PhysLight: Accurate rPPG Heart Rate Measurement with Adaptive Video Relighting

Menglin Zhang[1], Xiaoxin Guo[1,*], Bohao Qu[2,5], Xiaofeng Cao[3], Shuifa Sun[4,*], and Qing Guo[5]

[1]College of Computer Science and Technology, Jilin University, Changchun, China
[2]School of Artificial Intelligence, Jilin University, Changchun, China
[3]School of Computer Science and Technology, Tongji University, Shanghai, China
[4]College of Information Science and Technology, Hangzhou Normal University, Hangzhou, China
[5]IHPC and CFAR, Agency for Science, Technology and Research (A*STAR), Singapore
zhang_menglin@126.com, guoxx@jlu.edu.cn, watersun@hznu.edu.cn

***Abstract*—Facial video-based remote physiological measurement (rPPG) can non-invasively estimate vital signs, such as heart rate (HR), which often faces challenges under varying lighting conditions. We propose the PhysLight framework to enhance the accuracy of rPPG heart rate measurement through adaptive video relighting. Our approach subtly modifies illumination in video frames to improve detection accuracy while maintaining visual quality. The framework includes a GenLightNet to extract ideal lighting priors and a WipeLightNet module to refine poorly lit videos. Extensive evaluations on benchmark datasets show that our method significantly improves HR estimation reliability, outperforming existing baselines and enhancing non-contact physiological monitoring in diverse environments.**

***Index Terms*—remote physiological measurement (rPPG), heart rate measurement, facial video, adaptive relighting**

I. Introduction

Remote photoplethysmography (rPPG) is a well-established non-invasive technique for estimating heart rate (HR) using facial videos [1]–[5]. It introduces a convenient alternative to traditional contact-based devices, such as electrocardiograms, by avoiding direct skin contact, thereby enhancing user comfort and accessibility [6]–[8]. The principle of rPPG HR measurement relies on detecting changes in the intensity of light reflected from the skin, which is influenced by variations in blood volume [9]–[11]. However, rPPG methods are highly sensitive to changes in lighting conditions, which can significantly affect their accuracy and reliability [11]–[13].

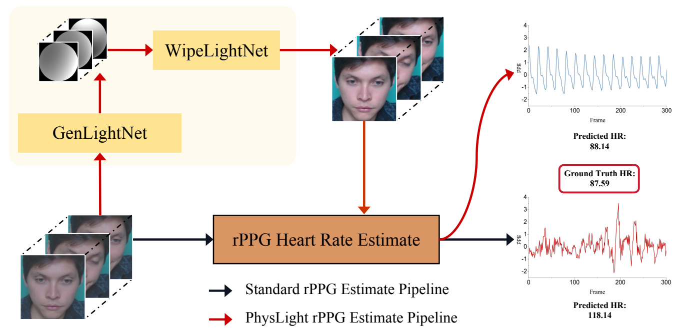

Fig. 1. Intuition of our method. Existing rPPG methods are highly susceptible to variations in lighting conditions, often resulting in inaccurate heart rate estimations. To address this, our approach focuses on regenerating lighting optimized for detection in poorly lit videos, thereby enhancing the accuracy of heart rate measurement.

This sensitivity poses a major challenge to the generalization of rPPG detection in real-world environments.

Existing research has explored various methods to improve the robustness of rPPG detection, including advanced signal processing and machine learning techniques [2]–[5], [14]–[20]. However, these approaches typically focus on improving the detection model itself, which introduces significant complexity and often lacks generalizability across different lighting conditions. To date, there has been little exploration into leveraging input modifications, such as adjusting the lighting in the video, to improve rPPG detection performance. This presents a novel opportunity for enhancing rPPG without altering the core detection algorithms.

Based on these observations, we identify several key challenges for improving rPPG detection. Firstly, rPPG systems are highly sensitive to changes in lighting, which can significantly impact their accuracy. Addressing this challenge requires a robust approach to enhance the generalization capability of rPPG under varying illumination conditions. Secondly, directly improving rPPG detection models is often complex and lacks a straightforward approach to addressing the impact of lighting. There has been no significant attempt to leverage lighting adjustment as a means to enhance rPPG detection performance,

* are corresponding authors.

This research is supported by the National Research Foundation, Singapore and Infocomm Media Development Authority under its Trust Tech Funding Initiative, the National Research Foundation, Singapore and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-GC-2023-008), and Career Development Fund (CDF) of Agency for Science, Technology and Research (A*STAR) (No. C233312028). It is also supported by the National Science Foundation of China (No. 62476192), Natural Science Foundation of Tianjin (No. 23JCQNJC02010). Any opinions, findings and conclusions or recommendations expressed in this material are those of the authors and do not reflect the views of the National Research Foundation, Singapore, Infocomm Media Development Authority. This research also supported by the National Natural Science Foundation of China under Grants 61871258 and 82071995, as well as the Key Research and Development Program of Jilin Province under Grant 20220201141GX.

presenting a gap in the current research landscape. Lastly, existing enhancement strategies are often limited to specific rPPG methods and lack general applicability. A universal approach that can improve the accuracy of different rPPG detection algorithms is still missing.

To tackle these challenges, we propose a novel framework, **PhysLight**, which integrates GenLightNet for extracting ideal lighting priors and WipeLightNet for refining poorly lit videos, thereby enhancing the accuracy of rPPG HR measurement through video relighting. Our contributions are as follows:

- **PhysLight Framework for Enhanced Generalization**. We present PhysLight, a novel framework for adaptive video relighting aimed at enhancing the generalization capability of rPPG detection under varying lighting conditions. By adjusting the illumination in video frames, PhysLight effectively mitigates the impact of lighting variations.
- **Innovative Use of Lighting Adjustment to Improve rPPG Detection**. PhysLight is the first framework to leverage adaptive illumination adjustment to enhance rPPG detection performance. By focusing on modifying the input video rather than altering the detection model itself, our approach reduces the complexity associated with model modifications and provides a new perspective for enhancing rPPG.
- **Generalizability Across Multiple rPPG Methods**. We demonstrate that PhysLight significantly improves the accuracy of various rPPG detection methods, proving its adaptability and generalizability. Our framework provides a universal enhancement strategy that can be applied to different rPPG algorithms, improving their performance in diverse lighting environments.

