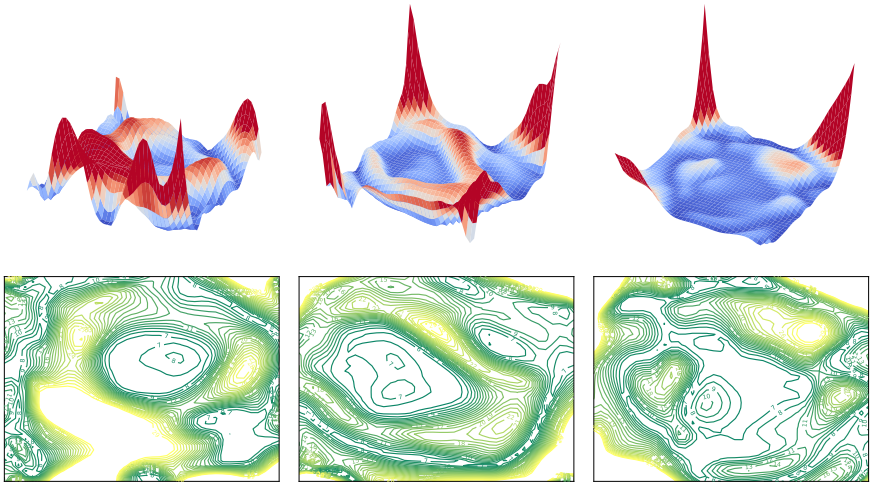
Bi-TTA: Bidirectional Test-Time Adapter for Remote Physiological Measurement

ECCV 2024 Under-review

https://bi-tta.github.io



(a) Pre-trained (b) Priors-based TTA (c) Priors-based Bi-TTA

Fig. 1: Visualization of the mean absolute error (MAE) result on down-sampled VIPL dataset[1] using **(a) the pre-trained rPPG model; (b) priors-based test-time adapted model; and (c) priors-based bidirectionally test-time adapted model.** These visualizations demonstrate that our proposed priors and the Bi-TTA framework significantly enhance the generalization performance (reflected by the flatness of MAE field [26]) beyond the pre-trained model.[2]

Abstract. Remote photoplethysmography (rPPG) is gaining prominence for its non-invasive approach to monitoring physiological signals using only cameras. Despite its promise, the adaptability of rPPG models to new, unseen domains is hindered due to the environmental sensitivity of physiological signals. To address this issue, we pioneer the Test-Time Adaptation (TTA) in rPPG, enabling the adaptation of pre-trained models to the target domain during inference, sidestepping the need for annotations or source data due to privacy considerations. Particularly, utilizing only the user's face video stream as the accessible target domain data, the rPPG model is adjusted by tuning on each single instance it encounters. However, 1) TTA algorithms are designed predominantly for classification tasks, ill-suited in regression tasks such as rPPG due to inadequate supervision. 2) Tuning pre-trained models in a single-instance manner

introduces variability and instability, posing challenges to effectively filtering domain-relevant from domain-irrelevant features while simultaneously preserving the learned information. To overcome these challenges, we present **Bi-TTA**, a novel expert knowledge-based **Bi**directional **T**est-**T**ime **A**dapter framework. Specifically, leveraging two expert-knowledge priors for providing self-supervision, our Bi-TTA primarily comprises two modules: a prospective adaptation (PA) module using sharpness-aware minimization to eliminate domain-irrelevant noise, enhancing the stability and efficacy during the adaptation process, and a retrospective stabilization (RS) module to dynamically reinforce crucial learned model parameters, averting performance degradation caused by overfitting or catastrophic forgetting. To this end, we established a large-scale benchmark for rPPG tasks under TTA protocol, promoting advancements in both the rPPG and TTA fields. The experimental results demonstrate the significant superiority of our approach over the state-of-the-art (SoTA).

Keywords: Bidirectional Test-Time Adaptation · Expert Knowledge-based Priors · Remote Photoplethysmography

1 Introduction

The cardiac cycle refers to the process experienced by the cardiovascular system from the start of one heartbeat to the beginning of the next. Through the analysis of the cardiac cycle, important physiological indicators such as heart rate (HR), respiration rate (RF), and heart rate variability (HRV) can be derived. These information are instrumental in assessing an individual's health and emotional condition, playing a crucial role in medical diagnostics, health surveillance, and forensic analysis [20, 36, 55].

Traditional physiological measurement devices like electrocardiograms (ECG), heart rate bands, and finger-clips, despite their accuracy, are often expensive and uncomfortable to wear. In contrast, rPPG offers a non-invasive and more convenient alternative by extracting blood volume pulses (BVP) from facial videos, analyzing the skin's light absorption variations to measure HR, HRV, and RF [44, 69, 70]. These monitors are especially important for tracking health status and sympathetic activity levels [20, 36, 55]. Nowadays, physiological measurements based on rPPG, using only ordinary cameras, are increasingly becoming the focus of research in computer vision community [5, 9, 33, 35, 43, 47, 50]. However, the periodic color changes in the skin captured by remote cameras are quite subtle, and variables like natural scene illumination and head movements can significantly impair the reliability of physiological indicator measurements.

[1] This dataset is 1/20 equally spaced down-sampled from the full VIPL dataset [42], and contains 8349 samples in total.

[2] This visualization is generated following [26].

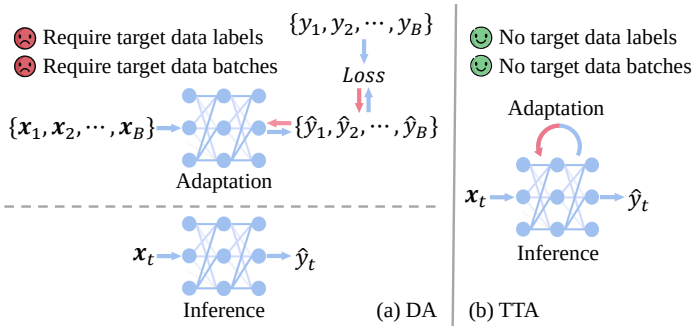


Fig. 2: Visualization of Domain Adaptation (DA) and Test-Time Adaptation (TTA) methodologies. DA utilizes batch learning with labeled target data, while TTA dynamically refines the model during inference without relying on target data labels or distribution. Both DA and TTA obviate the requirement for source data. Note that domain generalization (DG) is excluded as it does not focus on *specific* target domain adaptation.

Conventional rPPG methods [6, 25, 37, 46, 60, 64] perform well only in constrained environments. Recent deep learning (DL) methods [2, 43, 49, 53, 68] have demonstrate promising results, however, the most of training data are collected in controlled laboratory environments with limited variations in illumination, camera parameters, etc. Therefore, these methods still struggle to generalize to unseen domains, necessitating model adaptation.

For privacy considerations, the adaptation of rPPG models in unseen scenarios is required to be performed without accessing the source data, and target data annotations or distributions (*i.e.*, large batches of target data samples). Addressing this, we firstly implement Test-Time Adaptation (TTA) in the field of rPPG. TTA aims to fine-tune a pre-trained source model during inference time, as illustrated in Fig. 2, without accessing the distribution and labeling of both source data and target data and naturally eliminate the need for intensive re-training [61]. This innovation allows for the automatic adaptation of pre-trained rPPG models to unseen domains during inference, without the necessity for labelling or batches of target data.

