dataset  $D'$  using labels corrected by *SyncNet*  $h_\alpha$ , while the upper level updates *SyncNet* to minimize *TransNet*'s loss on the clean metaset  $D$ .

**K-step GD Meta-Gradient Approximation.** For  $k > 1$ , the backbone model updates  $k$  steps for each meta-network update. For  $\tau$  where  $t - k + 1 \leq \tau \leq t$ , the meta-parameters remain unchanged, i.e.,  $\alpha^t = \alpha^{t-1} = \dots = \alpha^{t-k+1}$ . Thus, the parameters  $\theta$  for the previous  $k$  steps depend on  $\alpha$ :

$$\frac{\partial \theta^{\tau+1}}{\partial \alpha} = \frac{\partial}{\partial \alpha} (\theta^\tau - \eta \nabla_{\theta} \mathcal{L}_{D'}(\alpha, \theta^\tau)) \approx (1 - \eta) \frac{\partial \theta^\tau}{\partial \alpha} - \eta H_{\theta, \alpha}^\tau, \quad (7)$$

where we approximate  $H_{\theta, \theta}^\tau = \nabla_{\theta, \theta} \mathcal{L}_{D'}(\alpha, \theta^\tau) \approx I$ . Expanding the recursion gives:

$$\frac{\partial \theta^{\tau+1}}{\partial \alpha} = -\eta H_{\theta, \alpha}^\tau - \eta \sum_{j=1}^{k-1} (1 - \eta)^j H_{\theta, \alpha}^{\tau-j}. \quad (8)$$

Substituting above  $\frac{\partial \theta^{\tau+1}}{\partial \alpha}$  to solve for meta-gradient gives:

$$\frac{\partial \mathcal{L}_D(\theta^{\tau+1})}{\partial \alpha} = g_{\theta^{\tau+1}} \frac{\partial \theta^{\tau+1}}{\partial \alpha} \approx -\eta g_{\theta^{\tau+1}} H_{\theta, \alpha}^\tau + \lambda \frac{\partial \mathcal{L}_D(\theta^\tau)}{\partial \alpha}, \quad (9)$$

where  $\lambda = 1 - \eta$ . The first term in Eq. 9 represents the current meta-gradient, consistent with the case  $k = 1$  in Eq. 5. The second term approximates the cumulative gradients the previous  $k - 1$  steps with a discount factor  $\lambda$ , requiring only the latest meta-gradient to be stored.

The meta-gradient for current step can be computed as:

$$\eta g_{\theta^{\tau+1}} H_{\theta, \alpha} = \nabla_{\alpha} (\eta \nabla_{\theta}^\top \mathcal{L}_{D'}(\alpha, \theta^\tau) \nabla_{\theta} \mathcal{L}_D(\theta^{\tau+1})). \quad (10)$$

For proofs of Eq. 7, 8 and 9, refer to App. A.

### 4.3 Correct Shifted Labels via Time Shifting Property of Fourier Transform

Returning to the task in this paper, our objective is to generate a corrected supervision signal  $y$  based on the output  $\hat{y} = f(x)$  *TransNet* and the label  $y'$ . A naive design for the *SyncNet* involves applying two separate 1D-CNNs to extract features  $\hat{y}$  and  $y'$ , followed by producing a corrected signal

$y_c$ . However, this approach has proven ineffective in learning the correct label, as it underemphasizes the richer supervisory information contained in  $y'$ .

Recalling the temporally misaligned sample pair  $(x, y')$ , we hypothesize that in the later stages of training, if  $f_\theta$  becomes sufficiently accurate, its output  $\hat{y}$  and the label  $y'$  should differ only by the time-offset  $s$ , with their underlying waveforms remaining highly similar. Even in the early training stages, when the waveform characteristics of  $\hat{y}$  and  $y'$  still differ significantly, we can still estimate the time offset  $s$  by exploiting the characteristics of the two periodic signals. Therefore, we modify the *SyncNet*'s task:  $h_\alpha$  is trained to learn the  $s$   $\hat{y}$  and  $y'$ . By trimming a segment of length  $s$  the beginning of  $\hat{y}$  and the end of  $y'$ , or vice versa, the remaining middle parts are the aligned segments, which can be used for loss calculation. However, the rounding operation applied to  $s$  and the slicing operation on  $\hat{y}$  make the loss non-differentiable, making it impossible for *SyncNet* to learn. To ensure that the learned  $s$  for label-correction contributes to the loss calculation and enables gradient back-propagation, we utilize the following time shifting property of Fourier transforms.

**Statement 1.** *The time shifting property of Fourier transform states that shifting a signal  $y'$  by  $t_0$  in time domain introduces a linear phase shift in frequency spectrum with a slope of  $-\omega t_0$ . Therefore, if*

$$y'(t) \xleftrightarrow{FT} Y'(\omega), \quad (11)$$

then based on time-shifting property of Fourier transform,

$$y'(t - t_0) \xleftrightarrow{FT} e^{-j\omega t_0} Y'(\omega). \quad (12)$$

Therefore, after *SyncNet* learns the time-offset  $s$ , we apply a *Fourier transform* to  $y'$  to obtain  $Y' = \text{FFT}(y')$ . In the frequency domain, a linear phase shift is applied as  $Y_c = Y' \cdot e^{-j2\pi f s}$ , followed by an inverse *Fourier transform* to obtain the corrected time-domain signal  $y_c = \text{IFFT}(Y_c)$ . Here,  $s$

is embedded in the exponential term of Fourier transform result, making it differentiable.

#### 4.4 Sample-selection-based Training Strategy

To better exploit the metaset, we initialize the backbone pre-trained on it. As shown in prior work (Zhang et al. 2021; Han et al. 2018), deep networks first learn clean, simple patterns, so early-stage losses can separate aligned corrupted samples. However, as training proceeds, shifted samples increasingly disrupt learning. Thus, we select only likely aligned samples for warmup in the first  $e$  epochs, using the following loss:

$$\mathcal{L}_{D'_a}(\theta) \triangleq \mathbb{E}_{(x,y') \in D'_a} \ell(y', f_\theta(x)), \quad (13)$$

where  $D'_a$  refers to the  $1 - r$  proportion of low-loss samples selected using the Co-teaching (Han et al. 2018) threshold, with the remaining samples discarded.

