SHINE-PPG: Non-Lambertian Intrinsic Decomposition for Illumination-Robust rPPG

Shih-Yu Yang[1⋆], Yen-Chun Chou[1⋆], Pei-Kai Huang[1,2†], and Chiou-Ting Hsu[1†]

[1] Department of Computer Science, National Tsing Hua University, Hsinchu, Taiwan
{edmond113062560,chouuuu,alwayswithme,cthsu}@gapp.nthu.edu.tw
[2] College of Computer and Cyber Security, Fujian Normal University, Fuzhou, China
alwayswithme@fjnu.edu.cn

Abstract. Remote photoplethysmography (rPPG) enables non-contact physiological monitoring by capturing subtle skin color variations induced by cardiac cycles. Despite its promise, rPPG remains highly sensitive to environmental illumination. Existing illumination-aware methods suffer from two key limitations: (1) limited out-of-distribution (OOD) generalization due to domain discrepancies between training and testing data, and (2) reliance on the Lambertian assumption, which neglects non-Lambertian specular highlights that frequently corrupt facial skin signals. In this paper, we propose SHINE-PPG (Specular-Highlight Intrinsic Network for rPPG Estimation), a novel framework that leverages non-Lambertian intrinsic decomposition to decouple facial videos into illumination, reflectance, and specular components in a self-supervised manner. By isolating physiological information within the intrinsic reflectance, our method effectively suppresses both ambient lighting variations and surface highlights to recover high-fidelity rPPG signals. To further enhance robustness, we introduce an adversarial illumination enhancement strategy that dynamically synthesizes challenging unseen lighting conditions during training, significantly improving OOD generalization. Extensive experiments on five benchmark datasets demonstrate that SHINE-PPG consistently outperforms previous methods, particularly under complex and dynamic illumination scenarios. The code is available at https://github.com/Edmond-Yang/SHINE-PPG.

Keywords: Remote photoplethysmography · Non-Lambertian Intrinsic Image Decomposition · Specular Highlights · Adversarial Learning

1 Introduction

Remote Photoplethysmography (rPPG) [10, 22, 32, 41, 44] is a non-contact physiological sensing technique that estimates cardiovascular signals from facial videos captured by commodity cameras. It operates by detecting subtle intensity variations induced by periodic blood volume changes, which modulate the skin's

⋆ Equal contribution; † Corresponding author.

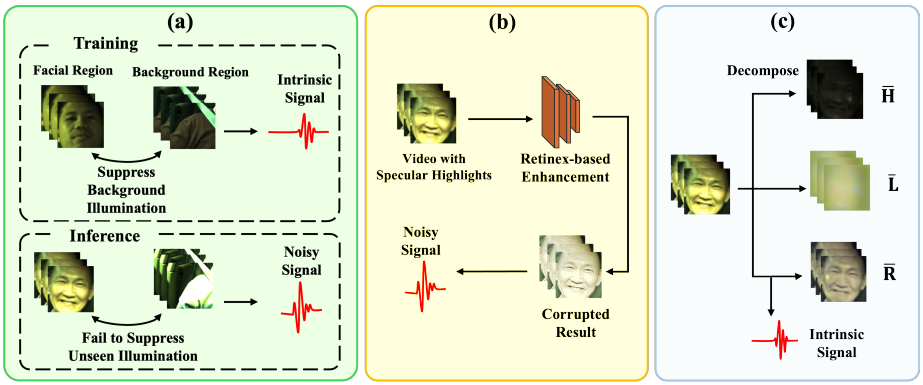

Fig. 1: Comparison of illumination-handling strategies for rPPG estimation. (a) **Background-based methods** leverages background regions as illumination priors for noise suppression, but often fail to generalize to unseen lighting conditions. (b) **Retinex-based methods** decompose illumination from reflectance to enhance signal extraction, yet ignore specular highlights due to the Lambertian assumption. (c) **SHINE-PPG (Ours)** leverages non-Lambertian intrinsic decomposition to explicitly separate illumination, reflectance, and specular components for robust rPPG extraction.

optical absorption and reflection properties. Despite its great potential, rPPG signals are inherently fragile and highly sensitive to environmental disturbances. In particular, illumination variations, such as flickering artificial lights, moving shadows, and fluctuating sunlight, introduce intensity changes that often dominate the underlying physiological signal. Therefore, effectively mitigating the impact of illumination interference remains a critical challenge for achieving robust rPPG estimation in real-world scenarios.

To mitigate illumination interference, recent approaches [10, 21, 32] attempt to model ambient lighting by exploiting background regions as illumination prior to suppress noise in facial regions, as illustrated in Fig. 1 (a). However, such background-dependent strategies are inherently limited by the diversity of lighting conditions present in the training data and often fail to generalize to unseen or complex illumination encountered during inference. Alternatively, other methods [4, 40] adopt Retinex theory [19] to decompose facial videos and mitigate illumination effects prior to rPPG extraction. While these approaches effectively model diffuse skin reflectance, they fundamentally rely on the Lambertian assumption [1], which treats the skin as an ideal matte surface. Consequently, they neglect **non-Lambertian specular reflections** [2, 35], the "shine" or highlight caused by the interaction between incident light and the skin surface. As shown in Fig. 1 (b), the presence of such specular highlights result in corrupted facial representation and noisy physiological signals.

To address the above limitations, we propose **SHINE-PPG** (**S**pecular-**H**ighlight **I**ntrinsic **N**etwork for rPPG **E**stimation), a novel framework inspired by non-Lambertian intrinsic image decomposition [35]. As illustrated in Fig. 1 (c),

our method explicitly decouples facial videos into three physically distinct components: illumination (**L**), reflectance (**R**), and specular highlights (**H**). By isolating the intrinsic reflectance from both ambient lighting and surface specularities, SHINE-PPG enables reliable rPPG extraction even under severe illumination interference. Since ground-truth intrinsic components are unavailable in existing rPPG datasets, we formulate the decomposition in a **self-supervised manner** using physics-grounded constraints. Specifically, we exploit spatial frequency-domain characteristics to distinguish slow-varying illumination and pulse-related reflectance, while incorporating spatial sparsity, dark-channel priors [18], and color consistency to isolate specular highlights. Furthermore, we introduce an **adversarial illumination enhancement strategy** that synthesizes challenging lighting conditions during training to substantially improve out-of-distribution generalization. Extensive experiments on multiple benchmark datasets demonstrate that SHINE-PPG consistently outperforms previous rPPG methods, particularly under complex and challenging illumination scenarios.

Our contributions are summarized as follows:

- We present SHINE-PPG, a novel rPPG estimation framework that explicitly models illumination, reflectance, and specular highlights via non-Lambertian intrinsic decomposition to tackle diverse lighting interference.
- We develop a self-supervised strategy that decomposes facial videos into intrinsic components to enable robust extraction of rPPG signals from reflectance while suppressing illumination- and highlight-induced noise.
- We introduce an adversarial illumination enhancement strategy to dynamically simulate unseen lighting conditions for effectively improving model robustness and out-of-distribution generalization.
- Extensive evaluations on multiple benchmarks demonstrate state-of-the-art performance and superior cross-dataset generalizability, particularly under challenging illumination scenarios.

