Anchoring Physiological Invariance, Expanding Domain Plasticity: Prior-Stabilized Dynamic Adaptation for Continual rPPG Measurement

Zhongtian He* **Shuyang Chu*** **Jingang Shi†**
School of Software Engineering, Xi'an Jiaotong University
{hezhongtian, xjtucsy}@stu.xjtu.edu.cn, jingang@xjtu.edu.cn

Abstract

Remote photoplethysmography (rPPG) enables non-contact physiological measurement from facial videos, but performance often degrades under unseen environments due to severe domain shifts. Continual learning offers a practical paradigm for updating rPPG models over sequential domains. However, it is challenged by two coupled issues: First, adapting to new domains may overwrite previously acquired physiological representations, resulting in the degradation of physiological consistency across domains. Second, drastic environmental variations induce heterogeneous facial video distributions, making the adaptation process unstable and prone to suboptimal convergence. To address these challenges, we propose **ApexPhys**, a continual rPPG framework that simultaneously **a**nchors **p**hysiological invariance and **ex**pands domain plasticity. Specifically, we introduce a Backbone Consistency Representation Preservation (BCRP) mechanism, which leverages the Fisher Information Matrix to identify and preserve parameters critical to physiological consistency. To enhance adaptation flexibility under diverse environmental shifts, we further propose a Hierarchical Expansion Strategy (HES) that autonomously perceives layer-wise distribution discrepancies and dynamically expands hierarchical adaptation branches. Additionally, we develop a Prior-Guided Pathfinding (PGP) strategy, which leverages parameter inheritance to guide the adaptation along a stable trajectory. Experiments on five public datasets demonstrate that ApexPhys achieves strong anti-forgetting performance and robust adaptation under continual learning. The source code is publicly available at https://github.com/HeZhongTian-xjtu/ApexPhys.

1 Introduction

Physiological signals such as heart rate (HR) and heart rate variability (HRV) are important for clinical diagnosis and health monitoring [3]. However, contact-based electrocardiography (ECG) and photoplethysmography (PPG) sensors are inconvenient for long-term use [31]. This limitation highlights the potential of remote photoplethysmography (rPPG) for HR estimation from facial videos [44, 10, 15]. Although recent deep learning-based rPPG methods have achieved strong performance [16, 45], their accuracy often drops sharply when deployed in unseen environments. This degradation mainly stems from distribution shifts in facial videos induced by uncontrolled factors such as illumination changes, background variations, and motion artifacts [8].

This practical challenge has motivated the exploration of continual learning (CL) [7, 24] for rPPG measurement, where models are updated sequentially as new domains arrive. However, continual

*Equal contribution.
†Corresponding author.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

rPPG measurement is typically hindered by two coupled challenges. First, existing rPPG methods often optimize all parameters for adaptation to the current domain, which may distort the consistency of acquired physiological information, causing performance degradation on previously seen domains. Second, real-world environmental domains typically induce diverse facial video distributions. rPPG models often struggle to adapt to the current domain in a fast and stable manner under drastic distribution shifts, exhibiting unstable optimization and suboptimal convergence. To mitigate this, Liang et al. propose ADDP [23], which alleviates forgetting via prototype-based augmentation that replays past feature characteristics and simplifies inference. However, this framework does not explicitly disentangle physiological and domain information. As a result, adaptation can be unstable when samples from a new domain are limited, making it difficult to balance adaptation and forgetting.

To address the above key challenges (i.e., *what to optimize* and *how to optimize* rPPG models during continual learning), we propose to simultaneously achieve two objectives: **1) Anchoring physiological invariance.** When learning in a novel domain, a pre-trained rPPG model tends to optimize all its parameters uniformly to adapt to the aforementioned environmental factors and user behaviors [23]. However, the periodicity of physiological signals like heart rate shares commonalities across domains [32]. To preserve the acquired physiological knowledge, the parameters encoding such knowledge in the pre-trained model should be explicitly isolated and anchored. **2) Expanding domain plasticity.** In practice, the distribution of facial videos typically experiences drastic and heterogeneous variations with domain shift. To achieve stable and rapid adaptation to such variations, rPPG models should flexibly leverage prior knowledge (e.g., domain features observed in previous domains) to adaptively update the parameters encoding domain knowledge.

Based on the above analysis, we propose ApexPhys, a novel framework designed for stable continual rPPG measurement. First, to explicitly isolate and anchor the parameters encoding physiological knowledge within the pre-trained model, the **Backbone Consistency Representation Preservation (BCRP)** mechanism leverages the Fisher Information Matrix (FIM) to measure the importance of backbone parameters. By freezing these high-importance parameters, BCRP effectively anchors physiological invariance to maintain performance on past domains. Second, to adapt to the drastic and heterogeneous variations in facial video distributions caused by domain shifts, we introduce the **Hierarchical Expansion Strategy (HES)**, which perceives layer-wise environmental distribution shifts to autonomously determine whether to expand new adapters. This distribution-shift-driven self-expansion strategy hierarchically decouples domain features and enables compositional generalization during inference, thereby expanding domain plasticity. This enables our model to rapidly learn new domain-specific patterns to fit diverse facial video distributions when facing continuous domain shifts. Furthermore, to avoid the unstable adaptation to diverse facial video distributions induced by drastic environmental variations, we propose **Prior-Guided Pathfinding (PGP)**, which leverages prior knowledge to find the most stable optimization trajectory. PGP first initializes the new adapter by inheriting parameters from the most similar historical domain to capture prior knowledge from past tasks. Subsequently, it employs a curriculum mechanism that progressively introduces training samples from easy to hard, while simultaneously relaxing constraints to allow the model to gradually deviate from the inherited prior. This explicitly guides the adaptation along a stable trajectory, increasing optimization stability under drastic shifts.

Our contributions are summarized as follows:

- We propose ApexPhys, a prior-stabilized dynamic adaptation framework for continual rPPG measurement. It utilizes the FIM to identify and lock parameters critical for physiological consistency while perceiving environmental shifts to autonomously trigger self-expansion for domain adaptation, thereby ensuring an optimal trade-off between physiological invariance and domain plasticity.

- We introduce a prior-guided optimization strategy. By inheriting priors from similar domains and employing a gradually releasing learning constraint, it explicitly guides the model along a low-forgetting trajectory, ensuring smooth convergence towards the new domain.

- Extensive experiments on a benchmark of five public rPPG datasets demonstrate that ApexPhys achieves state-of-the-art performance, exhibiting superior capabilities in both anti-forgetting and domain adaptation compared to existing methods.

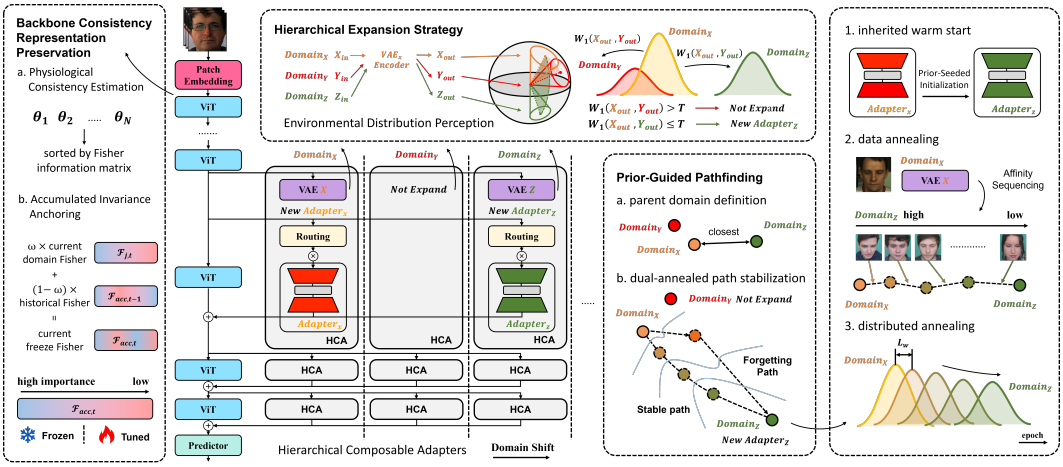


Figure 1: Overview of the **ApexPhys** framework. To anchor physiological invariance, BCRP utilizes the Accumulated Fisher Matrix to selectively freeze consistency-critical parameters, preventing catastrophic forgetting. To expand domain plasticity, HES autonomously perceives layer-wise distribution shifts to trigger adapter instantiation. Complementarily, PGP ensures stable optimization for these new modules via parameter inheritance and dual-annealing, guiding the model along a low-loss trajectory to minimize physiological deviation.

2 Related Work

Remote Physiological Measurement. rPPG measurement detects subtle physiological signals from heartbeat-induced facial color fluctuations. Early studies typically performed rPPG measurement with signal decomposition [2, 38] and color space transformation [35, 2]. However, they are sensitive to real-world noise [28], leading to performance degradation in complex scenarios. With advancements in deep learning (DL), numerous DL-based rPPG methods have been developed [43, 44, 25, 10, 9, 39]. Despite impressive results, they typically rely on the unrealistic assumption that rPPG models are trained and tested on a single dataset.

Continual Learning (CL). Continual learning defines a more realistic setting in which models learn from a stream of tasks sequentially [11]. Prevailing CL approaches generally include 1) replay-based methods [20, 1], which rely on data buffers but raise significant privacy concerns, 2) regularization-based methods [19, 40], which constrain important parameters but struggle to balance the stability-plasticity trade-off, and 3) architecture-based methods [7], which isolate task-specific knowledge but suffer from unbounded parameter growth. Recently, parameter-efficient fine-tuning (PEFT) with pre-trained backbones has become a dominant paradigm [37, 24]. This strategy effectively mitigates forgetting by preserving the backbone while ensuring efficiency through fine-tuned parameters. Following these advances, Liang et al. [23] proposed the first continual rPPG measurement framework, which utilizes domain prototype-based augmentation to replay past feature characteristics. However, it lacks a dedicated mechanism to protect parameters essential for physiological invariance.

3 Methodology

3.1 Problem Formulation

Given a sequence of K domains $(\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_K)$ with distinct acquisition conditions, continual learning requires the rPPG model to learn from this non-stationary domain sequence. Specifically, at the t-th domain, the rPPG model receives N_t labeled data $\mathcal{D}_t = \{(x_i, y_i)\}_{i=1}^{N_t}$, where $x_i \in \mathbb{R}^{3\times T\times H\times W}$ denotes the input video clip, and $y_i \in \mathbb{R}^T$ denotes the corresponding ground-truth PPG signal. The goal is to learn a rPPG model F_Θ parameterized by Θ which performs well on all seen domains $(\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_t)$.