In later epochs, we also partition each batch into potentially aligned and misaligned samples, computing  $\mathcal{L}_{D'_a}$  with  $y'$  for selected aligned ones and using corrected supervision  $y_c h_\alpha$  for selected misaligned ones:

$$\mathcal{L}_{D'_s}(\alpha, \theta) \triangleq \mathbb{E}_{(x,y') \in D'_s} \ell(h_\alpha(f_\theta(x), y'), f_\theta(x)). \quad (14)$$

The soft loss for updating  $f_\theta$  has two terms weighted by the ratios of selected aligned and misaligned samples:

$$\mathcal{L}'_{\text{soft}} = \beta \mathcal{L}'_{D'_a} + (1 - \beta) \mathcal{L}'_{D'_s}, \quad \beta = \frac{|D'_a|}{|D'_a| + |D'_s|}. \quad (15)$$

## 5 Experiments

### 5.1 Datasets

We evaluate our approach on three datasets listed in Table 1. The first is real-world industrial data OPPO Health Lab<sup>1</sup>, a provider of smart healthcare solutions. To assess generalizability, we also use two public datasets: VitalDB (Lee et al. 2022) and MIMIC II (Saeed et al. 2011).

Datasets	VitalDB	MIMIC II	OML
#Subjects	144	942	182
#Segments	517,200	364,774	11,529
SBP(mmHg)	120.38 ± 19.63	134.19 ± 22.93	115.09 ± 12.24
DBP(mmHg)	66.14 ± 11.45	60.21 ± 12.60	74.78 ± 6.80

Table 1: Dataset statistical description.

**OML.** Our industrial dataset OML (Bian et al. 2024), collected 180+ subjects via specialized wearables, includes 125 Hz BCG and PPG signals. We apply Butterworth bandpass filtering (BCG) and derivative-based mean filtering (PPG), segment signals into 6.14 s slices (768 points) and split subject into train/valid/test sets in an 8:1:1 ratio.

**VitalDB.** VitalDB (Lee et al. 2022) contains 6,388 PPG, ECG, and ABP records ICU patients at Seoul National University Hospital. We adopt the cleaned version in (Wang et al. 2023), downsampled to 125 Hz and split by subject into training and test sets. Signals are segmented into 6.14 s slices, with 10% of training data for validation.

<sup>1</sup><https://www.oppo.com/en/>

**MIMIC II.** A subset of MIMIC-II (Saeed et al. 2011) PhysioNet (Goldberger et al. 2000), collected at Beth Israel Deaconess Medical Center, using the preprocessed version by Kachuee et al. (Kachuee et al. 2015, 2016), includes 12,000 PPG, ECG, and ABP records 942 ICU patients at 125 Hz. Signals are sliced into 6.14 s windows and split into train/valid/test sets (8:1:1).

### 5.2 Experimental Settings

**Time-shift Settings.** As real time-shifts are often unavailable, we simulate it by injecting artificial shifts into the training data, enabling controlled evaluation without ground-truth alignment and improving robustness by exposing the model to more diverse, complex misalignments. To simulate complex time-shift scenarios, we vary: (1) shift magnitude: each corrupted sample is randomly shifted to left or right by  $s$  points ( $s \leq S$ ), with  $S \in \{5, 10, 15, 20, 30, 40\}$ ; (2) corruption proportion: a fraction  $r$  of samples is shifted, with  $r \in \{0.3, 0.5, 0.7, 0.9\}$ . These settings enable evaluation of method’s robustness under varying time-shift conditions.

**Evaluation Metrics.** We use Mean Squared Error (MSE), Percent Root Difference (PRD), and Mean Absolute Error (MAE) for waveform transformation evaluation.

**Baselines.** We use three baseline categories: basic backbone models (Swin-Transformer (Liu et al. 2021), ResNet (He et al. 2016), InceptionTime (Ismail Fawaz et al. 2020)), sample selection methods (MW-Net (Shu et al. 2019), Co-teaching (Han et al. 2018)), and label correction methods (U-correction (Arazo et al. 2019), DivideMix (Li, Socher, and Hoi 2020), MSLC (Wu et al. 2021), MLC (Zheng, Awadallah, and Dumais 2021) and C2MT (Zhang et al. 2024)).

### 5.3 Performance Evaluation

**Overall Evaluation.** Table 2 shows overall performance with  $S = 20$  and  $r = 0.7$ , reporting average and standard deviation over 5 trials. More results on time-shift tolerance under various settings are in the App. B. Among backbone models, InceptionTime and ResNet outperform SwinTransformer, with InceptionTime chosen for subsequent backbone due to its higher computational efficiency. Overall, our model achieves the best results on 7 of 9 metrics across three datasets and ranks second on the other two metrics.

Compared to label correction methods, *ShiftSyncNet* outperforms U-correction, DivideMix, MSLC, MLC, and C2MT, reducing MSE by 30.4%, 19.6%, and 32.2% over the second-best method across the three datasets. U-correction, DivideMix, and C2MT rely on semi-supervised strategies that depend heavily on over-parameterized backbone predictions, which are unreliable under noise. As shown in Fig. 3, these methods generate overconfident pseudo-labels that deviate true waveforms. MSLC combines historical predictions blended with shifted labels, which may distort soft signal waveform. MLC, in contrast, employs an independent meta-network to directly output corrected labels and performs better than most, but as seen in Fig. 3b, its pseudo-labels only approximate the waveform rather than provide entirely accurate supervision labels. This limitation becomes more pronounced under larger shifts (Fig. 3c).