However, applying TTA directly into rPPG faces two key limitations. 1) Predominantly designed for classification tasks, most TTA algorithms leverage entropy in normalization layers or pseudo labels. These strategies are not inherently suited for regression tasks like rPPG, leading to a lack of appropriate supervision. 2) In real-world scenarios, the pre-trained rPPG model is tuned during inference the user's face video stream, on an instance-by-instance context.This variability and instability introduced by biased instances pose challenges to effectively filtering out domain-irrelevant noise and simultaneously preserving crucial learned information as the model acquiring new knowledge. This nuanced balance is critical in preventing slow or ineffective adaptation under single-instance learning condition, ensuring an efficient and consistent adaptation process.

To navigate these challenges, we proposed a novel **Bi**directional **T**est-**T**ime **A**dapter (**Bi-TTA**) framework, grounded on two expert knowledge-based self-supervised priors. Firstly, two regularizations are proposed for offering effective supervision especially in rPPG tasks, addressing the aforementioned ill-suited problem of applying existing TTA algorithms into regression tasks. These two priors focus on temporal and spatial consistency respectively, which is derived from the characteristics of physiological signals. Then, two strategies are introduced for stable and effective fine-tuning process. The prospective adaptation (PA) strategy is designed for filtering out the irrelevant information of each encountered instance, aiming that only target-domain-essential feature is truly contributing to model adaptation. The retrospective stabilization (RS) strategy backtracks and consolidates essential model parameters dynamically, leveraging previous update gradients during the self-supervised tuning process, ensuring that the learned adaptation ability is preserved intact. We evaluate our method across various datasets, benchmarking against existing prevailing TTA algorithms, demonstrating the superiority of our Bi-TTA. Overall, our contributions can be summarized as follows:

- We introduce the TTA protocol to rPPG tasks, paving the way for a new paradigm in self-supervised adapting of rPPG models during inference.
- We introduce two novel self-supervised priors to regularize temporal and spatial consistency, providing appropriate supervision.
- We propose a novel Bidirectional Test-Time Adapter (Bi-TTA) framework grounded on the introduced priors, which promotes effective and efficient adaptation, with enhanced stability.
- We established a large-scale benchmark for rPPG tasks under TTA protocol. Extensive experiments demonstrate significantly improved adaptation capacity of our approach.

2 Related Works

2.1 Remote Physiological Measurement

Remote physiological measurement aims to quantify HR and HRV from BVP signals by interpreting skin chrominance fluctuations. Traditional techniques typically involve color space transformations or signal decomposition to obtain the BVP signal with high signal-to-noise ratio (SNR) [6, 7, 46, 65]. With powerful modeling capacity of deep learning (DL), there's been a significant shift towards employing DL models for remote physiological estimation, yielding impressive results. [4, 16, 41, 43, 52, 66, 69]. Many of these methods extract BVP signals from face videos typically by aggregating spatial and temporal information [44, 52, 69, 70]. Besides, many works utilize hand-designed 2D feature maps extracted from face videos to alleviate computational demands [16, 33, 43, 48, 51]. Recently, researchers have turned to exploring how to leverage task-specific knowledge and how to improve the generalization performance of the model into unseen scenarios.

Task-specific priors. rPPG is distinguished by its unique biological and physical mechanisms, which can be exploited to enhance the physiological signal extraction ability of the model [17, 31, 58, 62]. These task-specific priors can be summed up as 1) spatial consistency: BVP signals extracted from different face areas should exhibit similar; 2) temporal consistency: BVP signals typically exhibit smooth and gradual changes over short time period; 3) heart rate range: The heart rate typically ranges between 40 and 250 beats per minute (bpm). Different from existing works, we provide a more granular solution leveraging the multi-scale latent features, rather than purely focusing on the prediction output of BVP signals.

The generalization of the model. With the domain shifts (*e.g.*, lighting, head motion, environment, etc.), most methods tend to have significant performance degradation. To improve the generalization performance on unseen domains, NEST-rPPG established a large-scale DG evaluation protocol [34]. These DG methods, requiring the access of all source data, do not focus on the adaptation of a specific target domain, leading to sub-optimal performance [56]. Besides, some work utilizes the labels and distribution of the target domain data to improve the model's adaptation performance [10,24,30], which is impractical due to accessing the unseen domain in advance. Differently from the above rPPG methods, we address this domain gap by using the TTA protocol, facilitating the adaptation to specific target domains during inference, based solely on unlabeled instances from the target domain.

2.2 Test-Time Adaptation

Test-time Adaptation (TTA), an emerging protocol, aims to adjust a pre-trained source model during inference by learning from the unlabeled target data, without accessing the distributions and labels of both source and target data, thereby naturally eliminate the need of intensive re-training [27, 61, 63]. Current TTA methods typically require aggregating a large number of samples from the target data to form a batch, and perform adaptation through 1) normalization calibration [1, 21, 22, 39, 74], 2) pseudo-labeling [8, 14], 3) input adaptation [15, 73], 4) dynamic inference that learns the network weights from a set of model parameters [72], 5) meta-learning [12], etc. These approaches can be considered as different forms of source-free domain adaptation (SFDA) [28] and are infeasible for regression tasks like rPPG where data arrives continuously and in a sequential manner. Some instance-level TTA approaches may either rely on specialized network architectures [1, 22, 61, 74] or incorporate additional auxiliary tasks applied during both pre-training and adaptation [11, 57]. By imposing specific non-general requirement to the pre-training stage, these methods narrow application scope of TTA. Moreover, some works even introduce a memory mechanism that turns instances into batches [21], lacking an in-time adaptation to each inference with increased computational cost. In contrast with above mentioned TTA methods, our Bi-TTA adapts the model during inference under each encountered instance from target domain without specific pre-training

or network architecture requirements, offering a flexible, effective, and efficient solution for rPPG model adaptation.

3 Method

3.1 Overview

In this paper, we propose a novel expert knowledge-based Bi-directional Test-Time Adapter (Bi-TTA) to address the performance degradation in unseen target domains. In the following sections, we will firstly formulate the problem in Sec. 3.2. Secondly, Sec. 3.3 introduces the proposed self-supervised priors based on expert knowledge in the realm of rPPG. The bidirectional (prospective and retrospective) adaptation strategy of our Bi-TTA will be specified in Sec. 3.4.

3.2 Preliminaries

The source dataset $D_{\mathrm{S}} \triangleq \cup_{i=1}^{n_{\mathrm{S}}} \{(\boldsymbol{x}_i, y_i)\}$ is assumed to be drawn independently and identically distributed (i.i.d.) from distribution X_{S}. similarly, the target dataset D_{T} is assumed to be sampled from a different distribution X_{T}, reflecting the domain shift. With a model parameterized by $\boldsymbol{w} \in \mathcal{W}$ and considering the supervised loss function $L : \mathcal{W} \times \mathcal{X} \times \mathcal{Y} \to \mathbb{R}_+$ for evaluation, we define the observation loss over dataset D as an approximation of the population loss over distribution X:

$$\mathcal{L}_D(\boldsymbol{w}) = \frac{1}{n} \sum_{i=1}^{n} L\left(\boldsymbol{w}, \boldsymbol{x}_i, y_i\right) \approx \mathbb{E}_{(\boldsymbol{x},y)\sim X}\left[L(\boldsymbol{w}, \boldsymbol{x}, y)\right]. \tag{1}$$

The objective of TTA, including our Bi-TTA, is to minimize the target observation loss $\mathcal{L}_{D_{\mathrm{T}}}(\boldsymbol{w}_t)$ after fine-tuning the initial model weights from $\boldsymbol{w}_0$ to $\boldsymbol{w}_t$ after t observed samples.