3.2 Overview

As illustrated in Fig. 1, ApexPhys is designed for continual rPPG adaptation under domain shifts, where the model must both adapt to new environments and preserve previously learned physiological knowledge. To achieve a favorable stability–plasticity trade-off, we build ApexPhys on a strong backbone (i.e., Vision Transformer [12]) augmented with self-expanding hierarchical composable adapters (HCA), and organize the adaptation into three coherent stages.

Specifically, when adapting to a new environmental domain $\mathcal{D}_t$, ApexPhys proceeds as follows: **1) Backbone Consistency Representation Preservation**, which identifies and protects backbone parameters that are critical for physiological invariance, so that updates on $\mathcal{D}_t$ do not corrupt shared physiological representations (this step is skipped when $t = 1$). **2) Hierarchical expansion strategy**, which detects distribution shifts and expands HCA modules only when necessary, enabling layer-wise domain adaptation with sub-linear parameter growth (the initial domain $t = 1$ instantiates the first set of adapters). **3) Prior-guided pathfinding**, which initializes newly expanded modules from the most relevant prior and guides optimization along a stable trajectory, reducing training instability when $\mathcal{D}_t$ provides limited samples (this step is not performed when $t = 1$).

Overall, ApexPhys integrates *anchoring physiological invariance* with *expanding domain plasticity* during adaptation, mitigating catastrophic forgetting while enabling rapid domain adaptation. We describe these components below.

3.3 Backbone Consistency Representation Preservation

Remote physiological signals, stemming from skin color variations, are weak and submerged by environmental noise [35]. Consequently, unrestricted backbone updates lead to catastrophic forgetting of consistency representations. To **anchor physiological invariance**, we propose the Backbone Consistency Representation Preservation (BCRP) mechanism, which explicitly identifies and secures parameters critical for signal stability.

Physiological Consistency Estimation. To address physiological representation degradation from uniform updates during domain adaptation, anchoring invariance requires identifying parameters critical for rPPG extraction. This necessitates quantifying backbone parameter importance, for which we employ the Fisher Information Matrix (FIM) [19]. The FIM quantifies model sensitivity to parameter perturbations, where high values identify critical weights whose minor modification would drastically alter predictions. To capture the invariance of physiological representations against environmental changes, we propose deriving the FIM from a Physiological Consistency Loss (L_c).

$$L_c = \frac{1}{N_t} \sum_{i=1}^{N_t} \|\hat{y}_i^{(1)} - \hat{y}_i^{(2)}\|_2^2, \tag{1}$$

where $\hat{y}_i^{(1)}$ and $\hat{y}_i^{(2)}$ denote predictions from two augmented views (i.e., flipping, illumination, and Gaussian noise) to simulate domain shifts. Consequently, on the current domain $\mathcal{D}_t$, the Fisher information value $\mathcal{F}_{j,t}$ for parameter θ_j is approximated by the squared gradient of this loss:

$$\mathcal{F}_{j,t} = \mathbb{E}\left[\left(\frac{\partial L_c}{\partial \theta_j}\right)^2\right]. \tag{2}$$

This explicitly indexes parameters critical for physiological invariance, mapping the backbone's core capabilities.

Accumulated Invariance Anchoring. However, domain shifts can alter the Fisher Information of backbone parameters. Preserving parameters based solely on the current domain's importance may thus lead to forgetting rPPG knowledge from previous domains. To construct a robust, lifespan-aware metric, we extend the Fisher metric to an Accumulated Fisher Matrix ($\mathcal{F}_{acc,t}$) that integrates historical knowledge. It is updated via a parameter-wise moving average rule:

$$\mathcal{F}_{acc,t} \leftarrow \omega \mathcal{F}_{j,t} + (1 - \omega)\mathcal{F}_{acc,t-1}, \tag{3}$$

where ω is the accumulation factor. Here, $\mathcal{F}_{j,t}$ and $\mathcal{F}_{acc,t}$ denote the j-th parameter's estimated and accumulated importance at domain t. Before adapting to a new domain, we rank backbone parameters θ based on $\mathcal{F}_{acc,t-1}$ and freeze the top-ρ ratio of consistency-critical weights. This ensures that core physiological capabilities are explicitly **anchored**, shielding them from catastrophic forgetting while permitting plasticity in non-critical regions.

3.4 Hierarchical Expansion Strategy

While BCRP anchors physiological invariance, it does not explicitly address the adaptation to diverse environmental distributions. According to [23], environmental interference in rPPG manifests with distinct characteristics across different network depths, and a static or globally uniform adaptation strategy fails to decouple these hierarchical domain factors. To effectively **expand domain plasticity**, we propose a Hierarchical Expansion Strategy (HES) that autonomously perceives environmental distribution shifts at each network depth to trigger on-demand adaptation and instantiate composable adapters to structurally accommodate and integrate the new domain knowledge.

Environmental Distribution Perception. To achieve efficient and on-demand plasticity, it is essential to determine when the expansion is necessary. A critical dilemma arises: over-reacting to trivial variations wastes resources, whereas under-reacting to substantial shifts causes failure.

To resolve this, we introduce the Environmental Distribution Perception (EDP) module to autonomously perceive the magnitude of distribution shifts. To decide when to expand, we employ a lightweight distribution descriptor to model the statistics of the intermediate features $x^l \in \mathbb{R}^{patch \times d}$ entering the MLP blocks of the final three layers. This modeling explicitly captures the environmental domain distribution. For simplicity, we implement the descriptor as a standard VAE [18]. When instantiated for domain $\mathcal{D}_t$ at layer l, it is composed of an encoder $E_{\phi^l_{E,t}}$ and a decoder $D_{\phi^l_{D,t}}$. The descriptor is trained by minimizing the perception loss $\mathcal{L}^l_p$, which comprises a reconstruction term and a KL regularization term:

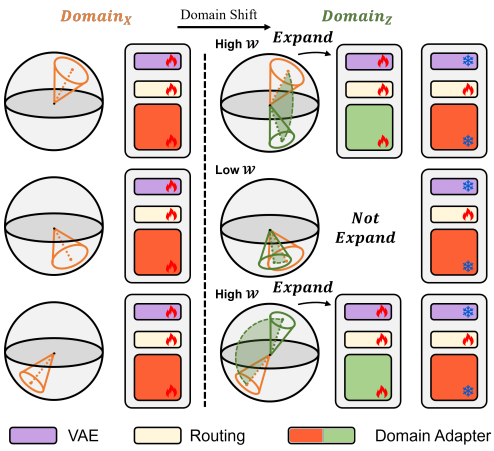


Figure 2: Hierarchical expansion strategy. EDP determines layer-wise adapter reuse or expansion from latent distribution shifts.

$$\mathcal{L}^l_p = \frac{1}{N_t}\sum_{i=1}^{N_t}\left(\mathbb{E}_{z\sim E_{\phi^l_{E,t}}(x^l_i)}\left[\left\|x^l_i - D_{\phi^l_{D,t}}(z)\right\|_2^2\right] + D_{KL}\left(E_{\phi^l_{E,t}}(x^l_i)\,\|\,\mathcal{N}(0,I)\right)\right), \tag{4}$$

where x^l_i denotes the input feature for sample $x_i \in \mathcal{D}_t$.

After perceiving domain $\mathcal{D}_k$, we archive its latent distributions as a historical reference for future comparisons. To capture structural characteristics while suppressing magnitude variations, we project latent features onto the unit hypersphere $\mathbb{S}^d$. The reference set $\mathcal{P}^l_k$ collects the L2-normalized mean vectors from the encoder:

$$\mathcal{P}^l_k = \left\{ \frac{\mu_i}{\|\mu_i\|_2} \mid \mu_i = E_{\phi^l_{E,k}}(x_i), \forall x_i \in \mathcal{D}_k \right\}, \tag{5}$$

where $E_{\phi^l_{E,k}}(\cdot)$ denotes the mean vector output of the encoder. This set $\mathcal{P}^l_k$ compactly represents the distribution manifold of domain $\mathcal{D}_k$ at layer l.

To quantify the distribution shift of the current domain $\mathcal{D}_t$, we map its features into the latent space of each historical encoder to form the query set $\mathcal{Q}^l_{t\to k}$. We measure the discrepancy between this query and the historical reference $\mathcal{P}^l_k$ using the sliced Wasserstein distance (SWD) [5]. This metric approximates the divergence by projecting high-dimensional features onto M unit vectors $\{u_m\}_{m=1}^M$ uniformly sampled from the hypersphere $\mathbb{S}^d$:

$$\mathcal{W}^l_{t,k} = \frac{1}{M}\sum_{m=1}^{M}\frac{1}{\hat{N}}\sum_{i=1}^{\hat{N}}\left|p_i^{(l,k,m)} - q_i^{(l,t\to k,m)}\right|, \tag{6}$$

where $p_i^{(\cdot)}$ and $q_i^{(\cdot)}$ denote the i-th smallest values of $\mathcal{P}^l_k$ and $\mathcal{Q}^l_{t\to k}$ projected onto direction u_m. $\mathcal{W}^l_{t,k}$ serves as a reliable proxy for the environmental variation at layer l. Consequently, the expansion

decision hinges on the minimum distance to the existing memory:

$$\begin{cases} \text{Expand}^l & \text{if } \min_k \mathcal{W}_{t,k}^l > T \\ \text{Not Expand}^l & \text{otherwise,} \end{cases} \tag{7}$$

where T is the expansion threshold. A distance larger than T indicates that the current domain is not well covered by existing descriptors, so HES instantiates a new adapter and descriptor at layer l. Otherwise, the routing mechanism reuses the existing branch.

Hierarchical Composable Adapters. Environmental noise affecting rPPG measurement manifests hierarchically across distinct network depths [23]. To address this, we design the Hierarchical Composable Adapters (HCA) module, aimed at decoupling these diverse environmental factors at their corresponding feature levels. Furthermore, since these factors often emerge in unseen combinations in real-world scenarios, the model is required to dynamically compose this decoupled knowledge for generalization, rather than merely switching between fixed adapters.