Methods	VitalDB			MIMIC II			OML		
	MSE ↓	PRD ↓	MAE ↓	MSE ↓	PRD ↓	MAE ↓	MSE ↓	PRD ↓	MAE ↓
SwinTransformer	0.065±0.001	5.492±0.027	0.205±0.002	0.070±0.001	5.700±0.026	0.214±0.001	0.055±0.001	5.522±0.026	0.187±0.001
ResNet	0.021±0.000	3.153±0.018	0.109±0.001	0.031±0.000	3.798±0.010	0.136±0.000	0.035±0.001	4.437±0.046	0.144±0.001
InceptionTime	0.021±0.000	3.100±0.016	0.109±0.001	0.031±0.000	3.771±0.009	0.135±0.000	0.036±0.001	4.466±0.033	0.144±0.001
MW-Net	0.019±0.001	2.972±0.058	0.104±0.002	0.028±0.000	3.591±0.021	0.127±0.001	0.035±0.000	4.393±0.025	0.140±0.001
Co-teaching	0.010±0.000	2.167±0.021	<b>0.065</b> ±0.001	0.019±0.000	2.945±0.015	<b>0.082</b> ±0.000	0.025±0.000	3.760±0.033	0.109±0.001
U-correction	0.019±0.000	2.985±0.000	0.103±0.000	0.028±0.000	3.623±0.000	0.128±0.000	0.039±0.000	4.622±0.000	0.156±0.000
DivideMix	0.017±0.001	2.779±0.097	0.089±0.007	0.022±0.001	3.167±0.095	0.100±0.004	0.035±0.000	4.379±0.012	0.142±0.001
MSLC	0.022±0.001	3.170±0.085	0.112±0.004	0.031±0.000	3.806±0.016	0.137±0.001	0.036±0.001	4.450±0.075	0.144±0.004
MLC	0.014±0.000	2.516±0.014	0.083±0.000	0.021±0.000	3.146±0.028	0.100±0.002	0.034±0.001	4.338±0.056	0.135±0.002
C2MT	0.018±0.001	2.860±0.123	0.089±0.004	0.020±0.000	3.071±0.014	0.094±0.002	0.034±0.001	4.341±0.054	0.135±0.002
Ours	<b>0.009</b> ±0.000	<b>2.097</b> ±0.015	0.066±0.001	<b>0.016</b> ±0.000	<b>2.755</b> ±0.009	0.083±0.000	<b>0.023</b> ±0.001	<b>3.572</b> ±0.038	<b>0.108</b> ±0.001

Table 2: Performance comparison of our method with baselines. **Bold** indicates the best results. ↓ means lower is better.

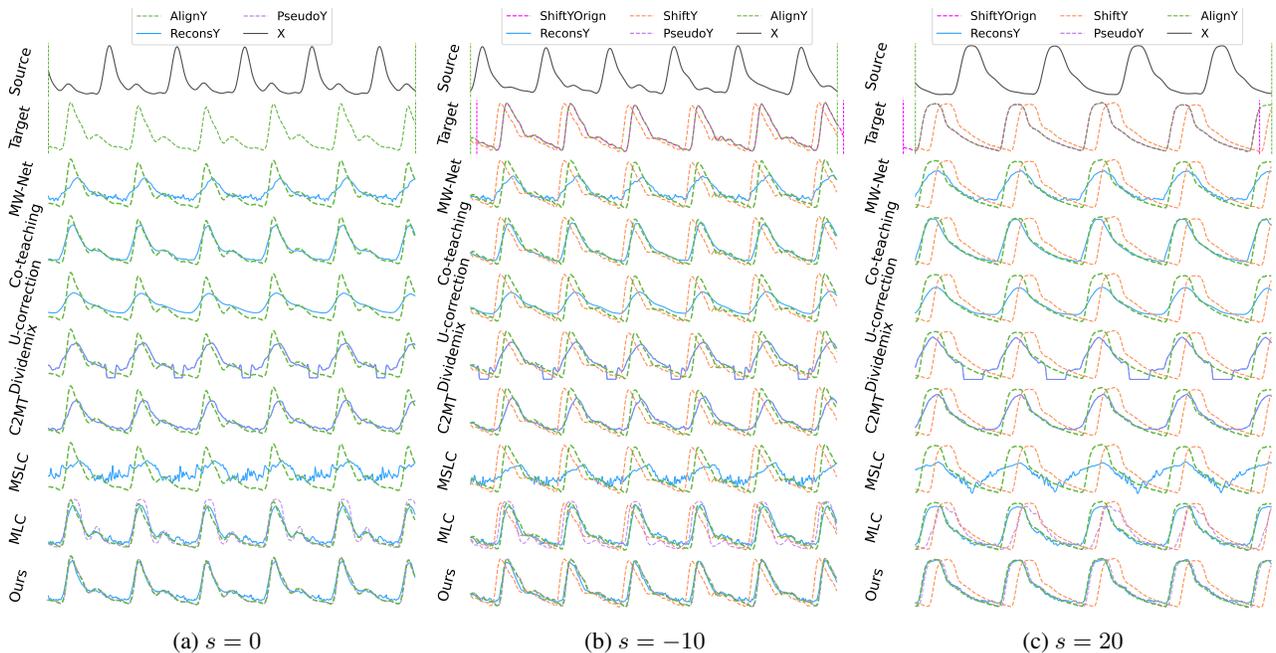


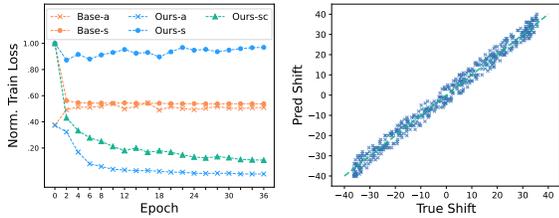
Figure 3: Signal transformation visualization under varying time-shift conditions. “Shift” indicates supervision with time-shift  $s$ ; “ReconsY” and “PseudoY” represent predictive signals and pseudo-labels, respectively. For clarity, we also show the true original position of “ShiftY” as “ShiftYOrigin” and the input-aligned target as “AlignY”, both unknown during training.