3.3 Expert Knowledge-based Priors

In the context of Bi-TTA, the unavailability of target data labels compels the use of self-supervised loss functions for providing the tuning gradient. Inspired by [17, 31, 58, 62], our Bi-TTA capitalizes on the characteristics of rPPG signals to craft these self-supervised loss functions, concentrating on the principles of spatial and temporal consistency. The introduced loss functions are termed as $L_{\mathrm{s}}\left(\boldsymbol{w}_{t-1}, \boldsymbol{x}_t\right)$ and $L_{\mathrm{t}}\left(\boldsymbol{w}_{t-1}, \boldsymbol{x}_t\right)$, focusing on spatial and temporal consistency respectively. These functions play a foundational role in the self-supervised tuning process, providing appropriate supervision based on the expert knowledge, despite the lack of direct label guidance.

Spatial-temporal feature map (STMap). In line with [16, 33, 43, 48, 51], 2D STMap is firstly constructed to encapsulate the face video, as specified in Fig. 3a. This STMap integrates both temporal and spatial dimensions: the temporal

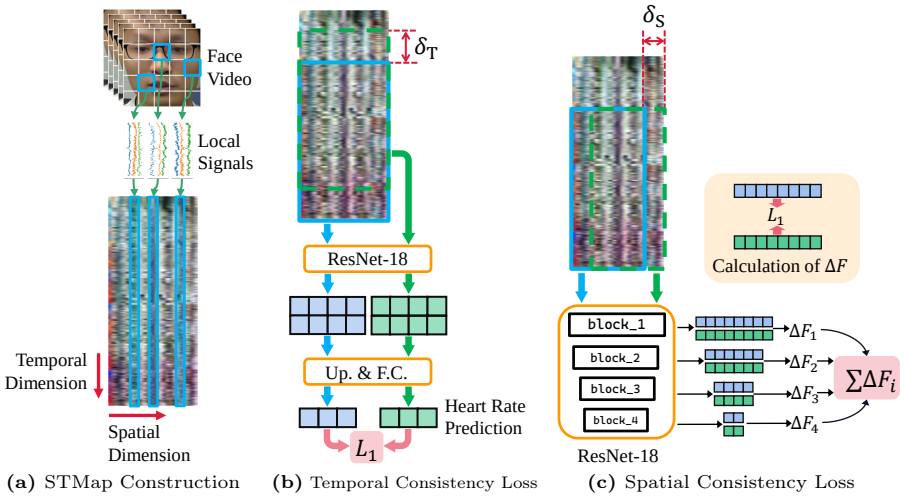


Fig. 3: Illustration of STMap construction and the implementation of our proposed expert knowledge-based priors. (a) The process of generating STMap, encompassing face alignment[3] and cropping, local signal extraction, and the subsequent integration. (b) The calculation process of TCL, aimed at minimizing significant prediction discrepancies between original and temporally shifted HR predictions. (c) The calculation process the SCL, focused on penalizing pronounced disparities across different facial regions. Note that the boxes colored in pink represent the loss outcomes.

dimension represents the time sequence and the spatial dimension corresponds to the patches of the cropped face video frames. The current sample $\boldsymbol{x}_t \in \mathbb{R}^{W \times H}$ is referred to as a sliding window on the STMap, where W represents the sequential length of the sliding window, and H indicates the size of spatial dimension in the STMap.

Temporal consistency loss (TCL). BVP signals exhibit smooth and gradual fluctuations over short time spans. This fundamental attribute, known as the temporal consistency, is essential for accurate physiological signal analysis and the key to guarantee the predictability of physiological signal changes. To preserving this critical characteristic and derive a tuning gradient accordingly, the TCL is meticulously designed to enforce the model's prediction deviations of BVP signals when subjected to minor time perturbations under an acceptable tolerance. To compute the TCL, a random minor temporal perturbation δ_T is introduced by shifting the original sample $\boldsymbol{x}_t$ along the STMap's temporal dimension. The shifted sample $\boldsymbol{x}_{t-\delta_T}$, in parallel with the origin sample, are then fed into the rPPG network $\boldsymbol{w}_{t-1}$ for predicting two sets of BVP signals. Then, for penalizing the model from large discrepancies in the HR estimation results between the original and shifted samples, L_1 regularization is applied on the original and shifted HR prediction results, incorporating with a tolerance

constant ξ_{T}. The specific calculation process of TCL is formulated in Eq. 2, as illustrated in Fig. 3b.

$$\begin{aligned} \Delta^{\mathrm{HR}}_{t,t-\delta_{\mathrm{T}}} &= \left\| \boldsymbol{w}^{\mathrm{HR}}_{t-1}(\boldsymbol{x}_t) - \boldsymbol{w}^{\mathrm{HR}}_{t-1}(\boldsymbol{x}_{t-\delta_{\mathrm{T}}}) \right\|_1 \in \mathbb{R}^{W\times 1} \\ L_{\mathrm{t}}\left(\boldsymbol{w}_{t-1}, \boldsymbol{x}_t\right) &= \sum_{i}^{W} \max\left(0, \Delta^{\mathrm{HR}(i)}_{t,t-\delta_{\mathrm{T}}} - \xi_{\mathrm{T}}\right). \end{aligned} \tag{2}$$

Spatial consistency loss (SCL). Heart rate variations induce chromatic changes across the face, which are periodic and consistent across different facial regions. This phenomenon is referred to as spatial consistency, alongside temporal consistency, are crucial attributes in the rPPG field. Inspired by this phenomenon, SCL is crafted to promote the spatial consistency, obtaining correct adaptation gradient by encouraging the BVP signals from different facial areas to exhibit similarities. In particular, the SCL is calculated covering a series of latent feature maps, corresponding to various depths, (*i.e.*, the number of residual blocks) in the adopted ResNet-18 [18] architecture. Given the current sample $\boldsymbol{x}_t \in \mathbb{R}^{W\times H}$, four latent feature maps $\left\{F_i \in \mathbb{R}^{W_i\times H_i} \mid i \in [1,4]\right\}$ could be extracted from the rPPG network $\boldsymbol{w}_{t-1}$. As illustrated in Fig. 3c, the calculation process of SCL is specify as:

$$L_{\mathrm{s}} = \sum_{i}^{n} \Delta F_i = \sum_{i}^{n} \sum_{j}^{W_i-\delta_{\mathrm{S}}} ||F_{i,j} - F_{i,j+\delta_{\mathrm{S}}}||_1, \tag{3}$$

where $F_{i,j} = F_i[:,j] \in \mathbb{R}^{W_i}$, $n = 4$ represents the number of residual blocks, and δ_S is the incremental spatial shift. Leveraging the fine-grained multi-level strategy, SCL effectively nurtures a comprehensive regularization of spatial features, ensuring that appropriate supervision is available for the adaptation during inference without annotations.