When expansion is triggered at layer l, a new adapter $f_{\varphi_t^l}$ is instantiated as a parameter-efficient bottleneck side-branch to the MLP. Defined by down-projection $W_{down}^{(l,t)} \in \mathbb{R}^{d\times r}$ and up-projection $W_{up}^{(l,t)} \in \mathbb{R}^{r\times d}$ parameters, the operation is:

$$f_{\varphi_t^l}(x^l) = \text{ReLU}\left(x^l \cdot W_{down}^{(l,t)}\right) \cdot W_{up}^{(l,t)}, \tag{8}$$

where $\varphi_t^l = \{W_{down}^{(l,t)}, W_{up}^{(l,t)}\}$ denotes the adapter's parameters. To enable compositional generalization, we employ an input-aware routing function $h_{\psi^l}(\cdot)$ parameterized by a learnable matrix $W_{route}^l \in \mathbb{R}^{d\times K_l}$. It computes dynamic weights $\alpha^l = \text{Softmax}(x^l \cdot W_{route}^l)$ to aggregate parallel adapters. The final HCA-enhanced output is:

$$x_{out}^l = \text{MLP}(x^l) + \sum_{k=1}^{K_l} \alpha_k^l \cdot f_{\varphi_k^l}(x^l). \tag{9}$$

The module is optimized end-to-end using a composite objective function $\mathcal{L}_T$. The total supervision loss is formulated as:

$$\mathcal{L}_T = \mathcal{L}_{NP}(y_i, \hat{y}_i) + \mathcal{L}_{CE}(\mathcal{P}(y_i), \mathcal{P}(\hat{y}_i)) + \mathcal{L}_{KL}(\mathcal{P}(y_i), \mathcal{P}(\hat{y}_i)), \quad \hat{y}_i = F_\Theta(x_i), \tag{10}$$

where x_i denotes the input facial video, $\hat{y}_i$ is the predicted rPPG signal, and y_i is the ground-truth signal. Here, $\mathcal{L}_{NP}$ denotes the negative Pearson correlation loss between $\hat{y}_i$ and y_i, which measures their temporal correlation in the time domain. $\mathcal{P}$ denotes the calculation of power spectral density, which aims to transform the rPPG signal from the time domain into the frequency domain. Based on these frequency representations, $\mathcal{L}_{CE}$ and $\mathcal{L}_{KL}$ jointly supervise the frequency-domain prediction.

3.5 Prior-Guided Pathfinding

Recent findings on Linear Mode Connectivity suggest that the optimization trajectory governs forgetting [29]. To guide the model along a stable trajectory and avoid distribution shock, we propose the Prior-Guided Pathfinding (PGP) module, which leverages parameter inheritance and a curriculum with gradually relaxing constraints.

Inherited Warm Start. To avoid the instability of random initialization, which risks convergence to dominant noise, IWS leverages the distributional shifts quantified by EDP to identify the optimal "parent domain" k^* with the minimal distance to the new domain $\mathcal{D}_t$:

$$k^* = \arg\min_{k\in\{1,\ldots,t-1\}} \mathcal{W}_{t,k}^l. \tag{11}$$

Subsequently, we initialize the new adapter via direct inheritance: $\varphi_t^l \leftarrow \varphi_{k^*}^l$. This ensures the pathfinding originates from a robust physiological prior, thereby stabilizing the initial trajectory and accelerating convergence.

Dual-Annealed Path Stabilization. Inheriting a prior is insufficient if the subsequent optimization is unconstrained. To prevent the model from drastically deviating into high-loss regions due to distribution shock, we introduce a dual-annealing strategy that synchronizes input-side data annealing with parameter-side constraint annealing.

Table 1: Comparison of intra-dataset HR estimation results. The symbol ↓ indicates lower is better, and ↑ indicates higher is better. The best and second-best results are marked in **bold** and underlined.

METHOD	VIPL			PURE			MMPD		
	MAE ↓	RMSE ↓	R ↑	MAE ↓	RMSE ↓	R ↑	MAE ↓	RMSE ↓	R ↑
DEEPPHYS [6]	11.0	13.8	0.11	0.83	1.54	**0.99**	—	—	—
PHYSNET [43]	10.8	14.8	0.20	2.95	3.67	**0.99**	7.54	13.0	0.65
GIDEON ET AL. [14]	4.93	7.68	0.81	2.30	2.90	**0.99**	—	—	—
PHYSFORMER [44]	6.23	8.09	0.78	0.50	0.71	**0.99**	8.23	12.1	0.67
UNIFORMER [21]	5.56	7.94	0.79	—	—	—	6.05	9.35	0.79
FACTORIZEPHYS [16]	5.25	9.09	0.68	2.84	5.87	**0.99**	6.71	12.2	0.69
RPPG-MAE [26]	4.52	7.49	0.81	<u>0.17</u>	<u>0.21</u>	**0.99**	—	—	—
RHYTHMMAMBA [45]	<u>4.30</u>	7.49	0.81	0.50	0.75	**0.99**	**3.16**	**7.27**	**0.84**
ADDP [23]	4.84	<u>7.34</u>	<u>0.83</u>	—	—	—	5.70	<u>8.72</u>	0.82
APEXPHYS (OURS)	**3.63**	**5.79**	**0.87**	**0.12**	**0.20**	**0.99**	<u>4.98</u>	10.29	<u>0.83</u>

On the data side, we implement Data Annealing to manage curriculum difficulty. Samples in $\mathcal{D}_t$ are sorted by their SWD distance to the parent distribution, creating a sorted list $\mathcal{D}_t^{sorted}$ that prioritizes samples most similar to the prior. To determine the effective training subset size, we define a linear curriculum schedule $q_{eff}(s)$ that progressively increases the training ratio from q_0 to 1:

$$q_{eff}(s) = \text{clip}\left(q_0 + (1 - q_0)s,\ q_0,\ 1\right), \tag{12}$$

where $s \in [0, 1]$ denotes the normalized progress within the curriculum stage. At each step, the model trains on the top $\lceil q_{eff}(s) \cdot N_t \rceil$ simplest samples from the sorted list, ensuring difficult examples are gradually introduced during the curriculum stage.

Simultaneously, on the distribution side, we apply Constraint Annealing to regulate plasticity. We define the drift statistic $\bar{\delta}$ to monitor the deviation of the current adapter $f_{\varphi_t^l}$ from its parent $f_{\varphi_{k^*}^l}$:

$$\bar{\delta} = \frac{1}{|\mathcal{D}_t(s)|} \sum_{x_i \in \mathcal{D}_t(s)} \mathcal{W}\left(f_{\varphi_t^l}(x_i^l), f_{\varphi_{k^*}^l}(x_i^l)\right). \tag{13}$$

To ensure the optimization path consists of small, gradual steps, we impose a path stability loss $\mathcal{L}_S$ with an exponentially decaying weight:

$$\mathcal{L}_S = \exp(-\kappa \cdot s) \cdot \bar{\delta}, \tag{14}$$

where κ controls the decay rate. This dual mechanism creates a protected optimization corridor, strictly constraining both input difficulty and parameter deviation in the early stages, while annealing these constraints to recover plasticity as the trajectory matures.

3.6 Training Strategy

For the initial domain $\mathcal{D}_1$, we optimize the framework by minimizing $\mathcal{L}_{Init} = \mathcal{L}_p + \mathcal{L}_T$. For subsequent domains $\mathcal{D}_t$, BCRP first applies $\mathcal{F}_{acc}$ to lock consistency-critical parameters. Optimization then adapts based on the expansion decision: if triggered, new modules and non-critical backbone parameters are optimized via $\mathcal{L}_{Total} = \mathcal{L}_p + \mathcal{L}_T + \mathcal{L}_S$, utilizing $\mathcal{L}_S$ for path stability. Conversely, if expansion is bypassed, the model is fine-tuned solely via $\mathcal{L}_T$. Post-training, $\mathcal{F}_{acc}$ is updated with the current domain's Fisher information to refine future preservation.

4 Experiments

4.1 Datasets and Evaluation Protocols

Datasets. We evaluate our method on five public datasets with diverse acquisition conditions. **VIPL** [30] characterizes scenarios with significant head movements and illumination variance, where we exclusively utilize RGB videos. **PURE** [33] specifically targets motion artifacts including talking and rotation. **UBFC** [4] features natural variations in sunlight and indoor lighting. **BUAA** [42] provides controlled illumination levels, from which we select samples with lighting greater than 6.3

Table 2: Comparison of HR estimation under rPPG DIL protocol [23]. ADDP (U) uses the original Uniformer backbone [23], while ADDP (V) uses our ViT backbone for a fair comparison.

METHODS	MAE ↓	RMSE ↓	STD ↓	R ↑
EWC (PNAS '17) [19]	4.38 ± 0.41	7.13 ± 0.83	6.67 ± 0.85	0.75 ± 0.04
LWF (TPAMI '17) [22]	4.26 ± 0.18	6.22 ± 0.65	6.17 ± 0.32	0.76 ± 0.03
S-PROMPTS (NEURIPS '22) [36]	3.77 ± 0.12	6.06 ± 0.25	5.98 ± 0.22	0.80 ± 0.02
LAE-ENSEMBLE (ICCV '23) [13]	7.95 ± 0.22	12.4 ± 0.48	12.2 ± 0.45	0.56 ± 0.02
LAE-OFFLINE (ICCV '23) [13]	7.78 ± 0.33	12.1 ± 0.92	11.9 ± 0.85	0.61 ± 0.04
LAE-ONLINE (ICCV '23) [13]	4.58 ± 0.36	7.02 ± 0.54	6.85 ± 0.51	0.79 ± 0.03
ANCL-EWC (CVPR '23) [17]	4.32 ± 0.21	6.89 ± 0.52	6.63 ± 0.55	0.75 ± 0.02
ANCL-LWF (CVPR '23) [17]	3.93 ± 0.11	6.10 ± 0.15	6.07 ± 0.16	0.79 ± 0.02
ADDP (U) (ECCV '24) [23]	3.70 ± 0.07	5.59 ± 0.02	5.56 ± 0.04	$\mathbf{0.83 \pm 0.01}$
ADDP (V) (ECCV '24) [23]	$\underline{3.64 \pm 0.15}$	$\underline{5.52 \pm 0.24}$	$\underline{5.48 \pm 0.22}$	$\mathbf{0.83 \pm 0.01}$
SD-LORA (ICLR '25) [41]	3.74 ± 0.16	5.81 ± 0.21	5.76 ± 0.28	$\underline{0.81 \pm 0.03}$
IDER (ICLR '26) [27]	3.67 ± 0.11	5.56 ± 0.31	5.52 ± 0.18	$\mathbf{0.83 \pm 0.02}$
APEXPHYS (OURS)	$\mathbf{3.20 \pm 0.13}$	$\mathbf{5.06 \pm 0.20}$	$\mathbf{4.60 \pm 0.16}$	$\mathbf{0.83 \pm 0.01}$

lux. **MMPD** [34] introduces critical domain shifts involving diverse Fitzpatrick skin types under complex mobile-captured lighting and motion conditions.