Compared to sample selection methods, *ShiftSyncNet* outperforms Co-teaching and MW-Net, reducing MSE by 6.0%, 12.8%, and 9.4% over second-best Co-teaching across the three datasets. MW-Net reweights all samples adaptively, which may mislead training and degrade performance. Co-teaching selects low-loss samples based on prior assumptions and performs better than MW-Net. However, as discussed in App. B, it suffers from selection bias and limited usable samples, hindering its ability to capture fine-grained waveform features such as signal peaks (Fig. 3a, 3b).

Unlike these baselines, our method redefines the meta-network objective to predict time shift  $s$  and reverse la-

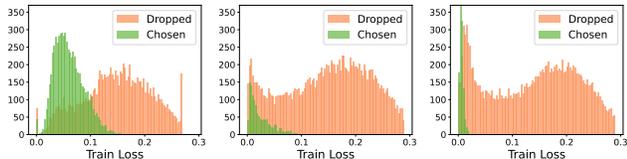
bel by phase shift, producing truly-aligned and waveform-consistent supervision that enhances data usability.

**Effectiveness of ShiftSyncNet.** We evaluate the effectiveness of *ShiftSyncNet* by analyzing its handling of aligned and misaligned samples. Fig. 4a compares training loss trajectories of *Base model* and *ShiftSyncNet*, both using a pre-trained backbone initialized on a small aligned metaset for fair comparison. Without correction, Base model notably suffers time shifts, with even aligned sample loss (Base-a) increasing over time. In contrast, *ShiftSyncNet* consistently reduces losses for aligned (Ours-a) and corrupted samples after correcting shifted labels (Ours-s → Ours-sc), indicat-

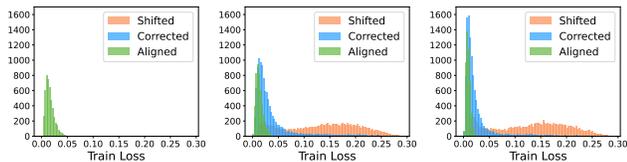


(a) Training loss history. (b) Time-shift prediction.

Figure 4: Effectiveness of *ShiftSyncNet*.



(a) Co-teaching, ep 5 (b) Co-teaching, ep 30 (c) Co-teaching, ep 60



(d) Ours, ep 5 (e) Ours, ep 30 (f) Ours, ep 60

Figure 5: Training loss distribution of aligned and shifted samples for Co-teaching and ours at epochs 5, 30, and 60.

ing effective label correction and improved supervision.

To verify *SyncNet*'s time-shift estimates, Fig. 4b plots true time-shifts  $s$  against predicted shifts  $\hat{s}$ , showing strong diagonal alignment  $\hat{s} = s$ . This enables phase-aligned pseudo-labels and improves waveform transformation.

To explore the performance gap, we compare how Co-teaching and our method handle aligned and shifted samples. By selecting low-loss samples and discarding high-loss ones based on a growing forgetting rate, Co-teaching gradually reduces useful data over time, hindering optimization. In contrast, our method recovers these discarded samples by correcting the supervision of shifted pairs. As shown in Fig. 5, both methods identify aligned samples (green), but Co-teaching gradually drops high-loss ones (orange in Fig. 5a, 5b, 5c). Our method corrects their labels (orange  $\rightarrow$  blue in Fig. 5d, 5e, 5f), significantly reducing loss.

#### 5.4 Downstream Task Evaluation

To demonstrate clinical applicability, we conduct downstream systolic and diastolic blood pressure (SBP/DBP) prediction under time-shift interference on VitalDB and MIMIC-II, as shown in Table 3. *ShiftSyncNet* achieves the best performance by mitigating time-shift effects. Compared to InceptionTime trained without time misalignment correction, it reduces SBP/DBP MAE by 80%/72% on VitalDB

and 71%/67% on MIMIC-II, respectively. It also meets the Association for the Advancement of Medical Instrumentation (AAMI) standard (MAE < 5 mmHg), confirming its utility for clinical prediction and continuous monitoring.

Methods	VitalDB		MIMIC II	
	SBP	DBP	SBP	DBP
InceptionTime	12.41	5.50	16.82	7.13
MW-Net	11.82	5.51	15.97	6.91
Co-teaching	3.22	2.44	5.82	2.99
U-correction	12.05	4.92	16.08	7.01
DivideMix	6.49	3.37	7.45	4.93
MSLC	13.83	1.93	19.50	7.48
MLC	5.70	3.60	7.49	4.91
C2MT	4.00	2.83	6.95	2.82
Ours	<b>2.43</b>	<b>1.49</b>	<b>4.83</b>	<b>2.36</b>
Impr.	80%	72%	71%	67%

Table 3: Downstream SBP/DBP prediction MAE (mmHg). "Impr." denotes improvement of ours over InceptionTime.

#### 5.5 Ablation Studies

To evaluate *sample-selection-based training strategy*, we compare our method with two variants in Table 4. Variant I (w/o SL) uses  $\mathcal{L}_{D'}$  in Eq. 3 instead of  $\mathcal{L}_{D'_{soft}}$  in Eq. 15, relying solely on pseudo-labels and ignoring aligned supervision. Variant II (w/o WU) removes warmup phase. Our method consistently outperforms both variants across time-shift settings. Variant I misses the benefits of supervision aligned samples, while our soft loss balances aligned and corrected shifted signals for improved robustness. Variant II performs worse due to the absence of warmup, which stabilizes training by selecting low-loss samples. Combining soft loss and warmup yields superior performance.

Dataset	Variation	$r = 0.5$	$r = 0.7$	$r = 0.9$
VitalDB	w/o SL	0.0094	0.0096	0.0098
	w/o WU	0.0096	0.0096	0.0096
	Ours	<b>0.0093</b>	<b>0.0095</b>	<b>0.0095</b>
OML	w/o SL	0.0237	0.0227	0.0238
	w/o WU	0.0242	0.0235	0.0245
	Ours	<b>0.0231</b>	<b>0.0226</b>	<b>0.0228</b>

Table 4: Test MSE comparison of ablation models.