After combining L_{t} and L_{s}, without direct target label reliance, the self-supervised loss is formulated by weighting and summing the spatial and temporal consistency priors as follows:

$$L_{\mathrm{p}} = \lambda_{\mathrm{s}} L_{\mathrm{s}} + \lambda_{\mathrm{t}} L_{\mathrm{t}}. \tag{4}$$

3.4 Bidirectional Test-Time Adaptation

As introduced above, based on the proposed priors, the Bi-TTA framework aims to ensure effective and robust model adaptation under the variability and instability of single-instance learning conditions, minimizing the impact of domain-irrelevant noise. Simultaneously adapting the model with new target samples, our Bi-TTA also focuses on preserving the acquired adaptation ability dynamically, preventing the performance degradation caused by overfitting or catastrophic forgetting, achieving a stable maintenance of the adaptation performance in the target domain.

[3] https://github.com/1adrianb/face-alignment.

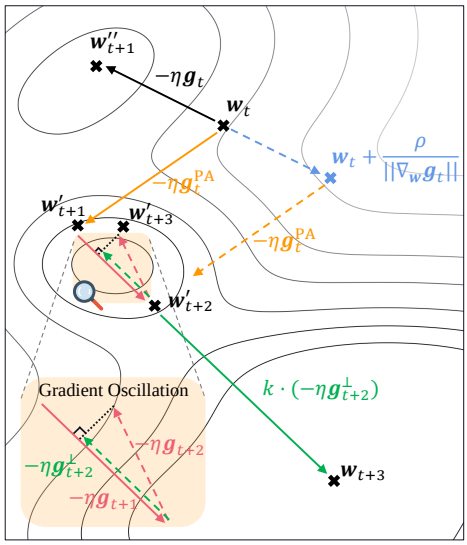


Fig. 4: Illustration of the proposed Bidirectional Test-Time Adapter (Bi-TTA). Black arrows → indicate the adaptation process purely with the proposed two priors, *i.e.*, TCL and SCL. Orange arrows → denote that the PA module adjusts model parameter using the gradient of representative neighborhood with a radius ρ. The green ones → show that the RS is activated when there is an oscillation, which is a sign of performance degradation, for maintaining the essential learned adaptation ability with former tuning gradients. The gradient and learning rate are formulated as $\boldsymbol{g}$ and η respectively.

Prospective adaptation (PA). During the process of model adaptation, especially under single-instance learning conditions, the noisy irrelevant information in biased target samples can easily interfere the model's tuning process, leading to a declined performance. Therefore, inspired by [13,38], rather than seeking out model parameter $\boldsymbol{w}_t$ that simply have low adaptation loss $L_{\mathrm{p}}(\boldsymbol{w}_t)$, we seek out model parameter whose entire neighborhoods to have uniformly low adaptation loss.

$$L'_{\mathrm{p}}(\boldsymbol{w}) = \max_{\|\boldsymbol{\epsilon}\|_2 \leqslant \rho} L_{\mathrm{p}}(\boldsymbol{w} + \boldsymbol{\epsilon}), \tag{5}$$

where the ρ is the disturbance range. This strategy, known as sharpness-aware minimization, has been proved to be effective in various areas [3, 19, 45, 67]. For computing efficiency and differentiability, we approximate the $\boldsymbol{\epsilon}^* = \arg\max_{\|\boldsymbol{\epsilon}\|_2 \leqslant \rho} L_{\mathrm{p}}(\boldsymbol{w}+\boldsymbol{\epsilon})$ leveraging the first-order Taylor Series: $\hat{\boldsymbol{\epsilon}} = \frac{\rho \,\mathrm{sign}(\nabla_{\boldsymbol{w}} L_{\mathrm{p}}(\boldsymbol{w}))}{\|\nabla_{\boldsymbol{w}} L_{\mathrm{p}}(\boldsymbol{w})\|}$. Consequently, we can derive the gradient approximation under this loss function (Eq. 5):

$$\begin{aligned} \boldsymbol{g}_t^{\mathrm{PA}} &= \nabla_{\boldsymbol{w}} L'_{\mathrm{p}}(\boldsymbol{w}) \approx \nabla_{\boldsymbol{w}} L_{\mathcal{S}}(\boldsymbol{w} + \hat{\boldsymbol{\epsilon}}) \\ &= \frac{d(\boldsymbol{w} + \hat{\boldsymbol{\epsilon}})}{d\boldsymbol{w}} \nabla_{\boldsymbol{w}} L_{\mathcal{S}}(\boldsymbol{w}) \Big|_{\boldsymbol{w}+\hat{\boldsymbol{\epsilon}}} \approx \nabla_w L_{\mathcal{S}}(\boldsymbol{w})|_{\boldsymbol{w}+\hat{\boldsymbol{\epsilon}}} \end{aligned} \tag{6}$$

As illustrated in Fig. 4, the gradient $\boldsymbol{g}_t^{\mathrm{PA}}$ is adopted to replace the original gradient merely derived from two self-supervised priors. This strategy boosts the model's robustness to the undesired variability from instance-level conditions, safeguarding the tuning process against the influence from target-domain-irrelevant information. Consequently, it secures both the efficacy and stability of the adaptation process.

Retrospective stabilization (RS). The PA module, by considering the model's entire neighborhood as the optimization objective, effectively enhances the model's robustness against irrelevant disturbances in target samples. However, after the model has processed a certain number of samples, it may still capture the detrimental noise feature instead of latent, universally domain-essential patterns, leading to performance degradation, as evidenced in Fig. 6. Therefore, to ensure reliable physiological signal predictions throughout the entire adaptation process, we propose the retrospective stabilization (RS) module. RS initially introduces the concept of *trend gradient*, termed as $\boldsymbol{g}^*$. Starting with $\boldsymbol{g}_0^* = \mathbf{0}$, the trend gradient after processing t samples is updated via: $\boldsymbol{g}_t^* = \frac{\boldsymbol{g}_{t-1}^* + (1/||L_{\mathrm{p}}||_1)\boldsymbol{g}_t^{\mathrm{PA}}}{1+1/||L_{\mathrm{p}}||_1}$. For each $\boldsymbol{g}_t^{\mathrm{PA}}$, its orthogonal projection on $\boldsymbol{g}_t^*$ is computed, formulated as $\boldsymbol{g}_t^{\perp} = \frac{\boldsymbol{g}_t^{\mathrm{PA}} \cdot \boldsymbol{g}_t^*}{\|\boldsymbol{g}_t^*\|^2}\boldsymbol{g}_t^*$. As illustrated in Fig. 4, RS is activated if this projection aligns in the opposite of the trend gradient, which signifies an gradient oscillation, indicative of a potential fall of the generalization performance. Specifically, upon detection of oscillation, the model's optimization gradient will be the product of $\boldsymbol{g}_t^{\perp}$ and the backtracking degree k. The orthogonal part of $\boldsymbol{g}_t$ on $\boldsymbol{g}_t^*$ is ignored.

$$\boldsymbol{g}_t^{\mathrm{RS}} = \begin{cases} k \cdot \boldsymbol{g}_t^{\perp}, \ \cos(\boldsymbol{g}_t^*, \boldsymbol{g}_t^{\mathrm{PA}}) < 0 \\ (1 - \lambda_t^{\mathrm{RS}})\boldsymbol{g}_t^{\mathrm{PA}} + \lambda_t^{\mathrm{RS}}\boldsymbol{g}_t^*, \ \cos(\boldsymbol{g}_t^*, \boldsymbol{g}_t^{\mathrm{PA}}) \geqslant 0 \end{cases} \tag{7}$$

where $\lambda_t^{\mathrm{RS}} = -1 + 2 \cdot \mathrm{Sigmoid}(-\frac{4t}{\Omega}) \in \mathbb{R}_+^{(0,1)}$, is the anneal scaling factor, Ω denotes the amount of target samples needed to ensure the fidelity of trend gradient, *i.e.*, λ_t^{RS} can be approximated as 1.0 when $t \geqslant \Omega$. The obtained $\boldsymbol{g}_t^{\mathrm{RS}}$ represents the gradient that ultimately contributes to fine-tune the model for target domain adaptation.