Evaluation Protocols. Following the domain-incremental learning (DIL) protocol in [23], we note that VIPL Fold 5 exhibits distinct distribution shifts (e.g., specific head motions) compared to Folds 1–4. We therefore use VIPL Fold 1–4 as the initial task, and then continually train the rPPG model on VIPL Fold 5, PURE, UBFC, BUAA, and MMPD in a random order.

For evaluation, we adopt mean absolute error (MAE), root mean squared error (RMSE), standard deviation (Std), and Pearson's correlation coefficient (R) for video-level HR estimation. We report the final incremental performance by:

$$P_N = \frac{1}{N} \sum_{j=1}^{N} p_{N,j}, \tag{15}$$

where $p_{N,j}$ denotes a specific metric evaluated on the test set of the j-th task after learning all N tasks. We repeat experiments three times with different task orders and report mean $\pm$ standard deviation.

Implementation Details. ApexPhys is implemented in PyTorch and optimized with the Adam optimizer using an initial learning rate of 1×10^{-4}, a weight decay of 5×10^{-5}, and a batch size of 8. Input videos are preprocessed into 128×128 facial clips with a temporal window of 160 frames. Training consists of 25 epochs for the initial domain $\mathcal{D}_1$ and 15 epochs for subsequent domains $\mathcal{D}_{2:K}$ to ensure efficient adaptation. The BCRP mechanism secures the top-$\rho = 0.6$ fraction of consistency-critical parameters and updates the Accumulated Fisher Matrix with an $\omega = 0.5$ weighting factor. For the Hierarchical Expansion Strategy, the SWD expansion threshold is set to $T = 0.05$. Regarding the PGP dual-annealing process, data annealing uses an initial ratio $q_0 = 0.3$, adopting curriculum learning in the first half of training and full data thereafter, while constraint annealing employs a path-stability parameter $\kappa = 1.0$.

4.2 Effectiveness of Backbone

We evaluate the fundamental performance of ApexPhys in a non-incremental setting to establish a robust baseline. As shown in Tab. 1, the results demonstrate superior feature-extraction capability in both controlled and complex environments. ApexPhys achieves a Pearson correlation of 0.99 with an MAE of 0.12 and an RMSE of 0.20 on the PURE dataset. Under the challenging conditions of VIPL and MMPD (e.g., varied lighting and head movements), ApexPhys remains robust and reduces the MAE on VIPL to 3.63, outperforming RhythmMamba [45] (4.30). These results confirm that our backbone provides a higher performance ceiling and a high-precision foundation for the continual learning process.

Table 3: Ablation studies on different strategies of BCRP and HES.

(a) Ablation study on backbone consistency representation preservation strategies.

Strategy	MAE ↓	RMSE ↓	R ↑
Backbone Frozen	$4.65_{\pm 0.28}$	$7.20_{\pm 0.45}$	$0.72_{\pm 0.03}$
Random Freezing	$4.05_{\pm 0.45}$	$6.35_{\pm 0.58}$	$0.78_{\pm 0.04}$
Fixed Fisher	$3.68_{\pm 0.19}$	$5.60_{\pm 0.25}$	$0.81_{\pm 0.02}$
BCRP (Ours)	$\mathbf{3.20}_{\pm \mathbf{0.13}}$	$\mathbf{5.06}_{\pm \mathbf{0.20}}$	$\mathbf{0.83}_{\pm \mathbf{0.01}}$

(b) Ablation study on the hierarchical expansion strategy. Exp. is short for Expansion.

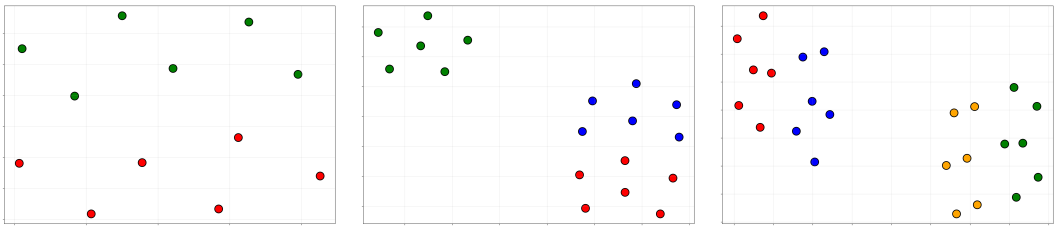
Strategy	MAE ↓	RMSE ↓	R ↑
w/o Exp.	$3.92_{\pm 0.22}$	$6.18_{\pm 0.35}$	$0.78_{\pm 0.02}$
w/ Exp. (Last)	$3.45_{\pm 0.16}$	$5.40_{\pm 0.28}$	$0.81_{\pm 0.02}$
w/ Exp. (Always)	$3.28_{\pm 0.14}$	$5.15_{\pm 0.22}$	$0.82_{\pm 0.01}$
HES (Ours)	$\mathbf{3.20}_{\pm \mathbf{0.13}}$	$\mathbf{5.06}_{\pm \mathbf{0.20}}$	$\mathbf{0.83}_{\pm \mathbf{0.01}}$

4.3 Benchmark Results

We evaluate ApexPhys on the rPPG DIL protocol against seven general CL baselines, i.e., EWC [19], LwF [22], ANCL [17], LAE [13], S-Prompts [36], SD-LoRA [41], and IDER [27], as well as ADDP [23], a prior continual rPPG method. We follow the training and evaluation protocol of ADDP [23] in all experiments. ADDP is assessed under two configurations: its native Uniformer backbone ADDP (U) and our ViT backbone ADDP (V) for architectural comparison.

The benchmark results in Tab. 2 show that ApexPhys outperforms all baselines, achieving an MAE of 3.20 ± 0.13 and an RMSE of 5.06 ± 0.20. Compared with general continual learning methods, ApexPhys achieves the best overall performance on the rPPG DIL protocol. Compared with the prior continual rPPG method ADDP, ApexPhys further reduces MAE and RMSE by 12.1% and 8.3%, respectively, under the same ViT backbone, while also yielding a lower error standard deviation.

These results highlight the limitations of existing methods for continual rPPG. General continual learning methods are less effective for rPPG, where subtle physiological signals are highly susceptible to environmental domain shifts. Although ADDP is tailored for continual rPPG and improves over generic CL baselines, its single-prototype inference strategy remains limited when handling complex overlapping factors. In contrast, ApexPhys combines BCRP, HES, and PGP to achieve a better invariance–plasticity trade-off.



(a) Layer 1 (2 distribution descriptors) (b) Layer 2 (3 distribution descriptors) (c) Layer 3 (4 distribution descriptors)

Figure 3: Visualization of Latent Distributions in HES. Distinct colors denote different Adapter paths activated by the HES decision logic. Points of the same color represent domains where the Sliced Wasserstein Distance (SWD) between descriptors was below the threshold T, triggering the *Reuse* mechanism. Separated color clusters indicate distinct latent distributions that triggered the *Expansion* mechanism.

4.4 Ablation Study

Effectiveness of BCRP. Tab. 3a compares our Accumulated Fisher strategy with Backbone Frozen, Random Freezing, and Fixed Fisher. Backbone Frozen yields the worst performance (MAE 4.65), indicating that completely restricting backbone updates limits the model's ability to adapt to new domains. Random Freezing improves the MAE to 4.05 but exhibits relatively high variance (± 0.45), suggesting that preserving an arbitrary subset of parameters is insufficient for stable continual adaptation. Fixed Fisher, which determines parameter importance using only the initial domain without subsequent updates, achieves a lower MAE of 3.68 but cannot account for changing parameter importance across domains. In contrast, BCRP dynamically accumulates Fisher importance throughout the

Table 4: Component-wise ablation of the prior-guided pathfinding, where PI denotes Parameter Inheritance, DA represents Data Annealing, and CA stands for Constraint Annealing.

PI	DA	CA	MAE ↓	RMSE ↓	R ↑
×	×	×	$3.50_{\pm 0.32}$	$5.65_{\pm 0.45}$	$0.80_{\pm 0.03}$
✓	×	×	$3.32_{\pm 0.18}$	$5.25_{\pm 0.26}$	$0.82_{\pm 0.02}$
✓	✓	×	$3.25_{\pm 0.15}$	$5.12_{\pm 0.23}$	$0.82_{\pm 0.01}$
✓	✓	✓	$\mathbf{3.20}_{\pm \mathbf{0.13}}$	$\mathbf{5.06}_{\pm \mathbf{0.20}}$	$\mathbf{0.83}_{\pm \mathbf{0.01}}$

domain sequence and achieves the best performance (MAE 3.20, ± 0.13). These results demonstrate the importance of selectively preserving consistency-critical parameters while retaining sufficient backbone plasticity.

Effectiveness of Hierarchical Expansion Strategy. Tab. 3b compares HES with No Expansion, Last-layer Expand, and Always Expand. No Expansion performs poorly (MAE 3.92, ± 0.22), suggesting that a static model struggles under domain shifts. Last-layer Expand improves the result (MAE 3.45, ± 0.16) but remains limited, and Always Expand still falls short of our method, indicating that indiscriminate expansion is suboptimal. In contrast, HES achieves the best performance (MAE 3.20, ± 0.13), showing that selective hierarchical expansion adapts more effectively to domain variation while maintaining better stability and parameter efficiency.

To further validate the effectiveness of the Environmental Distribution Perception (EDP) module within our Hierarchical Expansion Strategy (HES), we visualize the latent representations learned by the VAE across different network depths. Figure 3 presents the t-SNE projection of the latent distribution descriptors z extracted from the six evaluation domains. Closely clustered distributions indicate similar environmental characteristics, where a small SWD ($\mathcal{W} < T$) favors adapter reuse without unnecessary expansion. Conversely, separated clusters suggest larger distribution shifts ($\mathcal{W} > T$), motivating the expansion of domain-specific adapters. Moreover, the number of latent distributions increases from two in Layer 1 to four in Layer 3, indicating layer-dependent environmental variations. These observations support the proposed layer-wise expansion strategy, which adaptively allocates model capacity according to distribution shifts.

Effectiveness of PGP. Tab. 4 shows that PGP consistently improves both accuracy and stability. Without these components, random initialization produces an MAE of 3.50 with relatively high variance (± 0.32), reflecting the instability of adaptation without guidance from historical knowledge. Adding Parameter Inheritance reduces the variance to ± 0.18 and lowers MAE to 3.32, showing the benefit of initializing from a relevant prior. Data Annealing further improves optimization and reduces MAE to 3.25. The full PGP with Constraint Annealing achieves the best result (MAE 3.20, ± 0.13), indicating that these components provide a more stable and guided adaptation path. Together, these results demonstrate that PI, DA, and CA contribute complementary benefits to stable optimization under sequential domain shifts.