2 Related Works

rPPG Estimation Remote photoplethysmography (rPPG) extracts physiological signals from subtle facial color changes and provides valuable liveness cues, especially for face anti-spoofing applications [5, 11–14]. However, the underlying signals are inherently weak and highly susceptible to external interference. Early methods [6, 38] leveraged the Dichromatic Reflection Model (DRM) [31] to separate pulsatile diffuse reflection from physiologically irrelevant specular reflections. Recent learning-based methods [22, 41, 43] employed neural architectures to extract discriminative representations from facial videos under challenging illumination conditions. To mitigate lighting interference, several methods [7, 23, 26] introduced spatial–temporal maps (STMaps) combined with data augmentation strategies to suppress noise and enhance signal quality. Other methods [10, 21, 32] utilized background illumination priors to remove facial lighting effects for improving rPPG signal extraction. Despite these advances, previous methods often

suffer from limited generalization to unseen or highly dynamic illumination conditions. Consequently, effectively handling illumination variations remains a critical challenge for rPPG estimation.

Intrinsic Image Decomposition Intrinsic image decomposition decouples an image into its reflectance and environmental illumination components. Given its inherently ill-posed nature, early methods [30, 33, 34] relied on physical priors, while recent learning-based methods [20, 25] utilize deep neural networks to learn the decomposition directly from data. The Retinex theory [19] remains a cornerstone of this field and has been widely integrated into various computer vision tasks [8, 39]. Motivated by its illumination-tolerant properties, recent rPPG studies have adopted Retinex-based decomposition. For example, methods such as [4] assumed the rPPG signal resides in the reflectance component and employed an image enhancement module [24] to improve extraction robustness, while another method [40] utilized Retinex-based adaptive illumination to extract rPPG signals in extreme low-light conditions. However, these existing methods typically operated under the Lambertian assumption and did not account for non-Lambertian specular highlights that frequently corrupt facial skin signals.

3 Preliminary

To formalize the non-Lambertian decomposition adopted in our framework, we first revisit classical Retinex theory [19]. Under the Lambertian assumption, an observed image $\mathbf{I}$ can be decomposed into two components as:

$$\mathbf{I} = \mathbf{R} \circ \mathbf{L}, \tag{1}$$

where $\circ$ denotes the element-wise product, $\mathbf{R}$ represents the intrinsic surface reflectance, and $\mathbf{L}$ denotes the environmental illumination.

However, human facial skin inherently exhibits non-Lambertian characteristics due to surface specularities, particularly under oily or sweaty conditions. To model this realistic behavior, we adopt the formulation in [2], which incorporates the Dichromatic Reflection Model (DRM) [31]. Under DRM, the observed image is explicitly represented as the sum of diffuse and specular-highlight reflections, *i.e.*, $\mathbf{I} = \mathbf{D} + \mathbf{H}$, where $\mathbf{D}$ and $\mathbf{H}$ denote the diffuse and specular-highlight components, respectively. Since the diffuse reflection $\mathbf{D}$ follows the Lambertian assumption, it can be factorized as $\mathbf{D} = \mathbf{R} \circ \mathbf{L}$. The complete non-Lambertian intrinsic decomposition therefore becomes:

$$\mathbf{I} = \mathbf{R} \circ \mathbf{L} + \mathbf{H}. \tag{2}$$

For video-based rPPG estimation, we extend this spatial image formation model to temporal sequences. Let $\mathbf{V}_t$ denote the t-th frame of a facial video. The non-Lambertian model is written as:

$$\mathbf{V}_t = \mathbf{R}_t \circ \mathbf{L}_t + \mathbf{H}_t + \boldsymbol{\eta}_t, \tag{3}$$

where $\boldsymbol{\eta}_t$ represents a composite stochastic noise accounting for motion artifacts and sensor noise, and is therefore omitted from the optimization objective.

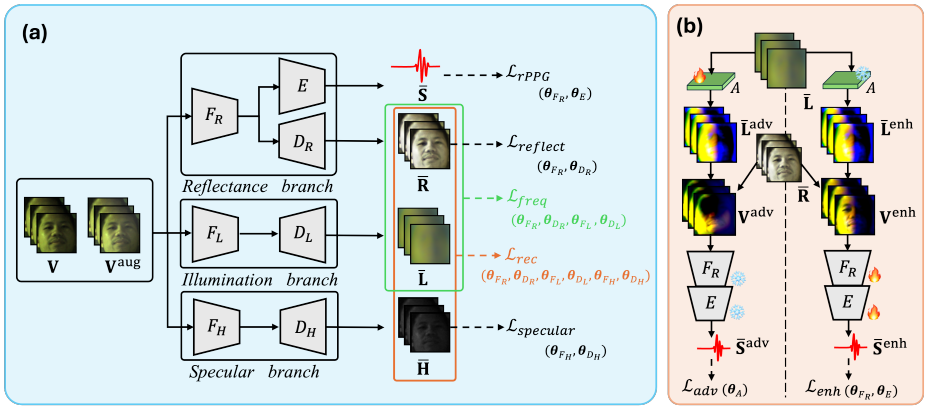

Fig. 2: Overview of SHINE-PPG. (a) **Self-Supervised Non-Lambertian Decomposition**: Facial videos are decoupled into intrinsic reflectance ($\bar{\mathbf{R}}$), illumination ($\bar{\mathbf{L}}$), and specular ($\bar{\mathbf{H}}$) to isolate intrinsic rPPG signals from environmental interference. (b) **Adversarial Illumination Enhancement**: Out-of-distribution lighting conditions are synthesized via AdaIN to improve model robustness and cross-dataset generalization.

Notably, the rPPG signal is primarily embedded in the intrinsic reflectance component $\mathbf{R}_t$, whereas the specular-highlight reflection $\mathbf{H}_t$ carries no physiological information [6, 38]. Therefore, accurately isolating $\mathbf{R}_t$ from illumination and specular-highlight components is a fundamental prerequisite for robust rPPG estimation under challenging lighting conditions.

4 Methodology

4.1 Overview

The overall architecture of SHINE-PPG is illustrated in Fig. 2. Given an input facial video $\mathbf{V} = \{\mathbf{V}_t\}_{t=1}^{T}$, we aim to decompose the observed signals into their fundamental physical components to enable robust rPPG estimation. As shown in Fig. 2 (a), SHINE-PPG consists of three parallel branches to explicitly model the non-Lambertian image formation process. Each branch is equipped with a task-specific feature extractor (F_L,F_R,F_H) and a corresponding decoder (D_L,D_R,D_H). The **reflectance branch** decouples the intrinsic reflectance $\bar{\mathbf{R}}$ from optical interference, serving as a purified physiological source for the rPPG estimator E. Simultaneously, the **illumination branch** captures the environmental lighting component $\bar{\mathbf{L}}$ to model spatial-temporal illumination changes, while the **specular branch** isolates surface specular highlights $\bar{\mathbf{H}}$ to suppress non-Lambertian corruption.