## 6 Conclusion

We tackle waveform transformation under time-shift interference in physiological signals with *ShiftSyncNet*, a meta-learning framework comprising *TransNet* for transformation and *SyncNet* for time-shift correction via Fourier phase shifts. A sample-selection strategy further enhances robustness against misaligned labels. Experiments on real-world and public datasets show that *ShiftSyncNet* achieves superior accuracy and robustness under time shifts, supporting its application in continuous health monitoring.

## 7 Acknowledgements

This work was supported by the Public Computing Cloud at Renmin University of China and the Fund for Building World-Class Universities (Disciplines) at Renmin University of China.

## References

- Arazo, E.; Ortego, D.; Albert, P.; O'Connor, N.; and McGuinness, K. 2019. Unsupervised label noise modeling and loss correction. In *International conference on machine learning*, 312–321. PMLR.
- Bian, C.; Li, X.; Bi, Q.; Zhu, G.; Lyu, J.; Zhang, W.; Li, Y.; and Zeng, Z. 2024. Constraint latent space matters: an anti-anomalous waveform transformation solution from photoplethysmography to arterial blood pressure. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 11087–11095.
- Boljanić, T.; Malešević, J.; Vujnović, S.; and Janković, M. M. 2023. Comparison of Time Domain Methods for Alignment of RR Signals Acquired by Different Sensor Systems. In *2023 10th International Conference on Electrical, Electronic and Computing Engineering (IcETRAN)*, 1–4. IEEE.
- Cao, T.; Tran, N.; Nguyen, L.; Nguyen, H.; and Pham, H. 2023. IncepSE: Leveraging InceptionTime's performance with Squeeze and Excitation mechanism in ECG analysis. In *Proceedings of the 12th International Symposium on Information and Communication Technology*, 578–584.
- Chen, Y.; Ke, M.; Sun, Y.; and Wang, L. 2024. An Improved InceptionTime Model for Mental Workload Assessment Based on EEG Signal. In *2024 5th International Seminar on Artificial Intelligence, Networking and Information Technology (AINIT)*, 1838–1842. IEEE.
- Dong, Y.; Li, G.; Tao, Y.; Jiang, X.; Zhang, K.; Li, J.; Deng, J.; Su, J.; Zhang, J.; and Xu, J. 2024. Fan: Fourier analysis networks. *arXiv preprint arXiv:2410.02675*.
- Eleveld, N.; Harmsen, M.; Elting, J. W. J.; and Maurits, N. M. 2024. Haemosync: A synchronisation algorithm for multimodal haemodynamic signals. *Computer Methods and Programs in Biomedicine*, 108298.
- Fortino, G.; and Giampà, V. 2010. PPG-based methods for non invasive and continuous blood pressure measurement: An overview and development issues in body sensor networks. In *2010 IEEE International Workshop on Medical Measurements and Applications*, 10–13. IEEE.
- Golany, T.; and Radinsky, K. 2019. Pgans: Personalized generative adversarial networks for ecg synthesis to improve patient-specific deep ecg classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, 557–564.
- Goldberger, A. L.; Amaral, L. A.; Glass, L.; Hausdorff, J. M.; Ivanov, P. C.; Mark, R. G.; Mietus, J. E.; Moody, G. B.; Peng, C.-K.; and Stanley, H. E. 2000. PhysioBank, PhysioToolkit, and PhysioNet: components of a new research resource for complex physiologic signals. *circulation*, 101(23): e215–e220.
- Goldberger, J.; and Ben-Reuven, E. 2017. Training deep neural networks using a noise adaptation layer. In *International conference on learning representations*.
- Goodwin, A. J.; Dixon, W.; Mazwi, M.; Hahn, C. D.; Meir, T.; Goodfellow, S. D.; Kazazian, V.; Greer, R. W.; McEwan, A.; Laussen, P. C.; et al. 2023. The truth Hertz—synchronization of electroencephalogram signals with physiological waveforms recorded in an intensive care unit. *Physiological Measurement*, 44(8): 085002.
- Han, B.; Yao, Q.; Yu, X.; Niu, G.; Xu, M.; Hu, W.; Tsang, I.; and Sugiyama, M. 2018. Co-teaching: Robust training of deep neural networks with extremely noisy labels. *Advances in neural information processing systems*, 31.
- 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.
- He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Hong, J. Y.; Park, S. H.; and Baek, J.-G. 2017. Solving the singularity problem of semiconductor process signal using improved dynamic time warping. In *2017 IEEE 11th International Conference on Semantic Computing (ICSC)*, 266–267. IEEE.
- Hospedales, T.; Antoniou, A.; Micaelli, P.; and Storkey, A. 2022. Meta-Learning in Neural Networks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(9): 5149–5169.
- Ibtehaz, 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.
- Jiang, Y.; Qi, Y.; Wang, W. K.; Bent, B.; Avram, R.; Olgin, J.; and Dunn, J. 2020. EventDTW: An improved dynamic time warping algorithm for aligning biomedical signals of nonuniform sampling frequencies. *Sensors*, 20(9): 2700.
- Kachuee, M.; Kiani, M. M.; Mohammadzade, H.; and Shabany, M. 2015. Cuff-less high-accuracy calibration-free blood pressure estimation using pulse transit time. In *2015 IEEE international symposium on circuits and systems (ISCAS)*, 1006–1009. IEEE.
- Kachuee, M.; Kiani, M. M.; Mohammadzade, H.; and Shabany, M. 2016. Cuffless blood pressure estimation algorithms for continuous health-care monitoring. *IEEE Transactions on Biomedical Engineering*, 64(4): 859–869.
- Lan, E. 2023. Performer: A novel ppg-to-ecg reconstruction transformer for a digital biomarker of cardiovascular disease detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1991–1999.
- 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.
- Li, J.; Socher, R.; and Hoi, S. C. 2020. Dividemix: Learning with noisy labels as semi-supervised learning. *arXiv preprint arXiv:2002.07394*.
- Li, X.; Hussein, R.; Zhu, G.; Sui, X.; Li, H.; Yang, X.; Zeng, Z.; and Li, Y. 2024. Continuous Blood Pressure Monitoring and Hypertension Risk Screening Using Smart Watch. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 1–6. IEEE.
- Lin, H.; Li, J.; Hussein, R.; Sui, X.; Li, X.; Zhu, G.; Katsaggelos, A. K.; Zeng, Z.; and Li, Y. 2025. Longitudinal Wrist PPG Analysis for Reliable Hypertension Risk Screening Using Deep Learning. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–5. IEEE.
- Liu, Y.; Yao, Y.; Wang, Z.; Plested, J.; and Gedeon, T. 2019. Generalized alignment for multimodal physiological signal learning. In