5 Conclusion and Limitations

In this paper, we presented ApexPhys, which aims address the challenges of *what to optimize* and *how to optimize* rPPG models in continual learning. Specifically, Backbone Consistency Representation Preservation (BCRP) leverages the Accumulated Fisher Matrix to protect consistency-critical backbone parameters, while the Hierarchical Expansion Strategy (HES) uses Environmental Distribution Perception (EDP) to trigger on-demand, layer-wise adapter expansion under distribution shifts. Finally, Prior-Guided Pathfinding (PGP) stabilizes optimization via parameter inheritance and a dual-annealing schedule. Following the current continual learning protocol, experiments on five public datasets demonstrate that ApexPhys achieves strong adaptation with reduced forgetting, providing a practical solution for non-contact physiological monitoring. Further validation on more diverse real-world domains would help strengthen the generality of these conclusions.

Acknowledgments and Disclosure of Funding

Funding in direct support of this work: This paper is funded by the National Natural Science Foundation of China (No. 62576273), the Key Research and Development Program of Shaanxi Province (No. 2026JC-ZDGG-20), and the Xiaomi Foundation.

Competing Interests: The authors declare no competing interests.

References

[1] Agil Aghasanli, Yi Li, and Plamen Angelov. Prototype-based continual learning with label-free replay buffer and cluster preservation loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 6611–6620, 2025.

[2] Guha Balakrishnan, Fredo Durand, and John Guttag. Detecting pulse from head motions in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3430–3437, 2013.

[3] Yannick Benezeth, Peixi Li, Richard Macwan, Keisuke Nakamura, Randy Gomez, and Fan Yang. Remote heart rate variability for emotional state monitoring. In *Proceedings of the IEEE-EMBS International Conference on Biomedical and Health Informatics*, pages 153–156, 2018.

[4] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 124:82–90, 2019.

[5] Nicolas Bonneel, Julien Rabin, Gabriel Peyré, and Hanspeter Pfister. Sliced and radon wasserstein barycenters of measures. *Journal of Mathematical Imaging and Vision*, 51(1):22–45, 2015.

[6] Weixuan Chen and Daniel McDuff. Deepphys: Video-based physiological measurement using convolutional attention networks. In *Proceedings of the European Conference on Computer Vision*, pages 356–373, 2018.

[7] De Cheng, Yusong Hu, Nannan Wang, Dingwen Zhang, and Xinbo Gao. Achieving plasticity-stability trade-off in continual learning through adaptive orthogonal projection. *IEEE Transactions on Circuits and Systems for Video Technology*, 35(8):7485–7498, 2025.

[8] Shuyang Chu, Jingang Shi, Xu Cheng, Haoyu Chen, Xin Liu, Jian Xu, and Guoying Zhao. To remember, to adapt, to preempt: A stable continual test-time adaptation framework for remote physiological measurement in dynamic domain shifts. In *Proceedings of the ACM International Conference on Multimedia*, pages 7307–7316, 2025.

[9] Shuyang Chu, Jingang Shi, Mengyao Yuan, Xuqi Li, Zhengdong Jiang, and Guoying Zhao. Satphys: Sandglass transformer for efficient video-based remote physiological measurement. *IEEE Transactions on Circuits and Systems for Video Technology*, 36(8):11402–11415, 2026.

[10] Shuyang Chu, Menghan Xia, Mengyao Yuan, Xin Liu, Tapio Seppänen, Guoying Zhao, and Jingang Shi. CodePhys: Robust video-based remote physiological measurement through latent codebook querying. *IEEE Journal of Biomedical and Health Informatics*, 29(7):4932–4945, 2025.

[11] Matthias De Lange, Rahaf Aljundi, Marc Masana, Sarah Parisot, Xu Jia, Aleš Leonardis, Gregory Slabaugh, and Tinne Tuytelaars. A continual learning survey: Defying forgetting in classification tasks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7):3366–3385, 2022.

[12] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, Jakob Uszkoreit, and Neil Houlsby. An image is worth 16x16 words: Transformers for image recognition at scale. In *Proceedings of the International Conference on Learning Representations*, 2021.

[13] Qiankun Gao, Chen Zhao, Yifan Sun, Teng Xi, Gang Zhang, Bernard Ghanem, and Jian Zhang. A unified continual learning framework with general parameter-efficient tuning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 11483–11493, 2023.

[14] John Gideon and Simon Stent. 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*, pages 3995–4004, 2021.

[15] Zhongtian He, Shuyang Chu, Xuqi Li, Zhengdong Jiang, Jian Xu, Guoying Zhao, and Jingang Shi. Resophys: Unsupervised plug-and-play remote physiological measurement via facial videos of arbitrary resolution. *IEEE Journal of Biomedical and Health Informatics*, 30(8):6904–6917, 2026.

[16] Jitesh Joshi, Sos S Agaian, and Youngjun Cho. Factorizephys: Matrix factorization for multidimensional attention in remote physiological sensing. In *Advances in Neural Information Processing Systems*, volume 37, pages 96607–96639, 2024.

[17] Sanghwan Kim, Lorenzo Noci, Antonio Orvieto, and Thomas Hofmann. Achieving a better stability-plasticity trade-off via auxiliary networks in continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 11930–11939, 2023.

[18] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *Proceedings of the International Conference on Learning Representations*, 2014.

[19] James Kirkpatrick, Razvan Pascanu, Neil Rabinowitz, Joel Veness, Guillaume Desjardins, Andrei A. Rusu, Kieran Milan, John Quan, Tiago Ramalho, Agnieszka Grabska-Barwinska, Demis Hassabis, Claudia Clopath, Dharshan Kumaran, and Raia Hadsell. Overcoming catastrophic forgetting in neural networks. *Proceedings of the National Academy of Sciences*, 114(13):3521–3526, 2017.

[20] Soochan Lee, Junsoo Ha, Dongsu Zhang, and Gunhee Kim. A neural dirichlet process mixture model for task-free continual learning. In *Proceedings of the International Conference on Learning Representations*, 2020.

[21] Kunchang Li, Yali Wang, Junhao Zhang, Peng Gao, Guanglu Song, Yu Liu, Hongsheng Li, and Yu Qiao. Uniformer: Unifying convolution and self-attention for visual recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 45(10):12581–12600, 2023.

[22] Zhizhong Li and Derek Hoiem. Learning without forgetting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 40(12):2935–2947, 2018.

[23] Qian Liang, Yan Chen, and Yang Hu. Continual learning for remote physiological measurement: Minimize forgetting and simplify inference. In *Proceedings of the European Conference on Computer Vision*, pages 126–144, 2024.

[24] Wenzhuo Liu, Fei Zhu, and Cheng-Lin Liu. Branch-tuning: Balancing stability and plasticity for continual self-supervised learning. *IEEE Transactions on Neural Networks and Learning Systems*, 36(10):19312–19326, 2025.

[25] Xin Liu, Brian Hill, Ziheng Jiang, Shwetak Patel, and Daniel McDuff. Efficientphys: Enabling simple, fast and accurate camera-based cardiac measurement. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5008–5017, 2023.

[26] Xin Liu, Yuting Zhang, Zitong Yu, Hao Lu, Huanjing Yue, and Jingyu Yang. rPPG-MAE: Self-supervised pretraining with masked autoencoders for remote physiological measurements. *IEEE Transactions on Multimedia*, 26:7278–7293, 2024.

[27] Zhanwang Liu, Yuting Li, Haoyuan Gao, Yexin Li, Linghe Kong, Lichao Sun, and Weiran Huang. Ider: Idempotent experience replay for reliable continual learning. In *Proceedings of the International Conference on Learning Representations*, 2026.

[28] Daniel J McDuff, Ethan B Blackford, and Justin R Estepp. Fusing partial camera signals for noncontact pulse rate variability measurement. *IEEE Transactions on Biomedical Engineering*, 65(8):1725–1739, 2018.

[29] Seyed Iman Mirzadeh, Mehrdad Farajtabar, Dilan Gorur, Razvan Pascanu, and Hassan Ghasemzadeh. Linear mode connectivity in multitask and continual learning. In *Proceedings of the International Conference on Learning Representations*, 2021.

[30] Xuesong Niu, Hu Han, Shiguang Shan, and Xilin Chen. Vipl-hr: A multi-modal database for pulse estimation from less-constrained face video. In *Proceedings of the Asian Conference on Computer Vision*, pages 562–576, 2018.

[31] Xuesong Niu, Shiguang Shan, Hu Han, and Xilin Chen. Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation. *IEEE Transactions on Image Processing*, 29:2409–2423, 2020.

[32] Xuesong Niu, Zitong Yu, Hu Han, Xiaobai Li, Shiguang Shan, and Guoying Zhao. Video-based remote physiological measurement via cross-verified feature disentangling. In *Proceedings of the European Conference on Computer Vision*, pages 295–310, 2020.

[33] Ronny Stricker, Steffen Müller, and Horst-Michael Gross. Non-contact video-based pulse rate measurement on a mobile service robot. In *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*, pages 1056–1062, 2014.

[34] Jiankai Tang, Kequan Chen, Yuntao Wang, Yuanchun Shi, Shwetak Patel, Daniel McDuff, and Xin Liu. Mmpd: Multi-domain mobile video physiology dataset. In *Proceedings of the Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1–5, 2023.

[35] Wenjin Wang, Albertus C Den Brinker, Sander Stuijk, and Gerard De Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2017.

[36] Yabin Wang, Zhiwu Huang, and Xiaopeng Hong. S-prompts learning with pre-trained transformers: An occam's razor for domain incremental learning. In *Advances in Neural Information Processing Systems*, volume 35, pages 5682–5695, 2022.

[37] Zifeng Wang, Zizhao Zhang, Chen-Yu Lee, Han Zhang, Ruoxi Sun, Xiaoqi Ren, Guolong Su, Vincent Perot, Jennifer Dy, and Tomas Pfister. Learning to prompt for continual learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 139–149, 2022.

[38] Christian Wiede, Julia Richter, André Apitzsch, Fajer KhairAldin, and Gangolf Hirtz. Remote heart rate determination in rgb data-an investigation using independent component analysis and adaptive filtering. In *Proceedings of the International Conference on Pattern Recognition Applications and Methods*, volume 1, pages 240–246, 2016.

[39] Junjie Wu, Shuyang Chu, Jian Xu, and Jingang Shi. Mstrans: A multi-scale spatio-temporal transformer-based framework for robust rPPG measurement. *Computer Vision and Image Understanding*, page 104983, 2026.