To enhance generalization under unseen lighting conditions, we further introduce an **Adversarial Illumination Enhancement** strategy, as illustrated in Fig. 2 (b). A learnable Adaptive Instance Normalization (AdaIN) [15] layer A synthesizes challenging adversarial illumination components $\bar{\mathbf{L}}^{\text{adv}}$, which are

re-composed with the intrinsic reflectance to generate augmented video samples $\mathbf{V}^{\text{adv}}$. This adversarial scheme forces the reflectance branch and estimator to remain stable under extreme lighting variations and effectively expands the training illumination distribution for improving overall robustness.

4.2 Self-Supervised Non-Lambertian Decomposition for rPPG

We describes our self-supervised strategy for decomposing facial videos into illumination, reflectance, and specular components, while ensuring that the **intrinsic rPPG signal** is faithfully preserved within the reflectance component.

Paired Data Augmentation and Modeling Since rPPG datasets lack paired samples of identical scenes under varying illumination, which is required for stable intrinsic decomposition, we simulate illumination variations by applying color jitter to the input video $\mathbf{V}$ to generate an augmented counterpart $\mathbf{V}^{\text{aug}}$. We employ three parallel branches to model these components:

$$\bar{\mathbf{L}} = D_L(F_L(\mathbf{V})), \quad \bar{\mathbf{R}} = D_R(F_R(\mathbf{V})), \quad \bar{\mathbf{H}} = D_H(F_H(\mathbf{V})), \tag{4}$$

where $F_{(\cdot)}$ and $D_{(\cdot)}$ denote the corresponding feature extractors and decoders. The same operations are applied to $\mathbf{V}^{\text{aug}}$.

Intrinsic Reflectance Invariance Reflectance is an intrinsic property of the subject and should remain invariant to illumination changes. To enforce this constraint, we define a reflectance consistency loss:

$$\mathcal{L}_{\text{reflect}}(\boldsymbol{\theta}_{F_R}, \boldsymbol{\theta}_{D_R}) = \sum_{t=1}^{T} \|\bar{\mathbf{R}}_t - \bar{\mathbf{R}}_t^{\text{aug}}\|_2^2. \tag{5}$$

Specular Highlight Modeling To isolate the non-Lambertian specular component $\bar{\mathbf{H}}$, we introduce three physics-grounded constraints:

Spatial Sparsity: Since specular highlights are typically localized and sparse, we impose an L_1 penalty:

$$\mathcal{L}_{\text{sparse}}(\boldsymbol{\theta}_{F_H}, \boldsymbol{\theta}_{D_H}) = \sum_{t=1}^{T} (\|\bar{\mathbf{H}}_t\|_1 + \|\bar{\mathbf{H}}_t^{\text{aug}}\|_1). \tag{6}$$

Distributional Prior: We adopt the normalized dark channel $\mathbf{W}_t$ as a probabilistic prior for highlight localization. Motivated by the observation that specular highlights tend to dominate the brightest responses across color channels [18], we compute the normalized dark channel prior $\mathbf{W}_t$ and spatially normalize it over all P pixels as:

$$\mathbf{W}_t(p) = \frac{\min_{c\in\{r,g,b\}} \mathbf{V}_{t,c}(p)}{\sum_{p=1}^{P} \min_{c\in\{r,g,b\}} \mathbf{V}_{t,c}(p)}, \tag{7}$$

where c indexes the RGB channels and p denotes the pixel location. To align the predicted highlights with this prior, we define the estimated specular magnitude $\bar{\mathbf{M}}_t$ as the spatially normalized L_2 norm of the predicted highlight channels:

$$\bar{\mathbf{M}}_t(p) = \frac{\sqrt{\bar{\mathbf{H}}_{t,r}^2(p) + \bar{\mathbf{H}}_{t,g}^2(p) + \bar{\mathbf{H}}_{t,b}^2(p)}}{\sum_{p=1}^{P} \sqrt{\bar{\mathbf{H}}_{t,r}^2(p) + \bar{\mathbf{H}}_{t,g}^2(p) + \bar{\mathbf{H}}_{t,b}^2(p)}}. \tag{8}$$

We then enforce consistency between the predicted distribution $\bar{\mathbf{M}}_t$ and the dark-channel prior $\mathbf{W}_t$ by minimizing their KL divergence:

$$\mathcal{L}_{\text{prior}}(\boldsymbol{\theta}_{F_H}, \boldsymbol{\theta}_{D_H}) = \sum_{t=1}^{T} (KL(\mathbf{W}_t \| \bar{\mathbf{M}}_t) + KL(\mathbf{W}_t^{\text{aug}} \| \bar{\mathbf{M}}_t^{\text{aug}})). \tag{9}$$

This encourages the estimated specular regions to align with the physically motivated highlight prior.

Color Consistency: Since specular reflections approximate the light source color, their chromaticity should align with illumination $\bar{\mathbf{L}}_t$. We therefore define $\mathcal{L}_{\text{color}}(\boldsymbol{\theta}_{F_H}, \boldsymbol{\theta}_{D_H})$ to enforce a $\mathbf{W}_t$-weighted cosine similarity:

$$\mathcal{L}_{\text{color}} = \sum_{t,p} \mathbf{W}_t(p)(1 - \cos(\bar{\mathbf{H}}_t(p), \bar{\mathbf{L}}_t(p))) + \sum_{t,p} \mathbf{W}_t^{aug}(p)(1 - \cos(\bar{\mathbf{H}}_t^{aug}(p), \bar{\mathbf{L}}_t^{aug}(p))). \tag{10}$$

The total specular loss is defined as:

$$\mathcal{L}_{\text{specular}}(\boldsymbol{\theta}_{F_H}, \boldsymbol{\theta}_{D_H}) = \lambda_{\text{sparse}}\mathcal{L}_{\text{sparse}} + \lambda_{\text{prior}}\mathcal{L}_{\text{prior}} + \lambda_{\text{color}}\mathcal{L}_{\text{color}}, \tag{11}$$

where $\lambda_{sparse}, \lambda_{prior}$, and λ_{color} are hyperparameters controlling the relative importance of each term.