Extensive evaluations on benchmark datasets demonstrate the effectiveness of our method in enhancing the accuracy of rPPG HR detection. Our results show that PhysLight significantly outperforms existing baselines, providing a practical solution for improving the reliability of non-contact physiological measurement systems in real-world settings.

The rest of this paper is organized as follows. In Section II, we introduce the preliminaries of this work and our motivation. The proposed methodology is detailed in Section III, and experimental results are presented in Section IV. Finally, conclusions and future directions are discussed in Section V.

II. Preliminaries and Motivation

rPPG heart rate (HR) estimates. Given a facial video sequence with N frames, we denote the frames as $\mathbf{V} = \{\mathbf{I}_t\}_{t=1}^{N}$, where $\mathbf{I}_t$ represents the frame at timestamp t. The goal of rPPG is to estimate the heart rate by analyzing these frames, which can be formulated as:

$$E_{\text{HR}} = f(\{\mathbf{I}_t\}_{t=1}^{N}), \tag{1}$$

where $f(\cdot)$ represents the rPPG detection algorithm that processes the video frames to estimate the HR.

The accuracy of rPPG detection is highly sensitive to the quality of the input video, which is influenced by various external factors. Among these factors, lighting variations play a crucial role. Changes in lighting, such as intensity fluctuations or different illumination directions, can introduce noise and artifacts, leading to significant degradation in rPPG signal quality. Such variations pose significant challenges for $f(\cdot)$ in generating reliable HR estimates, particularly in uncontrolled real-world environments.

Existing methods have attempted to address this challenge by making rPPG models more robust to noise through advanced signal processing techniques or by incorporating complex machine learning models. However, these approaches often come with increased computational complexity and are not easily generalizable to different lighting conditions. Moreover, modifying the rPPG detection model itself typically requires extensive retraining, which limits the adaptability of these methods across diverse settings.

Instead of focusing on modifying the detection algorithm $f(\cdot)$, our motivation is to explore a different perspective: improving the quality of the input video by adapting the illumination conditions. By formulating adaptive illumination adjustment, the goal is to create an enhanced version of the input frames, denoted as $\{\hat{\mathbf{I}}_t\}_{t=1}^{N}$, such that the resulting heart rate estimation:

$$E_{\hat{\text{HR}}} = f(\{\hat{\mathbf{I}}_t\}_{t=1}^{N}), \tag{2}$$

is more accurate and robust against varying lighting conditions. This approach not only aims to mitigate the impact of adverse lighting but also maintains the simplicity of the original detection model without adding additional computational burden. The motivation behind this work is to provide a straightforward yet effective solution to enhance rPPG performance by optimizing the input, rather than complicating the detection process itself.

By addressing the problem from the input enhancement perspective, we aim to improve the reliability and generalizability of rPPG HR estimation in diverse real-world environments. This motivates the development of our proposed method, which focuses on adaptive video relighting to standardize illumination and enhance the effectiveness of existing rPPG detection algorithms.

III. Methodology

In this study, we propose PhysLight, a framework that leverages adaptive video relighting to enhance rPPG heart rate detection. The core idea is rooted in the observation that adjusting video lighting conditions can significantly improve the accuracy of physiological measurements while maintaining the semantic integrity of the original video. By employing few-shot learning, PhysLight extracts ideal lighting priors from high-quality data and applies them to refine poorly lit videos.

A. Problem Formulation

Given a video sequence $\mathbf{V} = \{\mathbf{I}_t\}_{t=1}^{N}$ with N frames, where each frame $\mathbf{I}_t$ may have inconsistent or poor lighting conditions, the goal is to generate a new video sequence $\hat{\mathbf{V}} = \{\hat{\mathbf{I}}_t\}_{t=1}^{N}$, such that:

$$\hat{\mathbf{I}}_t = g(\mathbf{I}_t, \mathbf{L}_{\text{gen}}), \tag{3}$$

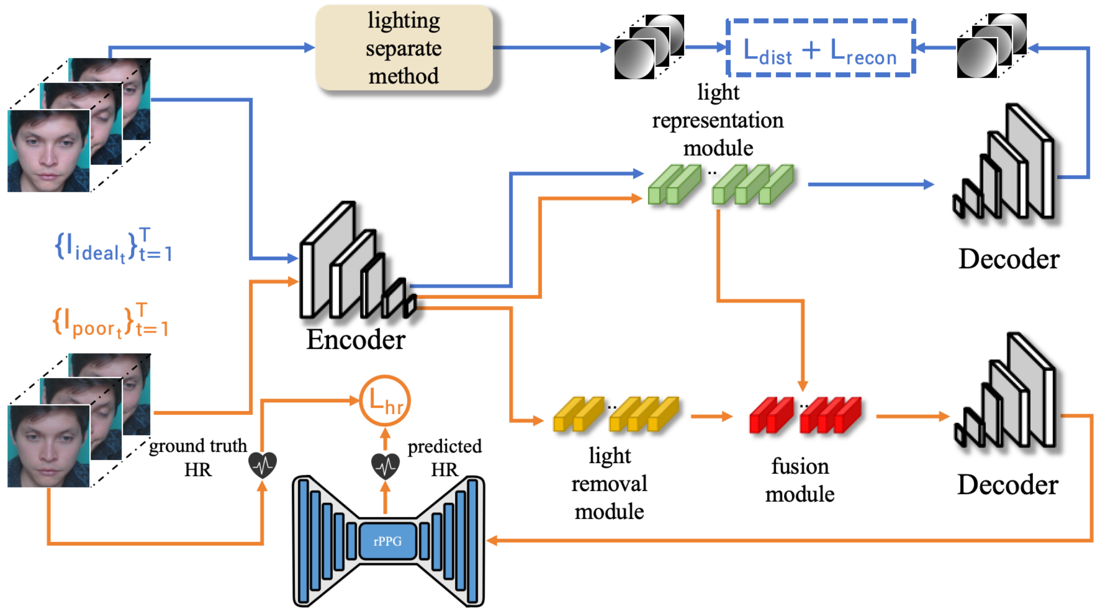

Fig. 2. Pipeline of PhysLight. GenLightNet (blue line) generates ideal lighting priors from well-lit videos and is trained during the first stage using reconstruction loss and distribution alignment loss. WipeLightNet (orange line) utilizes these priors to refine poorly lit videos and is trained in the second stage. The relit videos are subsequently processed by rPPG methods for heart rate estimation, with WipeLightNet training guided by the difference between the predicted and ground truth heart rates.

where $\mathbf{L}_{\text{gen}}$ represents the ideal lighting prior generated by GenLightNet, and $g(\cdot)$ is the transformation applied by WipeLightNet. The resulting video $\hat{\mathbf{V}}$ is optimized for accurate rPPG HR detection while preserving semantic consistency.

B. Few-shot Lighting Prior Learning with GenLightNet

GenLightNet is tasked with learning ideal lighting priors from a small set of high-quality video frames. These priors capture the statistical characteristics of well-lit conditions and are subsequently used to guide the relighting process.

We select input videos where heart rates can be accurately measured using basic rPPG methods as the ground truth, as the lighting in these videos is considered ideal. Given a small set of T ideal lit video frames $\{\mathbf{I}_{\text{ideal}_t}\}_{t=1}^{T}$. We use the method introduced in [21] to separate the ideal lighting from the frames, obtaining a set of ideal lighting $\{\mathbf{L}_{\text{ideal}_t}\}_{t=1}^{T}$.

We design a GenLightNet to learn ideal lighting priors $\mathbf{L}_{\text{gen}_t}$,

$$\hat{\mathbf{L}}_{gen_t} = \text{GenLightNet}(\mathbf{I}_{\text{ideal}_t}), \mathbf{I}_{\text{ideal}_t} \in \{\mathbf{I}_{\text{ideal}_t}\}_{t=1}^{T}. \quad (4)$$

The GenLightNet consists of three key components. First, an encoder, implemented as a convolutional neural network (CNN), extracts content features $\mathbf{F}_{\text{content}_t}$ from $\mathbf{I}_{\text{ideal}_t}$. Second, a light representation module inputs $\mathbf{F}_{\text{content}_t}$ and output a lighting representation $\mathbf{L}_{\text{gen}_t}$. Finally, the decoder reconstructs the ideal lighting map $\hat{\mathbf{L}}_{\text{gen}_t}$ from $\mathbf{L}_{\text{gen}_t}$, ensuring that the extracted features preserve the essential characteristics of ideal lighting.

GenLightNet is trained using the following two-part objectives: reconstruction loss and distribution alignment loss. The reconstruction loss,

$$\mathcal{L}_{\text{recon}} = \sum_{t=1}^{T} ||\hat{\mathbf{L}}_{\text{gen}_t} - \mathbf{L}_{\text{ideal}_t}||_1, \quad (5)$$

ensures that each single-frame lighting map $\hat{L}_{\text{gen}}$ generated by GenLightNet accurately reproduces the ideal lighting map L_{ideal}. It emphasizes pixel-wise consistency to guarantee the precise reconstruction of high-quality video frames.

The distribution alignment loss optimizes GenLightNet from a global perspective, enhancing the model's generalization ability under unseen lighting conditions. Specifically, we use the Kernel Density Estimation (KDE) [22], [23] method to model the ideal lighting and generated lighting, resulting in the Ideal Lighting Distribution $\mathrm{P}_{L_{\text{ideal}}} = \text{KDE}(\{\mathbf{L}_{\text{ideal}_t}\}_{t=1}^{T})$ and Generated Lighting Distribution $\mathrm{P}_{\hat{L}_{\text{gen}}} = \text{KDE}(\{\hat{\mathbf{L}}_{gen_t}\}_{t=1}^{T})$, respectively. To ensure consistency between the generated lighting distribution and the ideal lighting distribution, we introduce the KL divergence as the loss function,

$$\mathcal{L}_{\text{dist}} = \mathrm{D}_{\text{KL}}(\mathrm{P}_{\hat{L}_{\text{gen}}} || \mathrm{P}_{L_{\text{ideal}}}), \quad (6)$$

where D_{KL} is the KL divergence between the distributions of generated and ideal lighting maps, the KL divergence serves to minimize the difference between the two distributions.

C. Lighting Adjustment with WipeLightNet

WipeLightNet adjusts poorly lit video frames using the ideal lighting priors generated by GenLightNet. It removes undesirable lighting effects and seamlessly integrates the ideal lighting priors into the input frames.