[40] Yichen Wu, Long-Kai Huang, Renzhen Wang, Deyu Meng, and Ying Wei. Meta continual learning revisited: Implicitly enhancing online hessian approximation via variance reduction. In *Proceedings of the International Conference on Learning Representations*, 2024.

[41] Yichen Wu, Hongming Piao, Long-Kai Huang, Renzhen Wang, Wanhua Li, Hanspeter Pfister, Deyu Meng, Kede Ma, and Ying Wei. Sd-lora: Scalable decoupled low-rank adaptation for class incremental learning. In *Proceedings of the International Conference on Learning Representations*, 2025.

[42] Lin Xi, Weihai Chen, Changchen Zhao, Xingming Wu, and Jianhua Wang. Image enhancement for remote photoplethysmography in a low-light environment. In *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, pages 761–764, 2020.

[43] Zitong Yu, Xiaobai Li, and Guoying Zhao. Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks. In *Proceedings of the British Machine Vision Conference*, pages 1–12, 2019.

[44] Zitong Yu, Yuming Shen, Jingang Shi, Hengshuang Zhao, Philip HS Torr, and Guoying Zhao. Physformer: Facial video-based physiological measurement with temporal difference transformer. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 4186–4196, 2022.

[45] Bochao Zou, Zizheng Guo, Xiaocheng Hu, and Huimin Ma. Rhythmmamba: Fast, lightweight, and accurate remote physiological measurement. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, pages 11077–11085, 2025.

A Training Algorithm Detail

In this section, we provide the comprehensive training procedure of the proposed ApexPhys framework (Algorithm 1). The training process is structured into two distinct phases to ensure both a robust initialization and stable continuous adaptation.

In the **Initial Phase**, the model establishes a fundamental representation capability on the first domain by minimizing the joint initial objective $\mathcal{L}_{Init}$. The subsequent **Continual Phase** employs a cyclic mechanism: the *BCRP* module first applies a freezing mask to protect physiological invariance; the *HES* module then quantifies distribution shifts to make expansion decisions; finally, the *PGP* strategy guides the optimization of new components using a composite stability objective $\mathcal{L}_{Total}$ or a task-specific objective $\mathcal{L}_T$, balancing stability and plasticity.

Algorithm 1 ApexPhys Training Procedure

Input: Sequential Domain Stream $\mathcal{S} = \{\mathcal{D}_1, \ldots, \mathcal{D}_K\}$.
Input: Hyperparameters: Freezing Rate ρ, Accumulation Factor ω, Expansion Threshold T, Initial Curriculum Ratio q_0, and Constraint Decay Rate κ.
Output: Final Optimized Model Parameters Θ.
Phase 1: Initial Deployment ($\mathcal{D}_1$)
Optimization: Jointly train Backbone θ, initial Adapter $\mathcal{A}_1$, VAE Φ_1, and Routing Ψ on $\mathcal{D}_1$ by minimizing:
  $\mathcal{L}_{Init} = \mathcal{L}_p + \mathcal{L}_T$
Initialization: Compute Fisher Information Matrix (FIM) on $\mathcal{D}_1$ and initialize Accumulated Fisher $\mathcal{F}_{acc,1}$.
Phase 2: Continual Adaptation ($\mathcal{D}_t, t \geq 2$)
for $t = 2$ **to** K **do**
  // Step 1: BCRP Pre-training Protection
  Rank backbone parameters based on $\mathcal{F}_{acc,t-1}$ and generate mask to **freeze** top-ρ critical parameters.
  // Step 2: HES-Guided Expansion & Optimization
  for each adaptable layer l **do**
    Compute SWD shifts $\mathcal{W}^l_{t,k}$ between $\mathcal{D}_t$ and history.
    if $\min_k \mathcal{W}^l_{t,k} > T$ **then**
      Expand: Instantiate new Adapter $\mathcal{A}^l_t$ and VAE Φ^l_t.
      Inherited Warm Start: Initialize $\mathcal{A}^l_t$ from parent k^*.
      Dual-Annealed Optimization: Minimize composite stability objective: $\mathcal{L}_{Total} = \mathcal{L}_p + \mathcal{L}_T + \mathcal{L}_S$
    else
      Reuse: Reuse existing modules and fine-tune Routing Ψ^l by minimizing: $\mathcal{L}_T$
    end if
  end for
  Update trainable parameters via Adam.
  // Step 3: BCRP Post-training Consolidation
  Compute FIM $\mathcal{F}_t$ on current domain $\mathcal{D}_t$.
  Update global memory: $\mathcal{F}_{acc,t} \leftarrow \omega\mathcal{F}_t + (1-\omega)\mathcal{F}_{acc,t-1}$.
end for

Compute Resources. All experiments were conducted on a single NVIDIA GeForce RTX 3090 GPU. Under the reported configuration, the peak GPU memory usage was approximately 10000MB, and a complete continual training run typically took about 8 hours.

B Impact of Domain Arrival Order

In real-world deployments, the sequence in which new domains are encountered is often unpredictable. To evaluate the robustness of ApexPhys against such variations, we conducted experiments on six distinct permutations of the domain arrival order, as detailed in Tab. 5. The training sequences involve a diverse set of datasets: VIPL-HR (divided into Folds 1-4 for initial training and Fold 5 for testing adaptation), BUAA, PURE, UBFC-rPPG, and MMPD.

As presented in Tab. 5, ApexPhys demonstrates strong resilience to changes in task order. The Mean Absolute Error (MAE) fluctuates within a narrow margin (mean 3.19 ± 0.18 bpm) across all six permutations. Regardless of whether the domain sequence introduces significant distribution shifts early or late in the process, our framework consistently maintains high accuracy. This stability is primarily attributed to the *Prior-Guided Pathfinding (PGP)* module, which effectively identifies the most physiologically relevant historical path for initialization, and the *BCRP* mechanism, which preserves core physiological representations against catastrophic forgetting.

It is important to note that the results presented in Tab. 5 cover six specific, manually constructed permutation sequences to rigorously test robustness. These values may differ slightly from the metrics reported in the main text, where the results were averaged over three randomly sampled sequences.

While the exact numerical values vary due to the different sampling strategies, the overall trend of superior performance and stability remains consistent across both evaluation protocols.

Table 5: **Robustness against Task Order Permutations.** We evaluate ApexPhys on six distinct sequences of domain arrival. The sequences involve VIPL-HR (divided into Fold 1-4 and Fold 5), BUAA, PURE, UBFC, and MMPD. The results demonstrate that our method maintains stable performance regardless of the training order.

Domain Arrival Order	**MAE ↓**	**RMSE ↓**
$\text{VIPL}_{\text{Fold1-4}} \rightarrow \text{BUAA} \rightarrow \text{PURE} \rightarrow \text{UBFC} \rightarrow \text{MMPD} \rightarrow \text{VIPL}_{\text{Fold5}}$	3.01	4.78
$\text{VIPL}_{\text{Fold1-4}} \rightarrow \text{UBFC} \rightarrow \text{BUAA} \rightarrow \text{PURE} \rightarrow \text{MMPD} \rightarrow \text{VIPL}_{\text{Fold5}}$	3.03	5.01
$\text{VIPL}_{\text{Fold1-4}} \rightarrow \text{UBFC} \rightarrow \text{PURE} \rightarrow \text{BUAA} \rightarrow \text{MMPD} \rightarrow \text{VIPL}_{\text{Fold5}}$	3.09	5.03
$\text{VIPL}_{\text{Fold1-4}} \rightarrow \text{VIPL}_{\text{Fold5}} \rightarrow \text{BUAA} \rightarrow \text{PURE} \rightarrow \text{UBFC} \rightarrow \text{MMPD}$	3.18	5.19
$\text{VIPL}_{\text{Fold1-4}} \rightarrow \text{VIPL}_{\text{Fold5}} \rightarrow \text{BUAA} \rightarrow \text{MMPD} \rightarrow \text{PURE} \rightarrow \text{UBFC}$	3.31	5.25
$\text{VIPL}_{\text{Fold1-4}} \rightarrow \text{PURE} \rightarrow \text{UBFC} \rightarrow \text{BUAA} \rightarrow \text{VIPL}_{\text{Fold5}} \rightarrow \text{MMPD}$	3.53	5.51
Mean ± Std. Dev.	*3.19 ± 0.18*	*5.13 ± 0.23*

C Per-Domain Performance and Forgetting Analysis

While the main text reports the final average incremental performance (Tab. 2), it is equally important to analyze the micro-level forgetting dynamics. Providing a per-domain analysis helps to characterize how different environmental shifts independently impact catastrophic forgetting throughout the continuous adaptation process.

To evaluate this, we track two specific metrics for each domain during the continual learning stream: **Best MAE**, which represents the evaluation error immediately after the model learns the specific domain, and **Last MAE**, which represents the final evaluation error after the model has completed the entire domain sequence. The discrepancy between these two metrics, denoted as **Forgetting (Δ)**, formally quantifies the degree of knowledge corruption.

As demonstrated in Tab. 6, the magnitude of forgetting is intrinsically tied to the environmental complexity of the respective domains. Datasets featuring relatively controlled conditions (e.g., BUAA, PURE, and UBFC) exhibit marginal forgetting, maintaining their high precision throughout the sequence. Conversely, domains such as VIPL and MMPD introduce severe distribution shifts. VIPL contains drastic illumination variations and large head motions, while MMPD features diverse Fitzpatrick skin tones and mobile-captured artifacts. These complex, heterogeneous noise patterns naturally make the preservation of physiological invariance significantly more challenging, leading to larger intrinsic forgetting (e.g., VIPL Folds 1-4 yields the highest Δ of +2.52).

Despite the inherent difficulty of these severe distribution shifts, ApexPhys demonstrates strong resilience. By leveraging the Backbone Consistency Representation Preservation (BCRP) and Hierarchical Expansion Strategy (HES) mechanisms, our framework explicitly isolates and protects core physiological representations from being overwritten by subsequent environmental noise. Consequently, ApexPhys effectively bounds the catastrophic forgetting even on the most challenging domains, ensuring stable knowledge retention and ultimately achieving state-of-the-art anti-forgetting capabilities across the entire continual learning stream.

Table 6: Per-Domain Performance and Forgetting Analysis (evaluated via MAE ↓). **Best** denotes the performance immediately after learning the specific domain, while **Last** denotes the performance after learning the entire domain sequence. The **Forgetting (Δ)** row explicitly quantifies the performance degradation. Note: In this evaluated sequence, MMPD was the final task, hence its Best and Last MAEs are naturally identical.