Reconstruction and rPPG Learning We enforce the non-Lambertian formation model (Eq. 3) via the reconstruction loss $\mathcal{L}_{rec}(\boldsymbol{\theta}_{F_R}, \boldsymbol{\theta}_{D_R}, \boldsymbol{\theta}_{F_L}, \boldsymbol{\theta}_{D_L}, \boldsymbol{\theta}_{F_H}, \boldsymbol{\theta}_{D_H})$:

$$\begin{aligned}\mathcal{L}_{rec} = &\sum_{t=1}^{T}\sum_{p=1}^{P} \| \log(\mathbf{V}_t(p) - \bar{\mathbf{H}}_t(p)) - \log(\bar{\mathbf{L}}_t(p)) - \log(\bar{\mathbf{R}}_t(p)) \|_2^2 \\ &+ \sum_{t=1}^{T}\sum_{p=1}^{P} \| \log(\mathbf{V}_t^{aug}(p) - \bar{\mathbf{H}}_t^{aug}(p)) - \log(\bar{\mathbf{L}}_t^{aug}(p)) - \log(\bar{\mathbf{R}}_t^{aug}(p)) \|_2^2.\end{aligned} \tag{12}$$

Concurrently, the rPPG estimator E predicts the intrinsic rPPG signals ($\bar{\mathbf{S}}$ and $\bar{\mathbf{S}}^{aug}$) from reflectance features ($F_R(\mathbf{V})$ and $F_R(\mathbf{V}^{aug})$) by minimizing the negative Pearson correlation ($NP(\cdot,\cdot)$) with the ground-truth rPPG signal $\mathbf{S}$ by:

$$\mathcal{L}_{rPPG}(\boldsymbol{\theta}_{F_R}, \boldsymbol{\theta}_E) = NP(\bar{\mathbf{S}}, \mathbf{S}) + NP(\bar{\mathbf{S}}^{aug}, \mathbf{S}), \tag{13}$$

where $\bar{\mathbf{S}} = E(F_R(\mathbf{V}))$ and $\bar{\mathbf{S}}^{aug} = E(F_R(\mathbf{V}^{aug}))$.

Frequency-Constrained Decomposition To further decouple illumination and reflectance, we leverage the prior that reflectance contains richer high-frequency details, whereas illumination is spatially smooth [39]. We introduce a multi-frequency constraint:

$$\mathcal{L}_{freq}(\boldsymbol{\theta}_{F_R}, \boldsymbol{\theta}_{D_R}, \boldsymbol{\theta}_{F_L}, \boldsymbol{\theta}_{D_L}) = \sum_{t=1}^{T} \sum_{\sigma \in \mathcal{T}} \Big(\mathcal{G}(\bar{\mathbf{R}}_t, \bar{\mathbf{R}}_t^{aug}; \sigma) + \hat{\mathcal{G}}(\bar{\mathbf{L}}_t, \bar{\mathbf{L}}_t^{aug}; \sigma) \Big), \quad (14)$$

where the reflectance penalty $\mathcal{G}$ and the illumination penalty $\hat{\mathcal{G}}$ are defined as:

$$\begin{aligned} \mathcal{G}(\bar{\mathbf{R}}_t, \bar{\mathbf{R}}_t^{aug}; \sigma) &= w(\sigma) \cdot \Big(\|\mathcal{F}(\bar{\mathbf{R}}_t) \odot \mathcal{M}(\sigma)\|_1 + \|\mathcal{F}(\bar{\mathbf{R}}_t^{aug}) \odot \mathcal{M}(\sigma)\|_1 \Big), \\ \hat{\mathcal{G}}(\bar{\mathbf{L}}_t, \bar{\mathbf{L}}_t^{aug}; \sigma) &= (1 - w(\sigma)) \cdot \Big(\|\mathcal{F}(\bar{\mathbf{L}}_t) \odot (1 - \mathcal{M}(\sigma))\|_1 + \|\mathcal{F}(\bar{\mathbf{L}}_t^{aug}) \odot (1 - \mathcal{M}(\sigma))\|_1 \Big). \end{aligned} \quad (15)$$

Here, $\mathcal{F}$ denotes the Fast Fourier Transform, and $\mathcal{M}(u, v; \sigma) = \exp\left(-\frac{D^2(u,v)}{2\sigma^2}\right)$ represents a Gaussian low-pass filter in the frequency domain, where $D(u, v)$ is the distance to the frequency center and σ controls the cutoff frequency. The two penalty terms in Eq. (15) are specifically designed to enforce component decomposition. Specifically, $\mathcal{G}$ applies the low-pass filter $\mathcal{M}$ to penalize low-frequency energy in the reflectance component, encouraging preserving high-frequency texture details. Conversely, $\hat{\mathcal{G}}$ employs the complementary high-pass filter $(1 - \mathcal{M})$ to suppress high-frequency energy in the illumination component, ensuring spatial smoothness. We further introduce a set of frequency thresholds $\mathcal{T} = \{\sigma\}$ to impose multi-scale constraints, with adaptive weights $w(\sigma) \in [0, 1]$ balancing the penalties across different frequency bands.

4.3 Adversarial Illumination Enhancement

This section introduces the adversarial framework designed to improve robustness against out-of-distribution lighting. By leveraging the intrinsic reflectance decoupled in previous stages, we synthesize challenging illumination variations that force the rPPG estimator to generalize beyond the training distribution.

Adversarial Illumination Learning To simulate diverse lighting conditions, we adopt a learnable Adaptive Instance Normalization (AdaIN) layer A to transform the decomposed illumination $\bar{\mathbf{L}}$ into an adversarial counterpart $\bar{\mathbf{L}}^{\text{adv}}$:

$$\bar{\mathbf{L}}^{adv} = A(\bar{\mathbf{L}}) = \alpha \cdot \left(\frac{\bar{\mathbf{L}} - \mu_{\bar{\mathbf{L}}}}{\sigma_{\bar{\mathbf{L}}}}\right) + \beta, \quad (16)$$

where α and β are learnable parameters, $\mu_{\bar{\mathbf{L}}}$ and $\sigma_{\bar{\mathbf{L}}}$ denote the mean and standard deviation of $\bar{\mathbf{L}}$.

We then fix the reflectance branch F_R and estimator E and define an adversarial loss $\mathcal{L}_{\text{adv}}$ to optimize α and β. This objective seeks to maximize the rPPG

estimation error for enforcing A to generate increasingly difficult illumination conditions:

$$\mathcal{L}_{adv}(\boldsymbol{\theta}_A) = \arg\max_{\alpha,\beta} NP(\bar{\mathbf{S}}^{adv}, \mathbf{S}), \tag{17}$$

where $\bar{\mathbf{S}}^{\text{adv}} = E(F_R(\mathbf{V}^{\text{adv}}))$ is the rPPG signal estimated from adversarial samples $\mathbf{V}_t^{\text{adv}} = \bar{\mathbf{R}}_t \circ \bar{\mathbf{L}}_t^{\text{adv}}$. Notably, the specular component $\bar{\mathbf{H}}$ is intentionally excluded during synthesis. Since specular highlights carry no physiological information, removing them ensures the adversarial module focuses solely on simulating complex ambient variations rather than introducing irrelevant specular artifacts.