- 2019 *International Joint Conference on Neural Networks (IJCNN)*, 1–10. IEEE.
- Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.
- Ma, P.; Liu, Z.; Zheng, J.; Wang, L.; and Ma, Q. 2023. CTW: Confident Time-Warping for Time-Series Label-Noise Learning. In *IJCAI*, 4046–4054.
- Menon, A. K.; Rawat, A. S.; Reddi, S. J.; and Kumar, S. 2020. Can gradient clipping mitigate label noise? In *International Conference on Learning Representations*.
- Misawa, A.; Suzuki, A.; and Miura, H. 2022. Relationship analysis between BCG features and blood pressure. In *2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS)*, 1–2. IEEE.
- Nagaraj, S.; Gerych, W.; Tonekaboni, S.; Goldenberg, A.; Ustun, B.; and Hartvigsen, T. 2024. Learning from Time Series under Temporal Label Noise. *arXiv preprint arXiv:2402.04398*.
- Ogedegbe, G.; and Pickering, T. 2010. Principles and techniques of blood pressure measurement. *Cardiology clinics*, 28(4): 571–586.
- Pan, J.; Liang, L.; Liang, Y.; Tang, Q.; Chen, Z.; and Zhu, J. 2024. Robust modelling of arterial blood pressure reconstruction from photoplethysmography. *Scientific Reports*, 14(1): 1–13.
- Saeed, M.; Villarroel, M.; Reisner, A. T.; Clifford, G.; Lehman, L.-W.; Moody, G.; Heldt, T.; Kyaw, T. H.; Moody, B.; and Mark, R. G. 2011. Multiparameter Intelligent Monitoring in Intensive Care II: a public-access intensive care unit database. *Critical care medicine*, 39(5): 952–960.
- Sarkar, P.; and Etemad, A. 2021. Cardiogan: Attentive generative adversarial network with dual discriminators for synthesis of ecg from ppg. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 488–496.
- Shu, J.; Xie, Q.; Yi, L.; Zhao, Q.; Zhou, S.; Xu, Z.; and Meng, D. 2019. Meta-weight-net: Learning an explicit mapping for sample weighting. *Advances in neural information processing systems*, 32.
- Song, H.; Kim, M.; Park, D.; Shin, Y.; and Lee, J.-G. 2022. Learning from noisy labels with deep neural networks: A survey. *IEEE transactions on neural networks and learning systems*, 34(11): 8135–8153.
- Wang, W.; Mohseni, P.; Kilgore, K. L.; and Najafizadeh, L. 2023. PulseDB: A large, cleaned dataset based on MIMIC-III and VitalDB for benchmarking cuff-less blood pressure estimation methods. *Frontiers in Digital Health*, 4: 1090854.
- Wang, Y.; Zhou, X.; Noulas, A.; Mascolo, C.; Xie, X.; and Chen, E. 2018. Predicting the Spatio-Temporal Evolution of Chronic Diseases in Population with Human Mobility Data. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, 3578–3584. International Joint Conferences on Artificial Intelligence Organization.
- Wu, Y.; Shu, J.; Xie, Q.; Zhao, Q.; and Meng, D. 2021. Learning to purify noisy labels via meta soft label corrector. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 35, 10388–10396.
- Xiao, R.; Ding, C.; and Hu, X. 2022. Time Synchronization of Multimodal Physiological Signals through Alignment of Common Signal Types and Its Technical Considerations in Digital Health. *Journal of Imaging*, 8(5): 120.
- Xiao, R.; Dong, Y.; Wang, H.; Feng, L.; Wu, R.; Chen, G.; and Zhao, J. 2022. Promix: Combating label noise via maximizing clean sample utility. *arXiv preprint arXiv:2207.10276*.
- Yuan, X.; Wang, W.; Li, X.; Zhang, Y.; Hu, X.; and Deen, M. J. 2024. CATransformer: A Cycle-Aware Transformer for High-Fidelity ECG Generation From PPG. *IEEE Journal of Biomedical and Health Informatics*.
- Zeng, A.; Chen, M.; Zhang, L.; and Xu, Q. 2023. Are transformers effective for time series forecasting? In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, 11121–11128.
- Zhang, C.; Bengio, S.; Hardt, M.; Recht, B.; and Vinyals, O. 2021. Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, 64(3): 107–115.
- Zhang, H. 2017. mixup: Beyond empirical risk minimization. *arXiv preprint arXiv:1710.09412*.
- Zhang, Q.; Zhu, Y.; Yang, M.; Jin, G.; Zhu, Y.; and Chen, Q. 2024. Cross-to-merge training with class balance strategy for learning with noisy labels. *Expert Systems with Applications*, 249: 123846.
- Zheltonozhskii, E.; Baskin, C.; Mendelson, A.; Bronstein, A. M.; and Litany, O. 2022. Contrast to divide: Self-supervised pre-training for learning with noisy labels. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1657–1667.
- Zheng, G.; Awadallah, A. H.; and Dumais, S. 2021. Meta label correction for noisy label learning. In *Proceedings of the AAAI conference on artificial intelligence*, volume 35, 11053–11061.