Metric	VIPL1-4	VIPL5	BUAA	PURE	UBFC	MMPD
Best	4.01	2.77	1.74	0.20	0.04	5.29
Last	6.53	4.43	2.00	0.67	0.17	5.29
Forgetting (Δ)	+2.52	+1.66	+0.26	+0.47	+0.13	+0.00

D Sensitivity Analysis of BCRP Freezing Rate ρ

To validate the configuration of the Backbone Consistency Representation Preservation (BCRP) module, we investigated the impact of the freezing rate ρ on the stability-plasticity trade-off. As shown in Tab. 7, the model's performance exhibits a clear convex trend with respect to ρ. Lower freezing rates ($\rho < 0.4$) provide insufficient protection for the learned representations, leading to catastrophic forgetting where new gradients overwrite historical physiological priors (e.g., MAE degrades to 4.12 bpm at $\rho = 0.0$). Conversely, excessive freezing ($\rho > 0.8$) renders the backbone too rigid to adapt to domain-specific environmental shifts, resulting in underfitting. The optimal equilibrium is achieved at $\rho = 0.6$, where the mechanism effectively anchors critical physiological parameters while retaining sufficient flexibility for new domain alignment.

Table 7: Sensitivity analysis of the Freezing Rate ρ in the BCRP module. We investigate how the proportion of frozen backbone parameters affects the stability-plasticity trade-off. The optimal performance is achieved at $\rho = 0.6$.

FREEZING RATE (ρ)	MAE ↓	RMSE ↓	R ↑
$\rho = 0.0$ (TRAIN ALL)	4.12 ± 0.35	6.45 ± 0.42	0.76 ± 0.03
$\rho = 0.2$	3.65 ± 0.22	5.68 ± 0.28	0.80 ± 0.02
$\rho = 0.4$	3.35 ± 0.15	5.25 ± 0.21	0.82 ± 0.01
$\rho = 0.6$ (OURS)	$\mathbf{3.20 \pm 0.13}$	$\mathbf{5.06 \pm 0.20}$	$\mathbf{0.83 \pm 0.01}$
$\rho = 0.8$	3.88 ± 0.24	5.92 ± 0.30	0.79 ± 0.02
$\rho = 1.0$ (FREEZE ALL)	4.65 ± 0.28	7.20 ± 0.45	0.72 ± 0.03

E Sensitivity Analysis of Fisher Accumulation Factor ω

Complementary to the freezing rate, the accumulation factor ω in the BCRP module governs the temporal evolution of parameter importance. It determines the rate at which the global Fisher memory $\mathcal{F}_{acc}$ is updated with the current task's Fisher information $\mathcal{F}_t$, following the rule $\mathcal{F}_{acc} \leftarrow \omega\mathcal{F}_t + (1-\omega)\mathcal{F}_{acc}$. We evaluated the impact of ω on the model's lifelong learning capability, with results summarized in Tab. 8.

The performance trend reveals the critical importance of balancing historical priors with new knowledge. At the lower extreme ($\omega = 0.0$), the mechanism degrades to a static constraint (equivalent to the "Fixed Fisher" baseline), where the importance estimates remain frozen at their initial values. This prevents the model from registering new critical parameters found in subsequent domains, resulting in suboptimal adaptation (MAE 3.68 bpm). Conversely, at the higher extreme ($\omega = 1.0$), the system exhibits a "memoryless" behavior, aggressively overwriting historical importance with the current task's estimation. This leads to the loss of protection for previously learned features, causing a sharp performance drop due to catastrophic forgetting (MAE 3.95 bpm). The optimal performance is observed at $\omega = 0.5$, where the model maintains an equal trade-off, smoothly integrating new physiological evidence while preserving the structural integrity of historical representations.

Table 8: Sensitivity analysis of the Accumulation Factor ω. This hyperparameter controls the update momentum of the global Fisher Information Matrix. $\omega = 0.0$ corresponds to the Fixed Fisher baseline (never updating importance), while $\omega = 1.0$ implies discarding historical importance constraints.

ACCUMULATION FACTOR (ω)	MAE ↓	RMSE ↓	R ↑
$\omega = 0.0$ (FIXED FISHER)	3.68 ± 0.19	5.60 ± 0.25	0.81 ± 0.02
$\omega = 0.2$	3.35 ± 0.16	5.22 ± 0.22	0.82 ± 0.01
$\omega = 0.5$ (OURS)	$\mathbf{3.20 \pm 0.13}$	$\mathbf{5.06 \pm 0.20}$	$\mathbf{0.83 \pm 0.01}$
$\omega = 0.8$	3.42 ± 0.18	5.38 ± 0.24	0.81 ± 0.02
$\omega = 1.0$ (MEMORYLESS)	3.95 ± 0.25	6.15 ± 0.38	0.78 ± 0.03

F Sensitivity Analysis of Expansion Threshold T

To validate the robustness of the Hierarchical Expansion Strategy (HES), we investigated the impact of the expansion threshold T on the adaptive capacity of ApexPhys. The threshold T determines whether a new adapter should be instantiated according to the perceived distribution shift. As shown in Tab. 9, the model exhibits a clear trade-off with respect to T. A large threshold ($T = 0.5$) makes the expansion criterion overly conservative, preventing the model from allocating sufficient domain-specific capacity under environmental shifts and leading to degraded performance. In contrast, an excessively small threshold ($T \leq 0.01$) triggers more frequent expansion, which may introduce redundant adapters and fragment domain knowledge. The best performance is achieved at $T = 0.05$, where HES maintains an effective balance between compactness and adaptability.

Table 9: Sensitivity analysis of the Expansion Threshold T in the HES module. This threshold controls when a new adapter is instantiated according to the perceived distribution shift.

EXPANSION THRESHOLD (T)	MAE ↓	RMSE ↓	R ↑
$T = 0.5$	3.92 ± 0.22	6.18 ± 0.35	0.78 ± 0.02
$T = 0.1$	3.45 ± 0.16	5.40 ± 0.28	0.81 ± 0.02
$\boldsymbol{T = 0.05}$ **(OURS)**	$\mathbf{3.20 \pm 0.13}$	$\mathbf{5.06 \pm 0.20}$	$\mathbf{0.83 \pm 0.01}$
$T = 0.01$	3.24 ± 0.14	5.10 ± 0.21	0.82 ± 0.01
$T = 0.001$	3.28 ± 0.14	5.15 ± 0.22	0.82 ± 0.01

G Sensitivity Analysis of Initial Curriculum Ratio q_0

To evaluate the robustness of the data annealing strategy in PGP, we further analyze the effect of the initial curriculum ratio q_0. This ratio controls the proportion of training samples used at the beginning of curriculum learning. As shown in Tab. 10, ApexPhys is not sensitive to the choice of q_0. Across a wide range of values, the performance remains stable, indicating that the proposed data annealing strategy does not rely on fine-grained tuning. The best result is obtained at $q_0 = 0.3$, which provides a suitable starting point for gradually introducing harder samples. When $q_0 = 1.0$, curriculum learning is effectively disabled and the model uses all samples from the beginning, leading to a slight performance drop. This suggests that gradually increasing the training samples helps stabilize the early optimization process, while the overall method remains robust to different initial ratios.

Table 10: Sensitivity analysis of the Initial Curriculum Ratio q_0 in the PGP module.

INITIAL CURRICULUM RATIO (q_0)	MAE ↓	RMSE ↓	R ↑
$q_0 = 0.1$	3.36 ± 0.17	5.28 ± 0.24	0.82 ± 0.02
$\boldsymbol{q_0 = 0.3}$ **(OURS)**	$\mathbf{3.20 \pm 0.13}$	$\mathbf{5.06 \pm 0.20}$	$\mathbf{0.83 \pm 0.01}$
$q_0 = 0.6$	3.24 ± 0.15	5.11 ± 0.23	0.82 ± 0.01
$q_0 = 0.9$	3.28 ± 0.16	5.18 ± 0.24	0.82 ± 0.01
$q_0 = 1.0$	3.30 ± 0.18	5.22 ± 0.26	0.82 ± 0.02

H Sensitivity Analysis of Stability Decay Rate κ

To assess the robustness of the constraint annealing strategy in PGP, we analyze the effect of the stability decay rate κ. This parameter controls how fast the path-stability constraint is relaxed during training. As shown in Tab. 11, ApexPhys maintains stable performance across different values of κ, indicating that the proposed constraint annealing strategy is not sensitive to fine-grained tuning. The best performance is achieved at $\kappa = 1.0$, where the inherited prior provides sufficient guidance in the early stage while being gradually relaxed for new-domain adaptation. When κ is too small, the prior constraint decays slowly and may restrict adaptation to the new domain. Conversely, when κ is too large, the constraint is released too quickly, reducing the stabilizing effect of the inherited prior. Overall, the results demonstrate that ApexPhys is robust to a broad range of stability decay rates.

Table 11: Sensitivity analysis of the Stability Decay Rate κ in the PGP module.

Stability Decay Rate (κ)	MAE ↓	RMSE ↓	R ↑
$\kappa = 0.1$	3.62 ± 0.22	5.55 ± 0.30	0.80 ± 0.02
$\kappa = 0.5$	3.35 ± 0.16	5.25 ± 0.24	0.82 ± 0.01
$\kappa = 1.0$ (Ours)	$\mathbf{3.20 \pm 0.13}$	$\mathbf{5.06 \pm 0.20}$	$\mathbf{0.83 \pm 0.01}$
$\kappa = 1.5$	3.38 ± 0.17	5.30 ± 0.23	0.81 ± 0.02
$\kappa = 2.0$	3.65 ± 0.23	5.68 ± 0.31	0.80 ± 0.02

I Generalization on Unseen Domains (Zero-Shot Evaluation)

To further validate the robustness of ApexPhys and its ability to anchor domain-agnostic physiological invariance, we conduct zero-shot evaluations on unseen domains. Unlike the Domain Incremental Learning (DIL) protocol in the main text where models are eventually evaluated on all seen domains, this setting assesses the models on domains they have never encountered during the entire incremental training phase.

Following rigorous cross-dataset evaluation protocols, we establish two distinct zero-shot scenarios. We compare ApexPhys against a representative continual learning regularization method (EWC) and the state-of-the-art rPPG CL method (ADDP). For benchmarking purposes, we also provide the theoretical performance upper bound using a **Fully Supervised** baseline—a model trained and tested exclusively on the target subset. Note that this fully supervised baseline naturally achieves a lower error bound by overfitting the localized environmental patterns of its specific subset, whereas the zero-shot models must rely solely on generalized representations.