Illumination-Enhanced Feature Learning After optimizing A, we generate enhanced illumination components $\bar{\mathbf{L}}^{\text{enh}}$ and construct corresponding samples $\mathbf{V}_t^{\text{enh}} = \bar{\mathbf{R}}_t \circ \bar{\mathbf{L}}_t^{\text{enh}}$. These samples are then used to fine-tune F_R and E by minimizing the rPPG estimation error:

$$\mathcal{L}_{enh}(\boldsymbol{\theta}_{F_R}, \boldsymbol{\theta}_E) = NP(\bar{\mathbf{S}}^{enh}, \mathbf{S}), \tag{18}$$

where $\bar{\mathbf{S}}^{\text{enh}} = E(F_R(\mathbf{V}^{\text{enh}}))$. This two-step adversarial scheme effectively enlarges the illumination distribution and significantly enhances cross-dataset generalization under extreme lighting conditions.

4.4 Training and testing

Training To ensure stable convergence and reduce mutual interference among intrinsic components, we adopt a progressive four-stage training strategy:

Stage 1: Lambertian Initialization. We optimize the illumination, reflectance, and rPPG estimator parameters $\{\boldsymbol{\theta}_{F_R}, \boldsymbol{\theta}_{D_R}, \boldsymbol{\theta}_{F_L}, \boldsymbol{\theta}_{D_L}, \boldsymbol{\theta}_E\}$ using $\mathcal{L}_{\text{reflect}}$, $\mathcal{L}_{\text{rec}}$, $\mathcal{L}_{\text{rPPG}}$, and $\mathcal{L}_{\text{freq}}$, establishing a Lambertian baseline without specular modeling.

Stage 2: Specular Isolation. We freeze the previously learned parameters and update only the specular branch $\{\boldsymbol{\theta}_{F_H}, \boldsymbol{\theta}_{D_H}\}$ using $\mathcal{L}_{\text{specular}}$ and $\mathcal{L}_{\text{rec}}$ to isolate highlights without disturbing the Lambertian baseline decomposition.

Stage 3: Joint Refinement. All modules are unfrozen and jointly optimized using the full set of losses ($\mathcal{L}_{\text{reflect}}$, $\mathcal{L}_{\text{rec}}$, $\mathcal{L}_{\text{rPPG}}$, $\mathcal{L}_{\text{freq}}$, and $\mathcal{L}_{\text{specular}}$) for refinement and improved component consistency.

Stage 4: Adversarial Enhancement. We first update the AdaIN parameters $\boldsymbol{\theta}_A$ by maximizing $\mathcal{L}_{\text{adv}}$ while freezing $\{\boldsymbol{\theta}_{F_R}, \boldsymbol{\theta}_E\}$. Subsequently, $\boldsymbol{\theta}_A$ is fixed and $\{\boldsymbol{\theta}_{F_R}, \boldsymbol{\theta}_E\}$ are fine-tuned via $\mathcal{L}_{\text{enh}}$ to improve robustness under unseen illumination conditions.

Testing During inference, only the reflectance extractor F_R and the estimator E are utilized to extract the intrinsic rPPG signal, ensuring efficient heart rate estimation without the computational overhead of the auxiliary decomposition branches.

Table 1: Intra-domain testing within different illumination conditions.

Type	Method	UBFC			PURE			COHFACE			Indoor			Car		
		MAE↓	RMSE↓	R↑	MAE↓	RMSE↓	R↑	MAE↓	RMSE↓	R↑	MAE↓	RMSE↓	R↑	MAE↓	RMSE↓	R↑
	PhysFormer [41] (*CVPR 22*)	1.88	4.47	0.96	0.43	0.57	**0.99**	3.24	6.46	0.81	1.57	3.58	0.93	4.72	7.92	0.45
	EfficientPhys [22] (*WACV 23*)	0.72	1.77	**0.99**	0.30	0.36	**0.99**	2.75	6.99	0.81	1.16	3.06	0.96	4.79	9.76	0.48
	RhythmFormer [43] (*PR 25*)	0.49	1.37	**0.99**	0.16	0.21	**0.99**	**1.17**	**3.36**	**0.97**	0.67	1.57	0.98	2.87	6.16	0.72
	RhythmMamba [44] (*AAAI 25*)	0.70	1.02	**0.99**	0.48	0.55	**0.99**	7.70	14.18	0.61	0.46	0.66	**0.99**	4.90	10.22	0.53
Background-based	ND-DeeprPPG [21] (*TIP 23*)	0.78	1.94	0.98	0.24	0.32	**0.99**	1.56	3.95	0.93	0.52	0.93	**0.99**	4.28	8.76	0.54
	Shao *et al.* [32] (*CVPR 25*)	0.42	0.89	**0.99**	0.18	0.24	**0.99**	6.60	12.60	0.49	0.45	0.91	0.92	2.69	6.03	0.87
Retinex-based	Chen *et al.* [4] (*CVPRW 23*)	0.62	2.20	**0.99**	0.19	0.23	**0.99**	2.23	5.81	0.85	0.55	1.33	**0.99**	2.16	4.90	0.78
	Yang *et al.* [40] (*TCE 25*)	1.18	2.01	**0.99**	1.69	2.97	0.86	12.59	16.29	0.10	1.79	5.81	0.87	4.42	8.26	0.62
Non-Lambertian	**SHINE-PPG (Ours)**	**0.34**	**0.47**	**0.99**	**0.13**	**0.17**	**0.99**	1.47	4.03	0.93	**0.32**	**0.45**	**0.99**	**1.68**	**3.55**	**0.89**

5 Experiments

5.1 Experimental Setting

Network Architecture We adopt the feature extractors (F_L, F_R, F_H) and the rPPG estimator (E) from [37]. To reconstruct the intrinsic components, we design symmetric decoders (D_L, D_R, D_H) with skip connections following [29]. The adversarial module (A) is implemented using a learnable AdaIN layer [17] parameterized by two learnable variables to modulate illumination statistics.

Implementation Details The model is trained for 100 epochs on a single NVIDIA RTX 4090 GPU with a batch size of 1, using AdamW optimizer (learning rate 1×10^{-4}). Faces are detected using MTCNN [42] and cropped to 64×64. Following [32], each training clip consist of 320 consecutive RGB frames. The loss weights are set to: $(\lambda_{prior}, \lambda_{color}) = 2$; $(\lambda_{sparse}, \lambda_{adv}, \lambda_{enh}) = 1$; $\lambda_{specular} = 5$; $(\lambda_{reflect}, \lambda_{rec}, \lambda_{rPPG}, \lambda_{freq}) = 3$.

Datasets and Evaluation Metrics We evaluate SHINE-PPG on five public datasets: **UBFC** [3], **PURE** [36], **COHFACE** [9], **MR-NIRP Indoor** (denoted as **Indoor**) [27], and **MR-NIRP Car** (denoted as **Car**) [28]. The first four datasets are collected in a relatively constrained environments with stable illumination, whereas the **Car** dataset, specifically its Driving protocol, exhibits rapidly varying lighting conditions and serves as our primary challenge. Performance is evaluated using three standard metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Pearson Correlation Coefficient (R).