## A Proof

**Proof of Eq. 7.** To prove Eq. 7, we use the implicit dependence of  $\theta$  on  $\alpha$ , where  $\theta$ , updated  $k$  times per update of  $\alpha$ , can be expressed as a function of  $\alpha$ :

$$\frac{\partial\theta^{\tau+1}}{\partial\alpha} = \frac{\partial}{\partial\alpha}(\theta^\tau - \eta\nabla_{\theta}\mathcal{L}_{D'}(\alpha, \theta^\tau(\alpha))) \quad (16)$$

$$= \frac{\partial\theta^\tau}{\partial\alpha} - \eta\frac{\partial}{\partial\alpha}\nabla_{\theta}\mathcal{L}_{D'}(\alpha, \theta^\tau(\alpha)) \quad (17)$$

$$= (I - \eta H_{\theta, \theta}^\tau)\frac{\partial\theta^\tau}{\partial\alpha} - \eta H_{\theta, \alpha}^\tau \quad (18)$$

$$\approx (1 - \eta)\frac{\partial\theta^\tau}{\partial\alpha} - \eta H_{\theta, \alpha}^\tau. \quad (19)$$

**Proof of Eq. 8.** Eq. 19 establishes a recursive relation for  $\frac{\partial\theta^{\tau+1}}{\partial\alpha}$ , expressing it in terms of  $\frac{\partial\theta^\tau}{\partial\alpha}$  and higher-order gradients. Opening this recurrence, we derive:

$$\frac{\partial\theta^{\tau+1}}{\partial\alpha} = (1 - \eta)^k \frac{\partial\theta^{\tau-k+1}}{\partial\alpha} - \eta \sum_{j=0}^{k-1} (1 - \eta)^j H_{\theta, \alpha}^{\tau-j} \quad (20)$$

$$\approx -\eta \sum_{j=0}^{k-1} (1 - \eta)^j H_{\theta, \alpha}^{\tau-j} \quad (21)$$

$$= -\eta H_{\theta, \alpha}^\tau - \eta \sum_{j=1}^{k-1} (1 - \eta)^j H_{\theta, \alpha}^{\tau-j}. \quad (22)$$

**Proof of Eq. 9.** In the following proof,  $\frac{\partial\theta^{\tau+1}}{\partial\alpha}$  is substituted into the meta-gradient computation. The past  $k-1$  gradients  $g_{\theta^{\tau-j+1}}$  ( $1 \leq j \leq k-1$ ) are approximated by the current gradient  $g_{\theta^{\tau+1}}$ , avoiding storage of historical gradients.

$$\frac{\partial\mathcal{L}_D(\theta^{\tau+1})}{\partial\alpha} = g_{\theta^{\tau+1}} \frac{\partial\theta^{\tau+1}}{\partial\alpha} \quad (23)$$

$$= -\eta g_{\theta^{\tau+1}} H_{\theta, \alpha}^\tau + \eta \sum_{j=1}^{k-1} \gamma^{\tau-j} (-g_{\theta^{\tau-j+1}} H_{\theta, \alpha}^{\tau-j}) \quad (24)$$

$$\approx -\eta g_{\theta^{\tau+1}} H_{\theta, \alpha}^\tau + (1 - \eta) \frac{\partial\mathcal{L}_D(\theta^\tau)}{\partial\alpha}, \quad (25)$$

where  $\gamma^{\tau-j} = g_{\theta^{\tau+1}} \frac{g_{\theta^{\tau-j+1}}^\top}{\|g_{\theta^{\tau-j+1}}\|^2}$ .

## B Supplementary Experimental Material

### Baseline Details

Here’s a detailed introduction to the baselines:

Basic Backbone Models:

- The basic baselines are models for waveform transformation that does not handle corrupted data, including Swin-Transformer (Liu et al. 2021), InceptionTime (Islam Fawaz et al. 2020), and ResNet (He et al. 2016).

Sample Selection Methods:

- MW-Net (Shu et al. 2019): A meta-learning approach that adaptively learns loss weights to emphasize clean samples consistent with meta-knowledge.

- Co-teaching (Han et al. 2018): Trains two networks that mutually select likely clean samples in each mini-batch.

Label Correction Methods:

- U-correction (Arazo et al. 2019): Estimates label corruption probabilities, uses bootstrapping loss and adapts Mixup augmentation.
- DivideMix (Li, Socher, and Hoi 2020): Uses Gaussian Mixture Models to classify clean and noisy samples and uses two networks iteratively for label co-refinement and label co-guessing on labeled and unlabeled samples.
- MSLC (Wu et al. 2021): Learns a weight network based on meta-learning to generate soft labels by combining noisy labels with the backbone’s historical predictions.
- MLC (Zheng, Awadallah, and Dumais 2021): Directly generates corrected labels for noisy samples using a meta-learning-based label correction network.
- C2MT (Zhang et al. 2024): Based on DivideMix’s cross-training, C2MT periodically merges the two networks into a single model via federated parameter averaging.

Both sample selection and label correction methods are model-agnostic, using InceptionTime as the backbone for consistency in our experiments, though others can be used.

### Parameter Sensitivity

We evaluate the sensitivity of  $k$  on VitalDB and OML in Fig 6. Results show that  $k$  performs well when under 10. However, as Fig. 6a shows, for  $r = 0.7$ , smaller  $k$  yield better approximations, whereas with a higher corruption rate of  $r = 0.9$ , longer steps improve performance, supporting the effectiveness of k-step meta-gradient approximation.

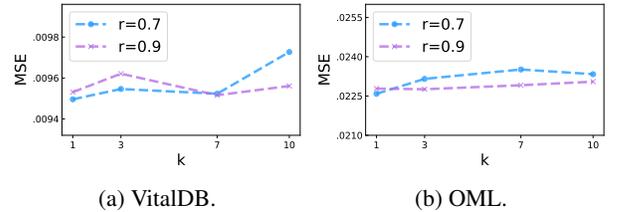


Figure 6: Parameter analysis on  $k$ .

We further analyzed the effect of metaset size  $M$ , a small practical set of manually aligned pairs (1-5% of the training data). As shown in Table 5, performance on the VitalDB dataset plateaus beyond  $M = 500$ , indicating *SyncNet*’s sufficient correction.

$M$	100	200	500	1000	2000
<b>MSE</b>	0.01064	0.01007	0.00945	0.00944	0.00944

Table 5: Parameter analysis on  $M$ .