Given a poorly lit video frame $\mathbf{I}_{\text{poor}_t}$ and the ideal lighting representation $\mathbf{L}_{\text{gen}_t}$ generated by GenLightNet. We design the WipeLightNet to obtain the relit frame $\hat{\mathbf{I}}_t$,

$$\hat{\mathbf{I}}_t = \text{WipeLightNet}(\mathbf{I}_{\text{poor}_t}, \mathbf{L}_{\text{gen}_t}). \quad (7)$$

WipeLightNet is composed of several key components. An encoder, the same one as the GenLightNet, extracts content features $\mathbf{F}_{\text{content}_t}$ from $\mathbf{I}_{\text{poor}_t}$. A light removal module, consisting of residual blocks, refines $\mathbf{F}_{\text{content}_t}$ by eliminating the effects of poor lighting. The fusion module utilizes an adaptive instance normalization (AdaIN) layer to integrate $\mathbf{L}_{\text{gen}_t}$ with $\mathbf{F}_{\text{content}_t}$, resulting in the fused features $\mathbf{F}_{\text{fused}_t}$. Finally, a decoder reconstructs the final relit frame $\hat{\mathbf{I}}_t$ from $\mathbf{F}_{\text{fused}_t}$.

The training of WipeLightNet is guided by:

$$\begin{aligned} \mathcal{L}_{\text{hr}} &= ||f(\{\mathbf{I}_t\}_{t=1}^{N}) - \text{HR}_{\text{gt}}||_1, \\ \text{subject to} &||\hat{\mathbf{I}}_t - \hat{\mathbf{I}}_{t+1}||_1 \leq \epsilon \end{aligned} \tag{8}$$

where HR_{gt} is the ground truth of the heart rate of the video, and Parameter ϵ controls the changing degrees of lighting ensuring smooth transitions between consecutive frames.

D. Overall Training Pipeline

We used the Adam optimizer with a learning rate of 1×10^{-4} All experiments were conducted on an NVIDIA RTX 3090 GPU. The training process consists of two stages:

1) **GenLightNet Training:** Few-shot learning on high-quality video frames to extract ideal lighting priors.
2) **WipeLightNet Training:** Refinement of poorly lit frames using the priors from GenLightNet. GenLightNet is frozen during this stage to ensure stable priors.

The Light Representation Module maps content features to a latent lighting representation with a Fully Connected Layers Architecture. WipeLightNet Encoder structure is identical to GenLightNet's encoder, and Light Removal Module include 3 ResBlocks, each block is Conv2D$(256 \rightarrow 256, kernel = 3)$ + BN + ReLU. Fusion Module integrates lighting priors with content features via Adaptive Instance Normalization.

IV. Experiments Results

A. Setups

Evaluation datasets. We run experiments on three datasets: UBFC-rPPG [24], PURE [25] and RLAP-rPPG [26]. RLAP-rPPG consists of MJPG-compressed videos from 58 participants recorded with a Logitech C930c webcam, while UBFC-rPPG includes uncompressed RGB videos of 43 participants captured with a Logitech C920 HD Pro webcam. PURE contains 59 videos of 10 subjects recorded using a robotic camera in lossless RGB format. All datasets provide ground truth physiological signals measured by the Contec CMS50E pulse meter, offering diverse settings for robust performance evaluation. Notably, all three datasets feature videos with a resolution of 640 × 480 pixels at 30 frames per second (FPS), ensuring consistency in spatial and temporal settings across evaluations. This uniform resolution and frame rate facilitate a fair comparison of rPPG models under varying lighting and motion conditions. Furthermore, the datasets cover a wide range of skin tones, genders, and head movements, providing comprehensive benchmarks for evaluating the effectiveness and generalizability of our proposed PhysLight framework.

rPPG methods. We evaluate the proposed PhysLight framework across a variety of rPPG models, including PhysNet [2], MTTS [3], RemotePPG [4], EfficientPhys [14], Physformer [15], and BigSmall [27]. To ensure fair comparison and optimal performance, all models undergo consistent face image pre-processing using MTCNN [28], which is applied uniformly throughout the evaluation. Our experiments focus on improving input quality through adaptive video relighting, rather than employing any adversarial attacks, thus enhancing the performance of different rPPG detection methods in diverse scenarios.

Evaluation protocol. Previous works calculate the HR from estimated rPPG signals and compare them to the corresponding ground truth for performance evaluation. We follow [4], [16], [29] to use mean absolute error (MAE), root mean square error (RMSE) and Pearson's correlation coefficient (ρ) as evaluation metrics for HR.

B. Comparison Results

We evaluate the PhysLight framework against baseline rPPG methods that do not utilize adaptive relighting, testing their performance on three benchmark datasets: UBFC-rPPG, PURE, and RLAP-rPPG. The results, as shown in Table I, demonstrate the following key findings: ❶ PhysLight consistently improves rPPG accuracy across all datasets and models. For example, PhysLight significantly reduces MAE and RMSE while increasing the ρ, demonstrating enhanced reliability and accuracy of heart rate predictions. ❷ The improvements are observed across diverse rPPG models, highlighting PhysLight's generalizability. Unlike conventional approaches, PhysLight maintains robust performance under varying lighting conditions, reducing errors without modifying the rPPG models themselves. ❸ The results also confirm that PhysLight enhances not only detection accuracy but also the reliability of non-contact physiological measurement across multiple challenging datasets, demonstrating its effectiveness as a universal input enhancement strategy. These findings demonstrate the effectiveness and advantages of PhysLight over traditional baseline methods, particularly in its ability to enhance accuracy under challenging lighting conditions while preserving performance on standard video inputs.

Figure 3 demonstrates the effectiveness of PhysLight in enhancing lighting conditions for poorly lit input frames. The figure compares three key components: the original poorly lit input, a reference image with ideal lighting, and the relighted output generated by PhysLight. The original input exhibits uneven lighting and loss of facial details, which are critical for accurate rPPG detection. In contrast, the PhysLight output shows smooth and consistent lighting that closely resembles the reference while maintaining natural facial features. The shade maps further highlight the improvement, where the PhysLight-generated shading achieves better uniformity compared to the uneven and noisy shading in the input. Additionally, the lighting maps illustrate that PhysLight corrects irregular lighting patterns, producing outputs with well-distributed and consistent lighting. These improvements directly enhance the temporal and spatial consistency of lighting, which are essential for reliable rPPG signal extraction, demonstrating the effectiveness of PhysLight in overcoming real-world lighting challenges.