4 Experiments

4.1 Datasets

We elaborately select five rPPG datasets, encompassing diverse races, levels of head motion, camera conditions, and ambient lighting, to establish a comprehensive large-scale benchmark of rPPG task under TTA protocol.

VIPL-HR [42] have nine scenarios, three RGB cameras, different illumination conditions, and different levels of movement. It is worth mentioning that the BVP

[4] The proposed self-supervised loss in ConPhys [58] is applied for a comparative analysis of our proposed priors.

Table 1: HR estimation results. Typically, the model selected for adaptation on each dataset is initially pre-trained on the remaining four datasets utilizing the NEST [34] framework. Ours w/o P.R. denotes that only the expert knowledge-based priors is adopted, without the bidirectional adaptation strategy. The best results are highlighted in **bold**, and the second-best results are underlined.

Method	UBFC [2]			PURE [53]			BUAA [68]			VIPL [42]			V4V [49]		
	M↓	R↓	r↑	M↓	R↓	r↑	M↓	R↓	r↑	M↓	R↓	r↑	M↓	R↓	r↑
*Methods based on **DG** protocol*															
Coral [54]	5.89	8.04	0.76	7.59	10.87	0.72	3.64	5.74	0.80	8.68	11.91	0.53	10.32	14.42	0.32
VREx [23]	5.59	7.68	0.81	7.24	10.14	0.78	3.27	5.01	0.86	8.37	11.62	0.54	9.82	14.16	0.37
NCDG [59]	5.31	7.56	0.82	7.32	10.35	0.77	3.12	5.16	0.85	8.47	11.81	0.52	10.14	14.46	0.34
NEST [34]	4.67	6.79	0.86	6.71	9.59	0.81	2.88	4.69	0.89	7.86	11.15	0.58	9.27	13.79	0.41
*Methods based on **TTA** protocol*															
Tent [61]	4.57	9.74	0.87	6.86	9.68	0.84	2.37	3.59	0.92	8.09	11.32	0.55	9.98	11.73	0.47
SAR [40]	4.56	9.64	0.87	6.33	9.30	0.86	2.07	2.93	0.95	8.13	11.45	0.53	9.68	11.46	0.47
TTT++ [32]	4.36	6.79	0.87	6.23	9.14	0.88	2.05	2.82	0.92	7.73	11.08	0.56	9.51	11.31	0.54
EATA [39]	4.25	6.86	0.88	6.13	9.23	0.86	1.89	2.64	0.94	7.69	11.03	0.57	9.36	11.12	0.51
SHOT [29]	4.05	6.45	0.87	5.81	8.86	0.87	1.87	2.47	0.96	7.75	11.07	0.56	9.55	11.32	0.51
AdaODM [72]	4.01	6.59	0.87	5.51	8.29	0.88	1.91	2.60	0.95	7.81	11.21	0.55	9.10	10.93	0.55
ConPhys[4] [58]	3.92	6.61	0.89	6.09	8.83	0.88	1.75	2.19	0.97	7.43	10.70	0.61	9.30	10.96	0.52
Ours w/o P.R.	3.89	6.64	0.88	5.56	8.67	0.88	1.68	2.04	0.99	7.31	10.64	0.62	8.97	10.82	0.55
Ours w/o RS	3.64	6.24	0.89	5.24	8.32	0.89	1.55	1.94	0.99	7.15	10.83	0.63	**8.93**	**10.79**	**0.55**
Ours w/o PA	3.59	9.73	0.87	5.39	8.45	0.88	1.51	1.98	0.99	7.24	10.78	0.62	9.04	10.97	0.54
Ours	**3.54**	**9.71**	**0.87**	**5.02**	**8.11**	**0.89**	**1.49**	**1.97**	**0.99**	**7.09**	**10.72**	**0.63**	9.10	11.02	0.54

signal and video of the dataset do not match in the time dimension. We used the output of HR head as the evaluation result. Besides, we normalize STMap to 30 fps by cubic spline interpolation to solve the problem of the unstable frame rate of video [33, 51, 70].

V4V [49] is designed to collect data with the drastic changes of physiological indicators by simulating ten tasks such as a funny joke, 911 emergency call, and odor experience. It is worth mentioning that There are only heart rate labels in the dataset and no BVP signal labels. We used the output of HR head as the evaluation result.

BUAA [68] is proposed to evaluate the performance of the algorithm against various illumination. We only use data with illumination greater than or equal to 10 lux because underexposed images require special algorithms that are not

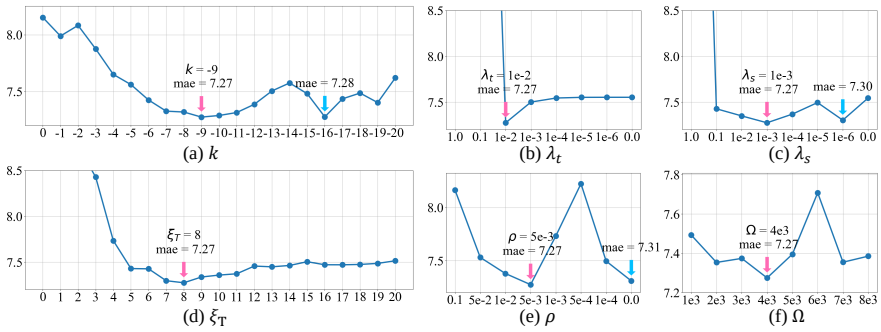


Fig. 5: Ablation experiments of hyper-parameters on down-sampled VIPL dataset[5]. Adheres to the default parameter configuration detailed in Sec.4.2, only one hyper-parameter is varied during each set of experiments. The pink arrows point to the lowest MAE result and the blue arrows mark the second if indiscernible.

considered in this article. We used the output of BVP head as the evaluation for BUAA, PURE, and UBFC-rPPG datasets.

PURE [53] contains 60 RGB videos from 10 subjects with six different activities, specifically, sitting still, talking, and four rotating and moving head variations. The BVP signals are down-sampled from 60 to 30 fps with cubic spline interpolation to align with the videos.

UBFC-rPPG [2] containing 42 face videos with sunlight and indoor illumination, the ground-truth BVP signals, and HR values were collected by CMS50E.

4.2 Implementation Details

We implement our proposed priors and Bi-TTA using the PyTorch framework. The STMap sample is denoted as $\boldsymbol{x} \in \mathbb{R}^{256\times25\times3}$, where 256 is the temporal length, *i.e.*, the number of video frames, of the sliding window, and 25 is the spatial dimension of the STMap. It is firstly resized to $\boldsymbol{x}' \in \mathbb{R}^{256\times64\times3}$ for network input. The network utilizes ResNet-18 [18] as the base architecture, followed by an additional fully connected layer for HR estimation. During the TCL calculation, the temporal perturbation δ_{T} is randomly sampled from a uniform distribution $\mathbb{U}_{(0,59]}$. Furthermore, we use stochastic gradient descent (SGD) with a momentum of 0.9 as the base optimizer. The bidirectional adaptation mechanism is implemented on top of the SGD. During the adaptation process, we set the learning rate to 0.0001 and the batch size is 1. The hyper-parameters are configured as follows: $\lambda_{\mathrm{s}} = 0.001$, $\xi_{\mathrm{T}} = 8.0$, $\lambda_{\mathrm{t}} = 0.01$, $\rho = 0.005$, $k = -9.0$, and $\Omega = 4000$. Please refer to Fig. 5 for detailed ablations.