## Time-shift Tolerance Analysis

As shown in Table 6, 7, and 8, we comprehensively evaluate the robustness of our method under varying corruption ratios ( $r$ ), time-shift magnitudes ( $S$ ), and training sample sizes ( $N_r$ ) on the VitalDB and OML datasets, comparing it against state-of-the-art baselines the sample selection (Co-teaching) and label correction (MLC) groups. Across all settings, our method consistently achieves the lowest MSE mean and standard deviation, demonstrating superior robustness and stability. Specifically, as  $r$  rises 0.3 to 0.9,  $S$  increases 5 to 40, or  $N_r$  drops 0.9 to 0.3, our method maintains optimal performance, whereas Co-teaching and MLC exhibit noticeable performance degradation due to their dependence on clean sample subsets or limited label correction capability. Interestingly, under fixed  $r = 0.7$  and  $N_r = 1.0$ , Co-teaching occasionally benefits larger  $S$ , possibly because larger shifts make it easier to identify and eliminate misaligned samples. However, its performance drops sharply when  $r$  or  $N_r$  reaches 0.9 or 0.3, owing to insufficient clean supervision. In contrast, our method leverages the full dataset via effective time-shift correction by *SyncNet*, consistently ensuring robust performance even under severe misalignment and data scarcity.

Dataset	$r$	Co-teaching	MLC	Ours
VitalDB	0.3	0.0086	0.0121	<b>0.0095</b>
	0.5	0.0092	0.0123	<b>0.0093</b>
	0.7	0.0103	0.0133	<b>0.0095</b>
	0.9	0.0159	0.0140	<b>0.0095</b>
	Avg	0.0110	0.0129	<b>0.0095</b>
	Std	0.0033	0.0009	<b>0.0001</b>
OML	0.3	0.0244	0.0309	<b>0.0241</b>
	0.5	0.0243	0.0314	<b>0.0221</b>
	0.7	0.0252	0.0324	<b>0.0226</b>
	0.9	0.0297	0.0337	<b>0.0228</b>
	Avg	0.0259	0.0321	<b>0.0229</b>
	Std	0.0026	0.0012	<b>0.0009</b>

Table 6: Test MSE under rising  $r$  ( $S = 20, N_r = 1.0$ ).

## Efficiency Analysis

Our method adds moderate overhead during training but incurs no additional cost during inference. Specifically, training involves two extra operations: (1) *SyncNet* forward on  $D'$ , and (2) *TransNet* updates on  $D$  combined with *SyncNet*'s meta-updates every  $k$  steps. These factors raise training time by 2-3 $\times$ . In terms of memory, storing one extra batch of  $D$  and the *SyncNet* parameters incurs limited overhead, and the space-efficient implementation of Eq. 9 requires storing only the latest meta-gradient. At inference, only *TransNet* is used, adding no extra time or memory cost.

## Implementation

Both our method and the baselines for handling corrupted samples utilize InceptionTime as the backbone network. In *SyncNet*, we stack convolutional layers with a kernel size of

Dataset	$S$	Co-teaching	MLC	Ours
VitalDB	5	0.0120	0.0117	<b>0.0096</b>
	10	0.0114	0.0122	<b>0.0093</b>
	15	0.0109	0.0126	<b>0.0095</b>
	20	0.0103	0.0133	<b>0.0095</b>
	30	0.0101	0.0144	<b>0.0097</b>
	40	<b>0.0095</b>	0.0141	0.0098
OML	Avg	0.0107	0.0131	<b>0.0096</b>
	Std	0.0009	0.0011	<b>0.0002</b>
	5	0.0255	0.0306	<b>0.0219</b>
	10	0.0271	0.0310	<b>0.0222</b>
	15	0.0261	0.0315	<b>0.0225</b>
	20	0.0252	0.0324	<b>0.0226</b>
OML	30	0.0241	0.0345	<b>0.0232</b>
	40	0.0250	0.0334	<b>0.0247</b>
	Avg	0.0255	0.0322	<b>0.0228</b>
	Std	0.0010	0.0015	<b>0.0010</b>

Table 7: Test MSE under rising  $S$  ( $r = 0.7, N_r = 1.0$ ).

Dataset	$N_r$	Co-teaching	MLC	Ours
VitalDB	0.9	0.0104	0.0117	<b>0.0095</b>
	0.7	0.0109	0.0124	<b>0.0095</b>
	0.5	0.0122	0.0115	<b>0.0097</b>
	0.3	0.0141	0.0121	<b>0.0101</b>
	Avg	0.0119	0.0119	<b>0.0097</b>
	Std	0.0016	0.0004	<b>0.0003</b>
OML	0.9	0.0262	0.0327	<b>0.0233</b>
	0.7	0.0262	0.0339	<b>0.0241</b>
	0.5	0.0276	0.0350	<b>0.0238</b>
	0.3	0.0287	0.0356	<b>0.0247</b>
	Avg	0.0272	0.0343	<b>0.0240</b>
	Std	0.0012	0.0013	<b>0.0006</b>

Table 8: Test MSE under decreasing  $N_r$  ( $S = 20, r = 0.7$ ).

15 and LeakyReLU activation across the two convolutional blocks, followed by a three-layer MLP for time-shift estimation. We use a batch size of 128, with initial learning rates of  $1.5e-3$  for *TransNet* and  $5e-5$  for *SyncNet*. The Adam optimizer is used for parameter updates, with a weight decay of  $5e-4$  for *TransNet*. A cosine annealing schedule is applied to adjust the learning rate. We use the first 10 epochs to warmup *TransNet*. For models that require the metaset, we randomly select 500, 500, and 300 signal segments the VitalDB, MIMIC, and OML datasets, respectively. Experiments are conducted on an NVIDIA A40 GPU with Ubuntu 20.04.6 LTS, using Python 3.10.0, PyTorch 1.13.0.

## Future Extensions

Our meta-learning bi-level optimization framework provides a basis for applications beyond time-series misalignment. Its modular, model-agnostic design supports extensions to other forms of label corruption, such as artifacts, missing points, or multimodal misalignment in diverse domains like audio-visual synchronization and biomedical signal processing.