TABLE I
PERFORMANCE OF SIX RPPG HR MEASUREMENT METHODS WITH AND WITHOUT PHYSLIGHT ON THREE DATASETS.

rPPG Model	w/PhysLight	UBFC-rPPG [24]			PURE [25]			RLAP-rPPG [26]		
		MAE↓	RMSE↓	$\rho\uparrow$	MAE↓	RMSE↓	$\rho\uparrow$	MAE↓	RMSE↓	$\rho\uparrow$
PhysNet [2]	✗	6.420	8.857	0.716	8.371	9.240	0.703	15.062	15.062	0.217
	✓	3.241	5.480	0.872	4.294	5.483	0.810	8.412	9.132	0.536
	Relative Change	-49.5%	-38.7%	+21.8%	-48.7%	-40.7%	+15.2%	-44.1%	-39.4%	+147.0%
MTTS [3]	✗	2.520	3.812	0.925	3.680	6.547	0.881	9.431	10.524	0.461
	✓	1.749	2.852	0.942	2.431	4.214	0.906	6.243	7.185	0.639
	Relative Change	-30.6%	-25.2%	+1.8%	-33.9%	-35.6%	+2.8%	-33.8%	-31.7%	+38.6%
RemotePPG [4]	✗	3.620	4.983	0.959	2.348	3.016	0.991	11.284	12.421	0.618
	✓	2.781	4.021	0.972	1.751	2.312	0.997	7.143	8.045	0.762
	Relative Change	-23.2%	-19.3%	+1.3%	-25.4%	-23.3%	+0.6%	-36.7%	-35.3%	+23.3%
EfficentPhys [14]	✗	3.270	4.875	0.957	5.308	9.243	0.870	10.748	11.932	0.552
	✓	2.263	3.821	0.965	3.954	6.184	0.897	6.743	8.213	0.662
	Relative Change	-30.8%	-21.7%	+0.8%	-25.5%	-33.1%	+3.1%	-37.3%	-31.2%	+20.0%
PhysFormer [15]	✗	5.504	10.879	0.623	4.543	10.376	0.694	5.431	8.013	0.896
	✓	3.312	7.254	0.834	2.943	6.712	0.772	3.092	5.390	0.912
	Relative Change	-39.8%	-33.3%	+33.9%	-35.2%	-35.3%	+11.2%	-43.1%	-32.7%	+1.8%
BigSmall [27]	✗	1.048	2.585	0.986	1.984	6.549	0.939	5.231	8.391	0.794
	✓	0.764	1.915	0.994	1.431	4.743	0.962	3.124	5.843	0.852
	Relative Change	-27.1%	-25.9%	+0.8%	-27.9%	-27.6%	+2.4%	-40.3%	-30.4%	+7.3%

TABLE II
ABLATION STUDY OF PHYSLIGHT ON MULTIPLE RPPG MODELS AND DATASETS. RANDOM PRIORS MEANS THAT TESTS THE EFFECT OF USING RANDOM PRIORS INSTEAD OF FEW-SHOT LEARNING.

rPPG Model	Ablation Configs	UBFC-rPPG				PURE				PLAP-rPPG			
		MAE↓	RMSE↓	$\rho\uparrow$	TS↓	MAE↓	RMSE↓	$\rho\uparrow$	TS↓	MAE↓	RMSE↓	$\rho\uparrow$	TS↓
RemotePPG [4]	PhysLight (Ours)	2.781	4.021	0.972	0.035	1.751	2.312	0.997	0.057	7.143	8.045	0.762	0.050
	Ours w/o GenLightNet	16.892	21.403	0.582	1.642	18.213	23.452	0.621	1.432	22.983	27.401	0.541	1.721
	Ours w/o WipeLightNet	19.721	24.103	0.503	1.524	17.912	22.371	0.542	1.389	21.874	26.501	0.573	1.612
	Random Priors	9.201	12.541	0.721	0.402	7.091	9.872	0.794	0.332	14.142	16.371	0.541	0.504
PhysFormer [15]	PhysLight (Ours)	3.312	7.254	0.834	0.039	2.943	6.712	0.772	0.045	3.092	5.390	0.912	0.051
	Ours w/o GenLightNet	18.402	23.521	0.521	1.743	17.812	22.712	0.591	1.592	19.412	24.501	0.502	1.631
	Ours w/o WipeLightNet	20.672	25.874	0.549	1.651	18.924	24.531	0.612	1.482	22.712	27.402	0.531	1.731
	Random Priors	9.572	12.841	0.732	1.152	7.472	10.972	0.694	1.024	15.132	17.491	0.521	1.382

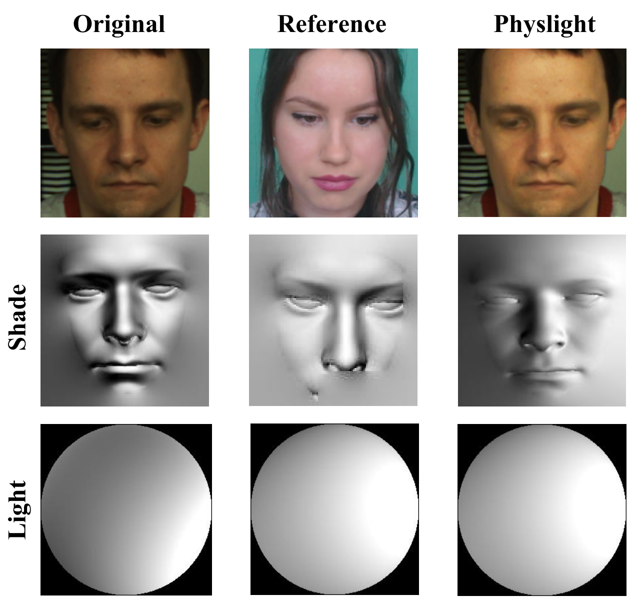

Fig. 3. Visualization of PhysLight's Adaptive Relighting.