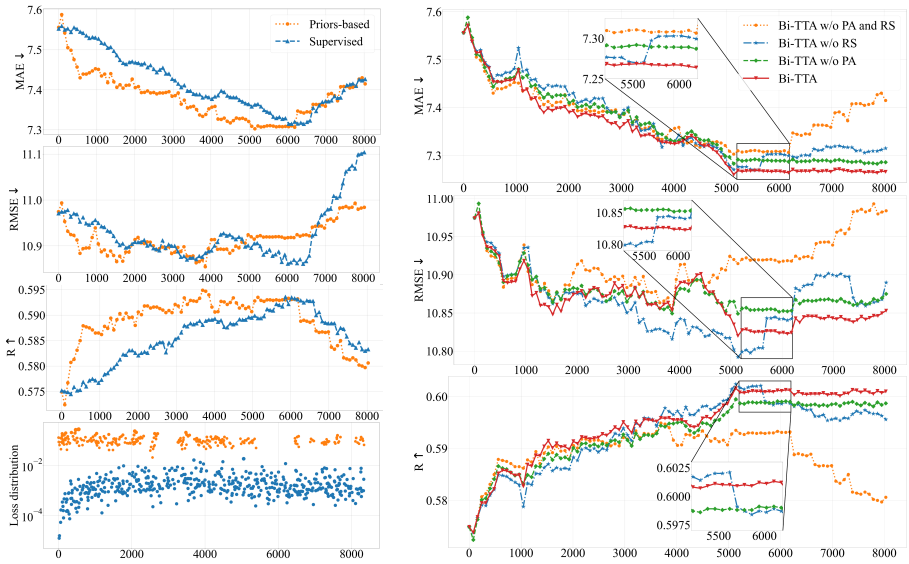
[5] Note that the performance of down-sampled VIPL dataset might be inferior than the full VIPL dataset, due to the larger interval between neighboring samples.

4.3 Results

Following existing rPPG methods [4,33,34,44,51,70], mean absolute error (MAE), root mean square error (RMSE), and Pearson's correlation coefficient (r) are used to evaluate the HR estimation, which inherently reflects the ability to extract physiological periodic signals. The results are organized and recorded in Tab. 1. It is evident that TTA methods generally perform better than DG approaches. Among these TTA methods, our Bi-TTA demonstrates significantly superior performance, showcasing our effectiveness especially when both prospective and retrospective adaptations are synergistically employed. This highlights Bi-TTA's robustness in adapting to various unseen domains, underscoring its potential for real-world applications.

4.4 Ablation Study

In this sub-section, we will mainly discuss about the effectiveness of the proposed priors and the bidirectional adaptation strategy.



(a) Comparison between priors and GT. **(b)** Ablation study of PA, RS, and, priors.

Fig. 6: (a) Comparison between task-specific priors-based TTA and fully supervised adaptation. The evaluation metrics are MSE, RMSE, and r. Alongside the visualization of the loss distribution for both priors-based and supervised adaptation method. Loss values smaller than 10^{-6} are ignored. Note that for better comparison, the learning rate of supervised adaptation is set to 0.0005, while for the priors-based TTA it is 0.0001. **(b) Ablation study of the adaptation performance among the proposed PA and RS modules in our Bi-TTA framework.** This Ablation study demonstrates the effectiveness of the PA and RS, in achieving a more effective and stable adaptation process.

Task-specific priors. As depicted in Fig. 6a, compared to direct ground-truth supervision, our task-specific priors provide effective adaptation supervision only on a significantly smaller number of samples, but induce considerably larger gradients. Thus, we can infer that the convergence rate under our self-supervised priors is markedly faster than ground-truth, despite a substantially lower learning rate. This phenomenon underscores the effectiveness of our priors.

As discussed in [71], modern deep learning models are prone to memorize training data due to their over-parameterized design, leading to overfitting risks. Merely minimizing standard loss functions on training data often falls short in ensuring satisfactory generalization ability [13]. Given that the rPPG model has undergone extensive pre-training with supervised learning, its parameters have likely been excessively optimized. In contrast, task knowledge-based priors are expected to offer fresh and accurate guidance, steering the model towards improved generalization capability.

Prospective and retrospective adaptation. As depicted in Fig. 6b, the initial convergence rate and adaptation performance of the four comparisons are basically consistent across the first 4000 samples. However, in the latter half, the priors-based method demonstrates noticeable performance degradation. In contrast, other three methods, through utilizing prospective or retrospective adaptations, substantially mitigate this problem, thereby enhancing the adaptation ability.

Notably, between the 5000^{th} and 6000^{th} samples, while methods employing prospective adaptation exhibit rapid convergence and superior performance, they also display a significant degree of performance decay due to overfitting or catastrophic forgetting, which even intensifies as more samples are processed. On the other hand, the retrospective adaptation strategy ensures a more stable post-convergence maintenance in the target domain. The Bi-TTA, integrating both prospective and retrospective strategies, adeptly balances convergence speed, effectiveness, and stability.

5 Conclusion

In this paper, we address the challenge of significant performance degradation when adapting rPPG models to unfamiliar target domains. Considering privacy, we implement TTA for rPPG tasks, enabling the adaptation of pre-trained models to the target domain using only unannotated, instance-level target data. The models are fine-tuned during the inference of target samples. In particular, we introduce Bi-TTA, a method that leverages spatial and temporal consistency for effective self-supervision, coupled with pioneering prospective and retrospective adaptation strategies. This methodology streamlines the adaptation process and outperforms existing techniques, demonstrating substantial promise for real-world applications in non-invasive physiological monitoring.