5.2 Experimental Results

Intra-Domain Testing As shown in Tab. 1, existing rPPG methods perform well under stable and controlled illumination like **UBFC** and **PURE**, but their accuracy degrades under challenging lighting conditions of the **Car** dataset. Notably, methods like PhysFormer [41] and EfficientPhys [22], without explicit mechanisms to handle illumination variations, yield MAE values exceeding 4.70 bpm on the **Car** dataset. While Retinex-based [4] and background-based [32]

Table 2: Cross-domain testing within different illumination conditions.

Type	Method	UBFC → Indoor			PURE → Indoor			UBFC → Car			PURE → Car		
		MAE↓	RMSE↓	R↑	MAE↓	RMSE↓	R↑	MAE↓	RMSE↓	R↑	MAE↓	RMSE↓	R↑
	PhysFormer [41] (*CVPR 22*)	2.67	4.43	0.92	0.69	1.42	0.98	16.33	22.85	0.11	6.25	9.07	0.41
	EfficientPhys [22] (*WACV 23*)	1.38	5.35	0.87	0.50	0.74	**0.99**	18.89	29.85	0.04	15.78	24.24	0.06
	RhythmFormer [43] (*PR 25*)	1.08	3.98	0.91	0.21	0.27	**0.99**	11.08	17.11	0.18	4.64	8.62	0.53
	RhythmMamba [44] (*AAAI 25*)	0.85	1.74	0.98	0.56	0.77	**0.99**	9.80	16.30	0.24	9.37	16.62	0.28
Background-based	ND-DeeprPPG [21] (*TIP 23*)	1.03	3.09	0.96	0.34	0.35	**0.99**	12.36	20.64	0.28	7.85	13.53	0.38
	DD-rPPGNet [10] (*TIFS 25*)	1.22	2.67	0.96	0.63	1.50	0.98	14.92	21.33	0.15	7.47	11.78	0.43
Retinex-based	Chen *et al.* [4] (*CVPRW 23*)	2.27	9.44	0.63	0.38	0.58	**0.99**	15.86	24.01	0.01	6.81	13.91	0.29
Non-Lambertian	**SHINE-PPG (Ours)**	**0.30**	**0.42**	**0.99**	**0.16**	**0.19**	**0.99**	**7.64**	**14.42**	**0.30**	**3.24**	**6.59**	**0.74**

methods show improved robustness ($R = 0.78$ and 0.87, respectively), SHINE-PPG achieves the best overall performance with an MAE of 1.68 bpm and an R value of 0.89. This demonstrates that explicitly modeling specular highlights is critical even when the testing and training data share the same domain.

Cross-Domain Testing Tab. 2 presents the cross-domain results, where A → B denotes that the model is trained on dataset A and tested on dataset B. While **UBFC → Indoor** and **PURE → Indoor** are well handled by most methods, since both the training and testing datasets exhibit relatively stable illumination, the transition to dynamic environments (**Car**) leads to substantial performance degradation in previous models. For example, in **PURE → Car**, EfficientPhys [22] drops to a near-zero correlation ($R = 0.06$). This degradation is largely due to severe illumination variations in the **Car** dataset, with approximately 11.9% of facial pixels belonging to specular highlight regions estimated by the pretrained M2-Net [16]. In contrast, SHINE-PPG maintains remarkable stability under dynamic lighting. In the particularly challenging **UBFC → Car** task, characterized by limited training data and highly variable test conditions, our method achieves an MAE of 7.64 bpm, significantly outperforming RhythmMamba [44] and RhythmFormer [43]. Moreover, in **PURE → Car**, SHINE-PPG attains an R value of 0.74, nearly doubling the performance of background-based [10, 21] and Retinex-based [4] methods. These results demonstrate that isolating intrinsic reflectance and incorporating adversarial illumination enhancement are essential for bridging the domain gap between constrained training conditions and real-world dynamic environments. Despite these unseen and different lighting conditions, SHINE-PPG consistently achieves strong performance, demonstrating robust cross-domain generalization under challenging OOD scenarios.

Robustness to Challenging Illumination Scenario In Tab. 3, we evaluate the generalization capability across diverse illumination conditions. Specifically, rPPG models are trained using the training videos of the Garage protocol from the **Car** dataset, and evaluated on three test sets: (1) the original Garage test videos (Garage), (2) Garage test videos with additional Color Jitter applied to simulate illumination variations (Garage + Color Jitter), and (3) the testing videos from

Table 3: Experimental comparisons on different challenging illumination scenarios.

Type	Method	Garage			Garage + Color Jitter			Driving		
		MAE↓	RMSE↓	R↑	MAE↓	RMSE↓	R↑	MAE↓	RMSE↓	R↑
	PhysFormer [41]	1.26	3.87	0.90	3.53	6.36	0.71	10.66	14.02	0.08
	EfficientPhys [22]	4.39	9.38	0.52	5.50	13.21	0.31	11.84	17.74	0.22
	RhythmFormer [43]	0.86	2.40	0.96	1.69	4.86	0.86	9.56	14.36	0.12
	RhythmMamba [44]	3.90	9.14	0.65	5.14	13.22	0.54	11.69	18.57	0.29
Background-based	DD-rPPGNet [10]	8.29	11.75	0.41	9.24	12.84	0.40	14.31	17.51	0.11
	ND-DeeprPPG [21]	1.04	4.39	0.88	1.72	5.12	0.83	8.18	12.75	0.27
Retinex-based	Chen *et al.* [4]	1.01	2.51	0.96	1.04	2.98	0.95	8.94	14.10	0.10
Non-Lambertian	**SHINE-PPG (Ours)**	**0.51**	**0.78**	**0.99**	**0.52**	**0.80**	**0.99**	**6.80**	**11.18**	**0.37**

Table 4: Ablation study of different loss combinations for SHINE-PPG.

Loss Terms					PURE → Car		
$\mathcal{L}_{\text{rPPG}}$	$\mathcal{L}_{\text{rec}}$	$\mathcal{L}_{\text{reflect}}$	$\mathcal{L}_{\text{freq}}$	$\mathcal{L}_{\text{specular}}$	MAE↓	RMSE	R↑
✓	–	–	–	–	8.65	14.03	0.32
✓	✓	–	–	–	9.97	17.81	0.19
✓	✓	✓	–	–	6.23	11.52	0.39
✓	✓	–	✓	–	6.02	10.49	0.43
✓	✓	✓	✓	–	4.17	8.77	0.48
✓	✓	✓	✓	✓	**3.76**	**7.76**	**0.57**

Table 5: Ablation study on different illumination enhancements.