Table 12: Zero-Shot Generalization under the **5-vs-1 Protocol**. The models are incrementally trained on five domains (*VIPL Folds 1-4* → *MMPD* → *BUAA* → *PURE* → *UBFC*) and evaluated on the held-out *VIPL Fold 5*. ApexPhys demonstrates minimal performance degradation, approaching the Fully Supervised limit.

Test Set	Method	MAE ↓	RMSE ↓	R ↑
VIPL Fold 5	EWC (Zero-Shot)	4.65	6.50	0.55
	ADDP (Zero-Shot)	3.85	5.25	0.61
	ApexPhys (Zero-Shot)	**3.33**	**4.60**	**0.66**
	Fully Supervised	2.65	3.67	0.70

Table 13: Zero-Shot Generalization under the **4-vs-2 Protocol**. The models are incrementally trained on four domains (*VIPL Folds 1-4* → *MMPD* → *BUAA* → *PURE*) and evaluated on both the held-out *UBFC* and *VIPL Fold 5* datasets.

Test Set	Method	MAE ↓	RMSE ↓	R ↑
UBFC	EWC (Zero-Shot)	0.55	0.85	0.92
	ADDP (Zero-Shot)	0.35	0.50	0.96
	ApexPhys (Zero-Shot)	**0.13**	**0.25**	**0.99**
	Fully Supervised	0.08	0.17	0.99
VIPL Fold 5	EWC (Zero-Shot)	4.75	6.60	0.53
	ADDP (Zero-Shot)	3.95	5.40	0.58
	ApexPhys (Zero-Shot)	**3.40**	**4.85**	**0.62**
	Fully Supervised	2.65	3.67	0.70

As shown in Tab. 12 and Tab. 13, ApexPhys significantly outperforms all baseline methods in unseen environments. By explicitly decoupling physiological invariance from environmental noise, our framework maintains robust zero-shot inference under severe distribution shifts, whereas existing methods suffer notable degradation.

J Analysis of Parameter Expansion and Sub-linear Growth

A common concern in continual learning is the critical trade-off between model performance and the risk of unbounded parameter explosion as the model encounters new tasks. To demonstrate the structural efficiency of ApexPhys, we analyze its cumulative parameter expansion trajectory across a continuous six-stage domain incremental training sequence (*VIPL Folds 1-4* $\rightarrow$ *VIPL Fold 5* $\rightarrow$ *BUAA* $\rightarrow$ *MMPD* $\rightarrow$ *PURE* $\rightarrow$ *UBFC*).

In this context, existing methods exhibit distinct architectural trade-offs. Regularization-based methods, such as EWC, maintain zero parameter growth (0MB) by strictly regularizing existing weights. However, without allocating new capacity to capture novel domain distributions, EWC suffers from severe performance degradation, yielding a high MAE of 4.38. Conversely, prototype-based methods like ADDP achieve better performance but suffer from strict linear parameter growth. By blindly allocating a new, fixed-size prototype pool for every incoming domain regardless of environmental similarity, ADDP adds 1.32MB per stage, eventually inflating the memory overhead to 7.92MB.

In contrast, ApexPhys achieves the optimal balance between performance and capacity through the Hierarchical Expansion Strategy (HES). Rather than growing linearly, ApexPhys exhibits a highly efficient sub-linear trajectory by actively measuring layer-wise distribution shifts and expanding adapters only when absolutely necessary. To illustrate this dynamic routing mechanism, Tab. 14 records the active adapter count (n_1, n_2, n_3) across the last three layers of the backbone at each stage. During the initial deployment on VIPL Folds 1-4, the model instantiates the first set of adapters, denoted as (1, 1, 1), with an initial overhead of 0.66MB. When adapting to the second domain (VIPL Fold 5), the Environmental Distribution Perception (EDP) module detects a significant distribution shift only in the deepest semantic layer. Consequently, HES triggers an expansion exclusively in the third layer, updating the architecture to (1, 1, 2) and slightly increasing the overhead to 0.87MB.

Subsequent adaptations to BUAA and MMPD trigger further localized expansions based on their specific distribution shifts, updating the adapter configuration to (1, 2, 3) and eventually (2, 3, 4). Crucially, when the model encounters the final two domains (PURE and UBFC), it recognizes that their noise characteristics largely overlap with historical distributions. Therefore, the HES bypasses expansion entirely and triggers the reuse mechanism. The architectural footprint plateaus at (2, 3, 4) with a total memory overhead of just 1.97MB. This intelligent, on-demand structural allocation enables ApexPhys to deliver the best measurement accuracy (lowest MAE of 3.20) while remaining highly compact compared to linear-growth alternatives.

Table 14: Parameter Expansion Analysis across Continual Learning Stages. The values represent the cumulative additional parameters introduced beyond the backbone. For ApexPhys, the tuple (n_1, n_2, n_3) indicates the number of active adapters in the last three layers of the backbone at each corresponding stage.

METHOD	VIPL1-4	VIPL5	BUAA	MMPD	PURE	UBFC	MAE ↓
EWC	0MB	0MB	0MB	0MB	0MB	0MB	4.38
ADDP (V)	1.32MB	2.64MB	3.96MB	5.28MB	6.60MB	7.92MB	3.64
APEXPHYS (OURS)	**0.66MB** (1, 1, 1)	**0.87MB** (1, 1, 2)	**1.31MB** (1, 2, 3)	**1.97MB** (2, 3, 4)	**1.97MB** (2, 3, 4)	**1.97MB** (2, 3, 4)	**3.20**

K Licenses and Terms of Use for Existing Assets

In this work, we use several public rPPG datasets whose official access conditions, agreements, or usage terms are summarized below.

VIPL-HR. The official dataset page provides an End User License Agreement (EULA) for dataset download and use. We use the dataset under the official agreement provided by the dataset owners.

PURE. The official dataset page provides access through the maintainers' data page and requests citation of the original work. We use the dataset following the official access route and citation requirement.

UBFC-rPPG. The official dataset page states that the videos are shared for research purposes. We use the dataset according to the conditions described on the official dataset page.

BUAA-MIHR. The official repository provides a dataset release agreement and states that the dataset is for research purposes only. We use the dataset under the official release agreement provided by the dataset authors.

MMPD. The official repository provides a data usage protocol and a release agreement for dataset access. We use the dataset according to the official access protocol and release agreement.

L Broader Impact Discussion

Non-contact physiological measurement may support more convenient health monitoring in settings where contact sensors are impractical, such as long-term observation, low-burden screening, and mobile measurement scenarios. In this context, ApexPhys provides an additional practical advantage by improving robustness under sequentially changing domains, which can reduce the need for repeated manual recalibration when capture environments evolve over time. This property is particularly relevant for real-world rPPG deployment, where illumination, motion, background conditions, and acquisition devices often vary continuously rather than remaining fixed.

At the same time, the use of facial videos and physiological signals may raise privacy concerns if data are collected or used without appropriate consent. Compared with replay-based continual learning strategies, our method offers a more privacy-aware design because it does not require storing raw historical facial videos for rehearsal, thereby reducing the exposure of sensitive visual and physiological information during continual adaptation. Nevertheless, this advantage does not eliminate the need for careful data governance in practice, and any deployment should still consider informed consent, secure data handling, and restrictions against unauthorized monitoring.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The abstract and introduction clearly state the core contribution of ApexPhys, including BCRP, HES, and PGP, and these claims are further detailed in the methodology and validated by the experimental results. The paper's scope is limited to continual rPPG measurement under domain shifts, which is consistent with the problem setting and conclusions presented throughout the paper.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper now explicitly discusses a limitation in the conclusion section. In particular, it clarifies that the current evaluation follows the present continual learning protocol and is conducted on five public rPPG datasets, while further validation on more diverse real-world domains is needed to strengthen the generality of the conclusions.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [N/A]

   Justification: The paper is primarily a methodological and empirical study for continual rPPG measurement. Although it presents mathematical formulations for BCRP, HES, and PGP, it does not state theorem-style theoretical results that would require formal assumptions and proofs.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper provides the continual learning protocol, the dataset usage and task construction, evaluation metrics, module definitions, optimization settings, and an explicit training algorithm in the appendix. Together, the Methodology, Experiments, and Appendix sections provide the main information required to reproduce and verify the core empirical results, while the submitted code package supplies the remaining implementation details.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.

(c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).

(d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: The source code for ApexPhys is publicly available at https://github.com/HeZhongTian-xjtu/ApexPhys. The repository provides the implementation of the proposed method and the resources needed to reproduce the reported experiments.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
   - Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Section "Datasets and Evaluation Protocols" specifies the five datasets, the VIPL fold usage, the incremental training order, evaluation metrics, optimizer, learning rate, weight decay, batch size, input resolution, epoch counts, and key hyperparameters used in the reported experiments. The appendix further supplements these details with the training algorithm, compute resources, and additional ablations.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The main benchmark and ablation tables report results as mean $\pm$ standard deviation, and the paper explicitly states that experiments are repeated three times with different task orders. This gives an appropriate measure of variability for the core empirical claims.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper now reports the experimental compute resources in the appendix, specifically in the "Training Algorithm Detail" section.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: Based on the content of the paper, the work uses public datasets for non-contact physiological measurement research, compares against prior methods, and does not describe procedures that obviously conflict with the NeurIPS Code of Ethics. The manuscript also

avoids storing replay buffers of sensitive facial videos and frames this as a privacy-related motivation for the proposed continual learning setting.

Guidelines:

- The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
- If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
- The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The paper now discusses broader impacts in the appendix, specifically in the "Broader Impact Discussion" section.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: The paper does not release a high-risk generative model, scraped dataset, or similarly misuse-prone asset. It studies continual adaptation for rPPG measurement on existing public datasets, so the safeguards question is not directly applicable in the current form of the manuscript.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.

- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The paper credits the original dataset sources and summarizes the official access conditions, agreements, and stated usage terms for the existing datasets in the appendix, specifically in the "Licenses and Terms of Use for Existing Assets" section.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: The current submission presents a new method, but it does not release a new dataset, benchmark, or documented public software package within the manuscript. Therefore, there are no newly released assets whose accompanying documentation must be evaluated here.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

Justification: The paper does not describe crowdsourcing experiments or newly conducted human-subject studies. All experiments are based on existing public rPPG datasets rather than participant instructions, annotations, or compensation managed by the authors for this work.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: The manuscript does not report a new human-subject study conducted by the authors, nor does it describe participant recruitment or intervention procedures requiring IRB discussion within this paper. The work evaluates on previously released public datasets only.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: LLMs are not part of the core methodology, experimental design, or scientific contribution of ApexPhys. The proposed method is based on continual learning mechanisms for rPPG measurement rather than any important or non-standard use of large language models.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.