C. Ablation Study

We performed ablation studies to evaluate the impact of removing key components of PhysLight, specifically GenLightNet, WipeLightNet, and the use of predictive priors. The results in Table II show significant performance degradation when any module is removed, validating the necessity of each component.

Ablating the GenLightNet module (Ours w/o GenLightNet) Removing GenLightNet causes the most severe performance drop across all datasets and models. For example, in the PURE dataset with RemotePPG, MAE increases from 1.751 to 18.213, and TS rises from 0.134 to 1.432. This demonstrates the importance of accurate lighting priors in enhancing rPPG accuracy and temporal consistency.

Removing the WipeLightNet module (Ours w/o WipeLightNet) Ablating WipeLightNet results in substantial degradation, such as increasing MAE from 3.312 to 20.672 and TS from 0.183 to 1.651 on the UBFC-rPPG dataset with PhysFormer. WipeLightNet plays a key role in refining lighting adjustments and maintaining temporal smoothness.

Using Random Priors instead of Predictive Priors (Random Priors) Using random priors instead of predictive priors leads to moderate performance decline. For instance, in the PLAP-rPPG dataset with PhysFormer, MAE increases from 3.092 to 15.132. This highlights the importance of Few-Shot Learning in enabling adaptive and effective relighting.

The significant performance degradation observed in the ablated versions highlights the necessity of both GenLight-

Net and WipeLightNet. GenLightNet provides the predictive lighting priors that are essential for accurate relighting, while WipeLightNet ensures smooth and consistent integration of these priors into the input frames. Together, these modules allow PhysLight to outperform existing methods and maintain robustness across diverse datasets and lighting conditions.

V. Conclusion

This paper introduced PhysLight, a novel framework for enhancing rPPG heart rate measurement through adaptive video relighting. We utilize few-shot learning to generate ideal lighting priors with GenLightNet, which are then applied by WipeLightNet to refine poorly lit videos. PhysLight improves rPPG accuracy across multiple datasets and models, addressing challenges posed by lighting variations. It also maintains temporal smoothness and visual consistency, ensuring compatibility with existing rPPG pipelines. Extensive experiments confirm its robustness and generalizability. By focusing on input optimization rather than modifying detection algorithms, PhysLight offers a practical plug-and-play solution for accurate and reliable non-contact physiological measurement.

References

[1] Weixuan Chen and Daniel McDuff, "Deepphys: Video-based physiological measurement using convolutional attention networks," in *ECCV*, 2018.

[2] Zitong Yu, Xiaobai Li, and Guoying Zhao, "Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks," in *BMVC*. British Machine Vision Association, 2019.

[3] Xin Liu, Josh Fromm, Shwetak Patel, and Daniel McDuff, "Multi-task temporal shift attention networks for on-device contactless vitals measurement," *NeurIPS*, 2020.

[4] John Gideon and Simon Stent, "The way to my heart is through contrastive learning: Remote photoplethysmography from unlabelled video," in *ICCV*, 2021.

[5] Hao Lu, Hu Han, and S Kevin Zhou, "Dual-gan: Joint bvp and noise modeling for remote physiological measurement," in *CVPR*, 2021.

[6] Chayakrit Krittanawong, Albert J Rogers, Johnson, et al., "Integration of novel monitoring devices with machine learning technology for scalable cardiovascular management," *Nature Reviews Cardiology*, 2021.

[7] Xuesong Niu, Shiguang Shan, Hu Han, and Xilin Chen, "Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation," *TIP*, 2019.

[8] Zijie Yue, Shuai Ding, Shanlin Yang, et al., "Deep super-resolution network for rppg information recovery and noncontact heart rate estimation," *TIM*, 2021.

[9] Philipp V Rouast, Marc TP Adam, Raymond Chiong, et al., "Remote heart rate measurement using low-cost rgb face video: a technical literature review," *Frontiers of Computer Science*, 2018.

[10] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *TBME*, 2010.

[11] Wenjin Wang, Albertus C Den Brinker, Sander Stuijk, and Gerard De Haan, "Algorithmic principles of remote ppg," *TBME*, 2016.

[12] Xiaobai Li, Jie Chen, Guoying Zhao, and Matti Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *CVPR*, 2014.

[13] Gerard De Haan and Vincent Jeanne, "Robust pulse rate from chrominance-based rppg," *TBME*, 2013.

[14] Xin Liu, Brian L Hill, Ziheng Jiang, Shwetak Patel, and Daniel McDuff, "Efficientphys: Enabling simple, fast and accurate camera-based vitals measurement," *arXiv preprint arXiv:2110.04447*, 2021.

[15] Zitong Yu, Yuming Shen, Jingang Shi, Zhao, et al., "Physformer: Facial video-based physiological measurement with temporal difference transformer," in *CVPR*, 2022.

[16] Zijie Yue, Miaojing Shi, and Shuai Ding, "Facial video-based remote physiological measurement via self-supervised learning," *TPAMI*, 2023.

[17] Pradyumna Chari, Anirudh Bindiganavale Harish, et al., "Implicit neural models to extract heart rate from video," in *ECCV*, 2024.

[18] Jingda Du, Si-Qi Liu, Bochao Zhang, et al., "Dual-bridging with adversarial noise generation for domain adaptive rppg estimation," in *CVPR*, 2023.

[19] Jun Seong Lee, Gyutae Hwang, et al., "Lstc-rppg: Long short-term convolutional network for rppg," in *CVPR*, 2023.

[20] Deivid Botina-Monsalve, Yannick Benezeth, and Johel Miteran, "Rtrppg: An ultra light 3dcnn for real-time rppg," in *CVPR*, 2022.