Enhancement strategies	PURE → Car		
	MAE	RMSE	R
Baseline	3.76	7.76	0.57
Color Jitter	3.56	7.74	0.58
Noise Injection	3.73	7.80	0.58
AdaIN	3.53	7.28	0.61
Ours	**3.24**	**6.59**	**0.74**

the Driving protocol, which contains substantial real-world lighting fluctuations (Driving). First, methods such as [41, 43], which are not explicitly designed to address illumination variations, exhibit significant performance degradation under varying lighting conditions, particularly in the **Garage + Color Jitter** and **Driving** settings. Background-based methods [10, 21], developed for interference suppression, partially alleviate illumination effects; however, their robustness strictly depends on the specific lighting conditions observed during training. Similarly, although Retinex-based methods [4] employ intrinsic decomposition to recover reflectance, they remains vulnerable to signal corruption induced by specular highlights. In contrast, SHINE-PPG leverages non-Lambertian intrinsic decomposition to explicitly separate specular and illumination components. By isolating the intrinsic reflectance, our framework effectively handles unknown and challenging lighting conditions and consistently outperforms prior methods across all evaluation protocols.

5.3 Ablation Studies

Effectiveness of Decomposition Constraints In Tab. 4, we evaluate the impact of different loss combinations on the challenging **PURE → Car** task. The baseline model, trained solely with $\mathcal{L}_{\text{rPPG}}$, is highly vulnerable to dynamic lighting and yields a low correlation ($R = 0.32$). Introducing the reconstruction loss $\mathcal{L}_{\text{rec}}$ without additional guidance further degrades performance ($R = 0.19$). This is

Table 6: Evaluation of rPPG in the decomposed $\bar{\mathbf{L}}$ and $\bar{\mathbf{H}}$.

MR-NIRP (Car)			
Source	MAE ↓	RMSE ↓	R ↑
Illumination ($\bar{\mathbf{L}}$)	15.91	21.32	-0.04
Specular ($\bar{\mathbf{H}}$)	18.46	23.64	-0.03
Original (V)	**1.68**	**3.55**	**0.89**

due to the ill-posed nature of intrinsic decomposition; that is, without sufficient constraints to separate illumination from reflectance, the network tends to corrupt subtle physiological signals. By introducing the frequency constraint ($\mathcal{L}_{\text{freq}}$) and reflectance consistency ($\mathcal{L}_{\text{reflect}}$), we incorporate key physical priors, namely the spatial smoothness of illumination and the invariance of skin reflectance, to stabilize the decomposition process. These terms jointly lead to a substantial improvement, increasing the correlation to 0.48. Finally, the specular constraint ($\mathcal{L}_{\text{specular}}$) provides the critical non-Lambertian refinement. By explicitly isolating sparse highlights through dark-channel and color-consistency priors, it removes residual optical interference and further recovers the intrinsic rPPG signal, achieving the best overall result with a Pearson correlation of 0.57.

Impact of Out-of-Distribution Illumination Enhancement In Tab. 5, we evaluate different strategies for expanding the decomposed illumination $\bar{\mathbf{L}}$ to enhance model training. Using the full model from Tab. 4 as the baseline, we apply Color Jitter, Noise Injection, AdaIN, and our proposed adversarial enhancement to $\bar{\mathbf{L}}$ to reconstruct augmented samples for Eq. (18). Conventional augmentations such as Color Jitter and Noise Injection provide only marginal improvements, slightly increasing R from 0.57 to 0.58, as they rely on simple global perturbations (e.g., brightness, contrast, saturation, and hue) and fail to capture the physically complex lighting variations required for robust generalization. Similarly, standard AdaIN, defined as $\text{AdaIN}(x_1, x_2) = \sigma(x_2)\frac{x_1-\mu(x_1)}{\sigma(x_1)} + \mu(x_2)$, relies on feature-wise mean and variance statistics to transfer lighting styles between samples, yet is inherently restricted to the lighting distributions observed in the training data. In contrast, our adversarial illumination enhancement replaces fixed statistics ($\mu(x_2)$ and $\sigma(x_2)$) with learnable parameters α and β in Eq. (16) and optimizes them adversarially to simulate out-of-distribution lighting. By forcing the model to maintain stability under these simulated extremes, the framework learns strong illumination-invariant physiological features, reducing the MAE from 3.76 to 3.24 bpm and increasing the correlation from 0.57 to 0.74.

rPPG Preservation in Decomposed Components In Tab. 6, we evaluate the rPPG information content of each estimated component to quantitatively address the concern of potential signal loss in the ill-posed decomposition. When feeding the decomposed illumination $\bar{\mathbf{L}}$ and specular $\bar{\mathbf{H}}$ into the trained rPPG model F_R–E, both yield high MAE/RMSE and near-zero Pearson correlation with ground truth. These results indicate that they do not contain meaningful pulsatile

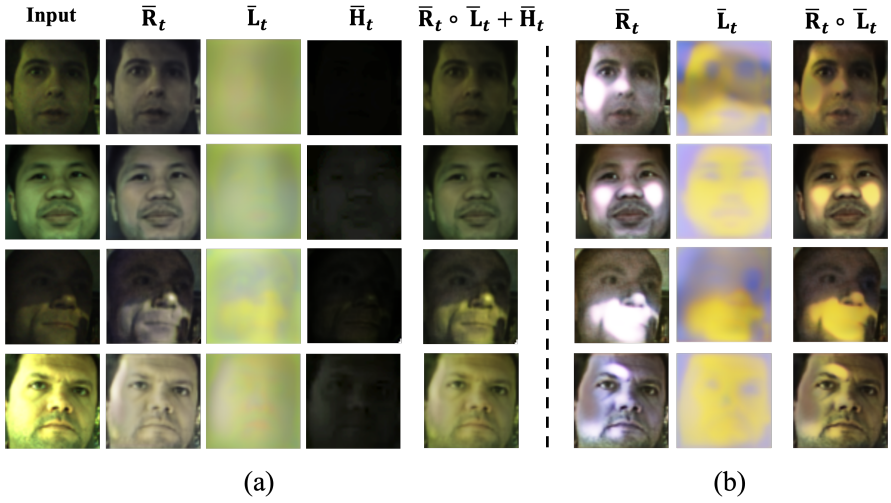

Fig. 3: Decomposition Visualizations. (a) **Non-Lambertian Decomposition**: Input frames are decoupled into intrinsic reflectance ($\bar{\mathbf{R}}_t$), illumination ($\bar{\mathbf{L}}_t$), and specular ($\bar{\mathbf{H}}_t$) components. (b) **Retinex-based Decomposition**: Conventional Retinex decomposition produces localized white artifacts in the reflectance components ($\bar{\mathbf{R}}_t$) due to its inability to model specular highlights.

signals but instead behave as noise, empirically supporting that rPPG-relevant information is primarily preserved in the reflectance component $\mathbf{R}$.

5.4 Visualizations

Non-Lambertian Decomposition To validate the effectiveness of SHINE-PPG under dynamic lighting, we visualize the decomposed components for different subjects from the **Car** dataset and compare them with conventional Retinex-based decomposition in Fig. 3.