References

1. Bahmani, S., Hahn, O., Zamfir, E., Araslanov, N., Cremers, D., Roth, S.: Semantic self-adaptation: Enhancing generalization with a single sample. arXiv preprint arXiv:2208.05788 (2022) 5
2. Bobbia, S., Macwan, R., Benezeth, Y., Mansouri, A., Dubois, J.: Unsupervised skin tissue segmentation for remote photoplethysmography. Pattern Recognition Letters **124**, 82–90 (2019) 3, 11, 12
3. Brock, A., De, S., Smith, S.L., Simonyan, K.: High-performance large-scale image recognition without normalization. In: International Conference on Machine Learning. pp. 1059–1071. PMLR (2021) 9
4. Chen, W., McDuff, D.: Deepphys: Video-based physiological measurement using convolutional attention networks. In: Proc. ECCV. pp. 349–365 (2018) 4, 13
5. Das, A., Lu, H., Han, H., Dantcheva, A., Shan, S., Chen, X.: Bvpnet: Video-to-bvp signal prediction for remote heart rate estimation. In: FG. pp. 01–08. IEEE (2021) 2
6. De Haan, G., Jeanne, V.: Robust pulse rate from chrominance-based rppg. IEEE Trans. Biomed. Eng. **60**(10), 2878–2886 (2013) 3, 4
7. De Haan, G., Van Leest, A.: Improved motion robustness of remote-ppg by using the blood volume pulse signature. Physiol. Meas. **35**(9), 1913 (2014) 4
8. D'Innocente, A., Bucci, S., Caputo, B., Tommasi, T.: Learning to generalize one sample at a time with self-supervision. arXiv preprint arXiv:1910.03915 (2019) 5
9. Du, J., Liu, S.Q., Zhang, B., Yuen, P.C.: Dual-bridging with adversarial noise generation for domain adaptive rppg estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 10355–10364 (June 2023) 2
10. Du, J., Liu, S.Q., Zhang, B., Yuen, P.C.: Dual-bridging with adversarial noise generation for domain adaptive rppg estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10355–10364 (2023) 5
11. D'Innocente, A., Borlino, F.C., Bucci, S., Caputo, B., Tommasi, T.: One-shot unsupervised cross-domain detection. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XVI 16. pp. 732–748. Springer (2020) 5
12. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: Precup, D., Teh, Y.W. (eds.) Proceedings of the 34th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 70, pp. 1126–1135. PMLR (06–11 Aug 2017), https://proceedings.mlr.press/v70/finn17a.html 5
13. Foret, P., Kleiner, A., Mobahi, H., Neyshabur, B.: Sharpness-aware minimization for efficiently improving generalization. In: International Conference on Learning Representations (2021), https://openreview.net/forum?id=6Tm1mposlrM 9, 14
14. Gandelsman, Y., Sun, Y., Chen, X., Efros, A.A.: Test-time training with masked autoencoders. In: Oh, A.H., Agarwal, A., Belgrave, D., Cho, K. (eds.) Advances in Neural Information Processing Systems (2022), https://openreview.net/forum?id=SHMi1b7sjXk 5
15. Gao, J., Zhang, J., Liu, X., Darrell, T., Shelhamer, E., Wang, D.: Back to the source: Diffusion-driven adaptation to test-time corruption. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11786–11796 (2023) 5

16. Gee-Sern Hsu, ArulMurugan Ambikapathi, M.S.C.: Deep learning with time-frequency representation for pulse estimation. In: Proc. IJCB. pp. 642–650 (2017) 4, 6
17. Gideon, J., Stent, S.: The way to my heart is through contrastive learning: Remote photoplethysmography from unlabelled video. In: Proceedings of the IEEE/CVF international conference on computer vision. pp. 3995–4004 (2021) 5, 6
18. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proc. IEEE CVPR. pp. 770–778 (2016) 8, 12
19. Jia, C., Yang, Y., Xia, Y., Chen, Y.T., Parekh, Z., Pham, H., Le, Q., Sung, Y.H., Li, Z., Duerig, T.: Scaling up visual and vision-language representation learning with noisy text supervision. In: International conference on machine learning. pp. 4904–4916. PMLR (2021) 9
20. Kessler, V., Thiam, P., Amirian, M., Schwenker, F.: Pain recognition with camera photoplethysmography. In: IEEE IPTA. pp. 1–5 (2017) 2
21. Khurana, A., Paul, S., Rai, P., Biswas, S., Aggarwal, G.: Sita: Single image test-time adaptation. arXiv preprint arXiv:2112.02355 (2021) 5
22. Klingner, M., Ayache, M., Fingscheidt, T.: Continual batchnorm adaptation (cbna) for semantic segmentation. IEEE Transactions on Intelligent Transportation Systems **23**(11), 20899–20911 (2022) 5
23. Krueger, D., Caballero, E., Jacobsen, J.H., Zhang, A., Binas, J., Zhang, D., Le Priol, R., Courville, A.: Out-of-distribution generalization via risk extrapolation (rex). In: ICML. pp. 5815–5826. PMLR (2021) 11
24. Lee, E., Chen, E., Lee, C.Y.: Meta-rppg: Remote heart rate estimation using a transductive meta-learner. In: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XXVII 16. pp. 392–409. Springer (2020) 5
25. Lewandowska, M., Rumiński, J., Kocejko, T., Nowak, J.: Measuring pulse rate with a webcam—a non-contact method for evaluating cardiac activity. In: FedCSIS. pp. 405–410. IEEE (2011) 3
26. Li, H., Xu, Z., Taylor, G., Studer, C., Goldstein, T.: Visualizing the loss landscape of neural nets. Advances in neural information processing systems **31** (2018) 1, 2
27. Liang, J., He, R., Tan, T.: A comprehensive survey on test-time adaptation under distribution shifts. arXiv preprint arXiv:2303.15361 (2023) 5
28. Liang, J., Hu, D., Feng, J.: Do we really need to access the source data? source hypothesis transfer for unsupervised domain adaptation. In: International Conference on Machine Learning (ICML). pp. 6028–6039 (2020) 5
29. Liang, J., Hu, D., Wang, Y., He, R., Feng, J.: Source data-absent unsupervised domain adaptation through hypothesis transfer and labeling transfer. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) (2021), in Press 11
30. Liu, X., Jiang, Z., Fromm, J., Xu, X., Patel, S., McDuff, D.: Metaphys: few-shot adaptation for non-contact physiological measurement. In: Proceedings of the conference on health, inference, and learning. pp. 154–163 (2021) 5
31. Liu, X., Zhang, Y., Yu, Z., Lu, H., Yue, H., Yang, J.: rppg-mae: Self-supervised pre-training with masked autoencoders for remote physiological measurement. arXiv preprint arXiv:2306.02301 (2023) 5, 6
32. Liu, Y., Kothari, P., van Delft, B.G., Bellot-Gurlet, B., Mordan, T., Alahi, A.: Ttt++: When does self-supervised test-time training fail or thrive? In: Thirty-Fifth Conference on Neural Information Processing Systems (2021) 11