[21] Qian Zhang, Qing Guo, Ruijun Gao, Felix Juefei-Xu, Hongkai Yu, and Wei Feng, "Adversarial relighting against face recognition," *TIFS*, 2024.

[22] Emanuel Parzen, "On estimation of a probability density function and mode," *The Annals of Mathematical Statistics*, 1962.

[23] Yen-Chi Chen, "A tutorial on kernel density estimation and recent advances," *Biostatistics & Epidemiology*, 2017.

[24] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," *PRL*, 2019.

[25] Ronny Stricker, Steffen Müller, and Horst-Michael Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," in *RO-MAN*. IEEE, 2014.

[26] Kegang Wang, Yantao Wei, Jiankai Tang, et al., "Camera-based hrv prediction for remote learning environments," *arXiv preprint arXiv:2305.04161*, 2023.

[27] Girish Narayanswamy, Yujia Liu, Yuzhe Yang, et al., "Bigsmall: Efficient multi-task learning for disparate spatial and temporal physiological measurements," in *WACV*, 2024.

[28] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *SPL*, 2016.

[29] Rencheng Song, Huan Chen, Juan Cheng, Chang Li, Yu Liu, and Xun Chen, "Pulsegan: Learning to generate realistic pulse waveforms in remote photoplethysmography," *JBHI*, 2021.

[30] Antony Lam and Yoshinori Kuno, "Robust heart rate measurement from video using select random patches," in *ICCV*, 2015.

[31] Sergey Tulyakov, Xavier Alameda-Pineda, Elisa Ricci, Lijun Yin, Jeffrey F Cohn, and Nicu Sebe, "Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions," in *CVPR*, 2016.

[32] Hua Qi, Qing Guo, et al., "Deeprhythm: Exposing deepfakes with attentional visual heartbeat rhythms," in *ACM MM*, 2020.

[33] Hao Zhou, Sunil Hadap, Kalyan Sunkavalli, and David Jacobs, "Deep single-image portrait relighting," in *ICCV*, 2019.

[34] Amnon Shashua and Tammy Riklin-Raviv, "Quotient image: Class-based re-rendering and recognition with varying illuminations," *TPAMI*, 2001.

[35] Zhibo Wang, Xin Yu, Ming Lu, Quan Wang, Chen Qian, and Feng Xu, "Single image portrait relighting via explicit multiple reflectance channel modeling," *TOG*, 2020.

[36] Haozhou Sun, Xue Qian, Jing Tang, and Heung-Yeung Shum, "Single image portrait relighting via explicit multiple reflectance modeling," in *ICCV*, 2019.

[37] Xiaoyan Zhang, Yukai Song, Zhuopeng Li, and Jianmin Jiang, "Pr-rl: Portrait relighting via deep reinforcement learning," *TMM*, 2021.

[38] Sai Bi, Zexiang Xu, Kalyan Sunkavalli, et al., "Deep reflectance volumes: Relightable reconstructions from multi-view photometric images," in *ECCV*, 2020.

[39] Yihao Huang, Liangru Sun, Qing Guo, et al., "Ala: Naturalness-aware adversarial lightness attack," in *ACM MM*, 2023.

[40] Binyu Tian, Felix Juefei-Xu, Qing Guo, et al., "Ava: Adversarial vignetting attack against visual recognition," in *IJCAI*, 2021.

[41] Yihao Huang, Felix Juefei-Xu, Qing Guo, et al., "Dodging deepfake detection via implicit spatial-domain notch filtering," *TCSVT*, 2023.

[42] Yue Cao, Tianlin Li, et al., "Irad: implicit representation-driven image resampling against adversarial attacks," in *ICLR*, 2024.

[43] Jianlang Chen, Xuhong Ren, et al., "Lrr: Language-driven resamplable continuous representation against adversarial tracking attacks," in *ICLR*, 2024.

[44] Sensen Gao, Xiaojun Jia, Xuhong Ren, et al., "Boosting transferability in vision-language attacks via diversification along the intersection region of adversarial trajectory," in *ECCV*, 2024.

APPENDIX

In the appendix, we provide additional details to support our main results. Sec. A discusses related work in the field of rPPG methods, offering context for our approach. Sec. B introduces the heart rate evaluation metrics used in our experiments, providing a clear understanding of the performance assessment criteria. These sections aim to enhance the completeness of our study.

A. Related Work

1) rPPG Measurement: Facial video-based remote physiological measurement (rPPG) has advanced significantly, enabling non-invasive extraction of heart rate and other vital signals from facial videos. Historically, rPPG techniques have evolved through two major phases. Initially, conventional methods analyzed subtle color variations in specific facial regions to estimate heart rate, relying on mathematical models such as blind signal separation [10], least mean square [12], majority voting [30], and self-adaptive matrix completion [31]. These approaches were grounded in skin reflection models [11], [13] and focused on signal extraction from predetermined areas of the face. The emergence of deep learning ushered in a transformative phase in rPPG research, leading to the development of more adaptive and powerful models. Contemporary deep learning-based approaches, such as DeepPhys [1], MTTS [3], and BigSmall [27], have significantly enhanced the accuracy of heart rate detection.

Prior studies have also utilized rPPG signals in downstream tasks such as DeepFake detection. For instance, DeepRhythm [32] leverages temporal inconsistencies in visual heartbeat rhythms to distinguish real and fake videos, highlighting the sensitivity of rPPG to subtle input variations. While our goal is to enhance rPPG robustness under natural lighting, their findings further confirm the importance of input quality for reliable rPPG estimation.

2) Input-level Manipulations in Facial Video Analysis: Input-level manipulations, particularly through relighting or adversarial perturbations, have attracted increasing attention in facial video analysis. In the realm of computer vision and graphics, facial video-based relighting has been widely studied to enhance the visual quality and aesthetic appeal of facial content. Techniques such as deep portrait relighting [33], image re-rendering [34], photorealistic portrait generation [35], and encoder-decoder-based CNN networks [36] have shown promising results. Recent approaches also include reinforcement learning [37] and dynamic lighting adjustment methods [38], with applications in video enhancement, editing, and virtual reality.

Beyond aesthetic enhancement, relighting has also been explored as an adversarial tool. Zhang et al. [21], [39] introduce adversarial relighting attacks that generate natural-looking lighting variations to deceive face recognition systems by reducing identity similarity scores. Similarly, several recent studies have proposed other forms of adversarial input perturbations, including lightness-based or vignetting attacks [39], [40], deepfake evasion via visual artifacts [41], and adversarial signal construction using implicit or language-driven representations [42]–[44]. While these works are not specifically developed for physiological signal processing, they share a common emphasis on structured input manipulation to influence model behavior.

Our work, although non-adversarial in nature, applies structured relighting to enhance rPPG robustness under real-world lighting variations. Unlike prior adversarial or aesthetic approaches, we focus on preserving subtle facial color rhythms crucial for heart rate estimation. To our knowledge, little attention has been paid to input-level robustness in rPPG systems under natural conditions, and our method addresses this important but underexplored aspect.

B. Heart Rate Evaluation Metrics

In assessing the accuracy of heart rate estimation in rPPG-based systems subjected to adversarial relighting, three primary metrics are employed: Mean Average Error (MAE), Root Mean Square Error (RMSE), and Pearson Correlation (r). These metrics provide a robust framework for evaluating the precision and reliability of heart rate measurements under different attack methods, thus highlighting the effectiveness and subtlety of adversarial interventions in compromising rPPG signal integrity.

Mean Average Error (MAE). The MAE as defined between the predicted signal rate R_{pred} and the ground truth signal rate R_{gt} for a total of N instances:

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|R_{gt} - R_{pred}|, \tag{9}$$

the ranges from 0 to infinity. The MAE measures the average magnitude of errors in a set of predictions, without considering their direction (*i.e.*, over or underestimations are treated equally). A lower MAE indicates that the post-attack videos are causing minimal deviation in heart rate predictions from the true values, suggesting your adversarial method maintains high fidelity in terms of not disturbing the heart rate measurement significantly. A higher MAE indicates greater error, implying that the adversarial relighting significantly impacts the accuracy of heart rate measurements.

Root Mean Square Error (RMSE). The RMSE as defined between the predicted signal rate R_{pred} and the ground truth signal rate R_{gt} for a total of N instances:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(R_{gt} - R_{pred})^2}, \tag{10}$$

the ranges from 0 to infinity. RMSE gives a measure of the magnitude of errors, with larger errors given more weight due to the squaring of each error term. This metric is sensitive to large errors, which can be crucial if your adversarial relighting significantly distorts heart rate measurement in some frames more than others. A lower RMSE suggests that on average, the perturbations introduced by your adversarial relighting are generally subtle and do not drastically affect heart rate measurement. Conversely, a higher RMSE indicates more pronounced errors introduced by the adversarial relighting.

Pearson Correlation (r). The Pearson correlation as defined between the predicted signal rate R_{pred} and the ground truth signal rate R_{gt} for a total of N instances, and R the mean of R over N instances:

$$r = \frac{\sum_{i=1}^{N}(R_{gt} - \overline{R_{gt}})(R_{pred} - \overline{R_{pred}})}{\sqrt{\sum_{i=1}^{N}(R_{gt} - \overline{R_{gt}})^2 \sum_{i=1}^{N}(R_{pred} - \overline{R_{pred}})}}, \tag{11}$$

the ranges from -1 to 1. This metric assesses the linear relationship between the true and measured heart rates. A correlation of 1 indicates a perfect positive linear relationship, -1 indicates a perfect negative linear relationship, and 0 indicates no linear relationship. In the context of your study, a high positive Pearson correlation (close to 1) suggests that despite the adversarial relighting, the measured heart rates still maintain a strong linear association with the true rates, indicating that the integrity of rPPG measurements is preserved. A low or negative correlation would suggest that the adversarial relighting disrupts the linear predictive ability of rPPG methods significantly.

C. Algorithm

Algorithm 1 PhysLight: Adaptive Video Relighting for rPPG

Input: Original video $V = \{I_t\}_{t=1}^{N}$ with N frames
Output: Relighted video $\hat{V} = \{\hat{I}_t\}_{t=1}^{N}$

```
Stage 1: Learn ideal lighting priors with GenLightNet
for t = 1 to T do
    Extract lighting label L_t^ideal ← LightingExtract(I_t^ideal)
    Compute predicted lighting L̂_t ← GenLightNet(I_t^ideal)
    Compute reconstruction loss L_recon += ||L̂_t − L_t^ideal||_1
end for
Compute KDE-based distributions P_L̂, P_L from {L̂_t}, {L_t^ideal}
Compute distribution loss L_dist ← D_KL(P_L̂ || P_L)
Update GenLightNet with L_recon + L_dist
Stage 2: Refine poorly-lit videos with WipeLightNet
for t = 1 to N do
    Extract content feature F_t ← Encoder(I_t)
    Refine lighting feature F̃_t ← LightRemoval(F_t)
    Fuse lighting prior F_t^fused ← AdaIN(F̃_t, L_t^gen)
    Decode relighted frame Î_t ← Decoder(F_t^fused)
    Append Î_t to V̂
end for
// Training WipeLightNet
Compute HR prediction loss L_HR = ||f(V̂) − HR_gt||_1
Enforce smooth lighting transition ∀t : ||Î_t − Î_{t+1}||_1 ≤ ε
Update WipeLightNet with L_HR and smoothness constraint

return V̂
```