33. Lu, H., Han, H., Zhou, S.K.: Dual-gan: Joint bvp and noise modeling for remote physiological measurement. In: Proc. IEEE CVPR. pp. 12404–12413 (2021) 2, 4, 6, 11, 13
34. Lu, H., Yu, Z., Niu, X., Chen, Y.C.: Neuron structure modeling for generalizable remote physiological measurement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). pp. 18589–18599 (June 2023) 5, 11, 13
35. McDuff, D.: Camera measurement of physiological vital signs. CSUR (2021) 2
36. McDuff, D.: Applications of camera-based physiological measurement beyond healthcare. In: CVSM, pp. 165–177. Elsevier (2022) 2
37. McDuff, D., Gontarek, S., Picard, R.W.: Improvements in remote cardiopulmonary measurement using a five band digital camera. IEEE Trans. Biomed. Eng. **61**(10), 2593–2601 (2014) 3
38. Mi, P., Shen, L., Ren, T., Zhou, Y., Sun, X., Ji, R., Tao, D.: Make sharpness-aware minimization stronger: A sparsified perturbation approach (2022) 9
39. Niu, S., Wu, J., Zhang, Y., Chen, Y., Zheng, S., Zhao, P., Tan, M.: Efficient test-time model adaptation without forgetting. In: The Internetional Conference on Machine Learning (2022) 5, 11
40. Niu, S., Wu, J., Zhang, Y., Wen, Z., Chen, Y., Zhao, P., Tan, M.: Towards stable test-time adaptation in dynamic wild world. In: Internetional Conference on Learning Representations (2023) 11
41. Niu, X., Han, H., Shan, S., Chen, X.: Synrhythm: Learning a deep heart rate estimator from general to specific. In: Proc. IEEE ICPR. pp. 3580–3585 (2018) 4
42. Niu, X., Han, H., Shan, S., Chen, X.: VIPL-HR: A multi-modal database for pulse estimation from less-constrained face video. In: Proc. ACCV. pp. 562–576 (2018) 2, 10, 11
43. Niu, X., Shan, S., Han, H., Chen, X.: Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation. IEEE Trans. on Image Process. **29**, 2409–2423 (2020) 2, 3, 4, 6
44. Niu, X., Yu, Z., Han, H., Li, X., Shan, S., Zhao, G.: Video-based remote physiological measurement via cross-verified feature disentangling. In: Proc. ECCV (2020) 2, 4, 13
45. Pham, H., Dai, Z., Xie, Q., Le, Q.V.: Meta pseudo labels. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 11557–11568 (2021) 9
46. Poh, M.Z., McDuff, D.J., Picard, R.W.: Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Opt. Express **18**(10), 10762–10774 (2010) 3, 4
47. Qiu, Y., Liu, Y., Arteaga-Falconi, J., Dong, H., El Saddik, A.: Evm-cnn: Real-time contactless heart rate estimation from facial video. IEEE Trans. Multimedia (2019) 2
48. Reiss, A., Indlekofer, I., Schmidt, P., Van Laerhoven, K.: Deep ppg: large-scale heart rate estimation with convolutional neural networks. Sensors **19**(14), 3079 (2019) 4, 6
49. Revanur, A., Li, Z., Ciftci, U.A., Yin, L., Jeni, L.A.: The first vision for vitals (v4v) challenge for non-contact video-based physiological estimation. In: Proc. CVPR workshop. pp. 2760–2767 (2021) 3, 11
50. Song, R., Chen, H., Cheng, J., Li, C., Liu, Y., Chen, X.: Pulsegan: Learning to generate realistic pulse waveforms in remote photoplethysmography. IEEE J-BHI pp. 1–1 (2021) 2

51. Song, R., Zhang, S., Li, C., Zhang, Y., Cheng, J., Chen, X.: Heart rate estimation from facial videos using a spatiotemporal representation with convolutional neural networks. IEEE Trans. Instrum. Meas. (2020) 4, 6, 11, 13
52. Špetlík, R., Franc, V., Matas, J.: Visual heart rate estimation with convolutional neural network. In: Proc. BMVC. pp. 3–6 (2018) 4
53. Stricker, R., Müller, S., Gross, H.M.: Non-contact video-based pulse rate measurement on a mobile service robot. In: Proc. IEEE ISRHIC. pp. 1056–1062 (2014) 3, 11, 12
54. Sun, B., Saenko, K.: Deep coral: Correlation alignment for deep domain adaptation. In: ECCV. pp. 443–450. Springer (2016) 11
55. Sun, B., Wei, Q., Li, L., Xu, Q., He, J., Yu, L.: Lstm for dynamic emotion and group emotion recognition in the wild. In: Proc. ACM ICMI. pp. 451–457 (2016) 2
56. Sun, W., Zhang, X., Lu, H., Chen, Y., Ge, Y., Huang, X., Yuan, J., Chen, Y.: Resolve domain conflicts for generalizable remote physiological measurement. In: Proceedings of the 31st ACM International Conference on Multimedia. pp. 8214–8224 (2023) 5
57. Sun, Y., Wang, X., Liu, Z., Miller, J., Efros, A., Hardt, M.: Test-time training with self-supervision for generalization under distribution shifts. In: International conference on machine learning. pp. 9229–9248. PMLR (2020) 5
58. Sun, Z., Li, X.: Contrast-phys: Unsupervised video-based remote physiological measurement via spatiotemporal contrast. In: European Conference on Computer Vision. pp. 492–510. Springer (2022) 5, 6, 10, 11
59. Tian, C.X., Li, H., Xie, X., Liu, Y., Wang, S.: Neuron coverage-guided domain generalization. IEEE Trans. on Pattern Anal. Mach. Intell. (2022) 11
60. Verkruysse, W., Svaasand, L.O., Nelson, J.S.: Remote plethysmographic imaging using ambient light. Opt. Express **16**(26), 21434–21445 (2008) 3
61. Wang, D., Shelhamer, E., Liu, S., Olshausen, B., Darrell, T.: Tent: Fully test-time adaptation by entropy minimization. In: International Conference on Learning Representations (2021), https://openreview.net/forum?id=uXl3bZLkr3c 3, 5, 11
62. Wang, H., Ahn, E., Kim, J.: Self-supervised representation learning framework for remote physiological measurement using spatiotemporal augmentation loss. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 36, pp. 2431–2439 (2022) 5, 6
63. Wang, J., Lan, C., Liu, C., Ouyang, Y., Qin, T., Lu, W., Chen, Y., Zeng, W., Yu, P.: Generalizing to unseen domains: A survey on domain generalization. IEEE Transactions on Knowledge and Data Engineering (2022) 5
64. Wang, W., den Brinker, A.C., Stuijk, S., de Haan, G.: Algorithmic principles of remote ppg. IEEE Trans. Biomed. Eng. **64**(7), 1479–1491 (2017) 3
65. Wang, W., Stuijk, S., De Haan, G.: Exploiting spatial redundancy of image sensor for motion robust rppg. IEEE Trans. Biomed. Eng. **62**(2), 415–425 (2015) 4
66. Wang, Z.K., Kao, Y., Hsu, C.T.: Vision-based heart rate estimation via a two-stream cnn. In: Proc. IEEE ICIP. pp. 3327–3331 (2019) 4
67. Wu, D., Xia, S.T., Wang, Y.: Adversarial weight perturbation helps robust generalization. Advances in Neural Information Processing Systems **33**, 2958–2969 (2020) 9
68. Xi, L., Chen, W., Zhao, C., Wu, X., Wang, J.: Image enhancement for remote photoplethysmography in a low-light environment. In: FG. pp. 1–7. IEEE (2020) 3, 11

69. Yu, Z., Peng, W., Li, X., Hong, X., Zhao, G.: Remote heart rate measurement from highly compressed facial videos: an end-to-end deep learning solution with video enhancement. In: Proc. IEEE ICCV. pp. 151–160 (2019) 2, 4
70. Yu, Z., Shen, Y., Shi, J., Zhao, H., Torr, P., Zhao, G.: Physformer: Facial video-based physiological measurement with temporal difference transformer. IEEE CVPR (2022) 2, 4, 11, 13
71. Zhang, C., Bengio, S., Hardt, M., Recht, B., Vinyals, O.: Understanding deep learning (still) requires rethinking generalization. Communications of the ACM **64**(3), 107–115 (2021) 14
72. Zhang, X., Chen, Y.C.: Adaptive domain generalization via online disagreement minimization. IEEE Transactions on Image Processing (2023) 5, 11
73. Zhao, X., Liu, C., Sicilia, A., Hwang, S.J., Fu, Y.: Test-time fourier style calibration for domain generalization. arXiv preprint arXiv:2205.06427 (2022) 5
74. Zou, Y., Zhang, Z., Li, C.L., Zhang, H., Pfister, T., Huang, J.B.: Learning instance-specific adaptation for cross-domain segmentation. In: European Conference on Computer Vision. pp. 459–476. Springer (2022) 5