As shown in Fig. 3 (a), the illumination components ($\bar{\mathbf{L}}_t$) accurately capture smooth and spatially varying lighting distributions, while the specular components ($\bar{\mathbf{H}}_t$) successfully isolate localized facial highlights. Consequently, the intrinsic reflectance components ($\bar{\mathbf{R}}_t$) preserve fundamental skin color and fine textures, providing a reliable source for robust rPPG extraction. Moreover, the high fidelity between the reconstructed frames ($\bar{\mathbf{R}}_t \circ \bar{\mathbf{L}}_t + \bar{\mathbf{H}}_t$) and the original inputs confirms the effectiveness of our self-supervised decomposition without requiring ground-truth supervision. In contrast, conventional Retinex decomposition in Fig. 3 (b) fails to model surface specularities, leaving residual highlight artifacts in the reflectance component that contaminate the underlying physiological signals. By explicitly modeling non-Lambertian reflections, SHINE-PPG effectively removes these interferences and produces cleaner intrinsic representations for more accurate physiological sensing.

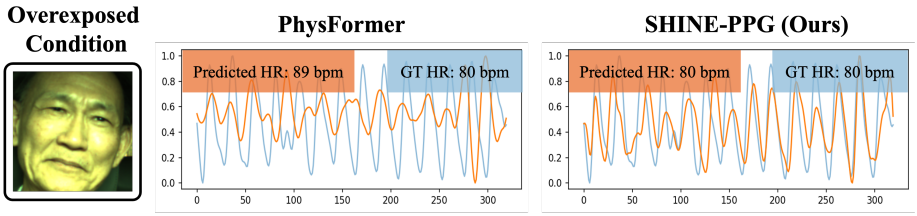

Fig. 4: Estimated Signal Visualization. Comparison between PhysFormer and the proposed SHINE-PPG under overexposed illumination condition.

Estimated signal visualization Fig. 4 visualizes the reconstructed rPPG waveforms under overexposure condition and shows that our method achieves significantly greater consistency with the ground truth than PhysFormer [41].

6 Conclusion

In this paper, we present SHINE-PPG, a novel framework designed to tackle the long-standing challenge of illumination interference in remote photoplethysmography. Moving beyond the conventional Lambertian assumption, our method leverages non-Lambertian intrinsic decomposition to explicitly decompose facial videos into illumination, reflectance, and specular components in a self-supervised manner. This physically grounded separation enables the isolation of intrinsic reflectance, effectively suppressing ambient light variations and surface specularities to recover high-fidelity rPPG signals. To further enhance robustness, we propose an adversarial illumination enhancement strategy that dynamically synthesizes challenging out-of-distribution lighting conditions. This design encourages the learning of illumination-invariant physiological representation, significantly improving cross-domain generalization. Extensive experiments across five benchmark datasets demonstrate that SHINE-PPG achieves state-of-the-art performance under complex illumination while exhibiting superior cross-dataset robustness. Overall, our work establishes a physically interpretable pathway for reliable physiological monitoring. While the current AdaIN-based augmentation effectively improves robustness, modeling spatially non-uniform and dynamically varying illumination remains an important direction for future research.

Acknowledgements

This work was supported in part by the National Science and Technology Council grant 112-2221-E-007-082-MY3 of Taiwan and the Natural Science Foundation of Fujian Province, China under Grant 2026J008139.

References

1. Barrow, H., Tenenbaum, J., Hanson, A., Riseman, E.: Recovering intrinsic scene characteristics. Comput. vis. syst **2**(3-26), 2 (1978)
2. Baslamisli, A.S., Le, H.A., Gevers, T.: Cnn based learning using reflection and retinex models for intrinsic image decomposition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 6674–6683 (2018)
3. Bobbia, S., Macwan, R., Benezeth, Y., Mansouri, A., Dubois, J.: Unsupervised skin tissue segmentation for remote photoplethysmography. Pattern recognition letters **124**, 82–90 (2019)
4. Chen, S., Ho, S.K., Chin, J.W., Luo, K.H., Chan, T.T., So, R.H., Wong, K.L.: Deep learning-based image enhancement for robust remote photoplethysmography in various illumination scenarios. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6077–6085 (2023)
5. Chong, J.X., Hsu, Y., Hsu, M.T., Lin, Y.T., Chien, K.H., Hsu, C.T., Huang, P.K.: Multi-modal face anti-spoofing via cross-modal feature transitions. Expert Systems with Applications p. 131292 (2026)
6. De Haan, G., Jeanne, V.: Robust pulse rate from chrominance-based rppg. IEEE transactions on biomedical engineering **60**(10), 2878–2886 (2013)
7. Du, J., Liu, S.Q., Zhang, B., Yuen, P.C.: Dual-bridging with adversarial noise generation for domain adaptive rppg estimation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 10355–10364 (2023)
8. Fu, H., Zheng, W., Meng, X., Wang, X., Wang, C., Ma, H.: You do not need additional priors or regularizers in retinex-based low-light image enhancement. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 18125–18134 (2023)
9. Heusch, G., Anjos, A., Marcel, S.: A reproducible study on remote heart rate measurement. arxiv 2017. arXiv preprint arXiv:1709.00962 **2**(6)
10. Huang, P.K., Chen, T.H., Chan, Y.T., Chen, K.W., Hsu, C.T.: Dd-rppgnet: Deinterfering and descriptive feature learning for unsupervised rppg estimation. IEEE Transactions on Information Forensics and Security (2025)
11. Huang, P.K., Chiang, C.H., Chen, T.H., Chong, J.X., Liu, T.L., Hsu, C.T.: One-class face anti-spoofing via spoof cue map-guided feature learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 277–286 (2024)
12. Huang, P.K., Chin, M.C., Hsu, C.T.: Face anti-spoofing via robust auxiliary estimation and discriminative feature learning. In: Asian Conference on Pattern Recognition. pp. 443–458. Springer (2021)
13. Huang, P.K., Chong, J.X., Hsu, M.T., Hsu, F.Y., Hsu, C.T.: Channel difference transformer for face anti-spoofing. Information Sciences **702**, 121904 (2025)
14. Huang, P.K., Chong, J.X., Hsu, M.T., Hsu, F.Y., Lin, Y.T., Chien, K.H., Hsu, C.T.: Enhancing learnable descriptive convolutional vision transformer for face anti-spoofing. Pattern Recognition p. 113289 (2026)
15. Huang, X., Belongie, S.: Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of the IEEE international conference on computer vision. pp. 1501–1510 (2017)
16. Huang, Z., Hu, K., Wang, X.: M2-net: multi-stages specular highlight detection and removal in multi-scenes. arXiv preprint arXiv:2207.09965 (2022)
17. Karras, T., Laine, S., Aila, T.: A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4401–4410 (2019)

18. Kim, H., Jin, H., Hadap, S., Kweon, I.: Specular reflection separation using dark channel prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1460–1467 (2013)
19. Land, E.H.: The retinex theory of color vision. Scientific american **237**(6), 108–129 (1977)
20. Li, Z., Snavely, N.: Cgintrinsics: Better intrinsic image decomposition through physically-based rendering. In: Proceedings of the European conference on computer vision (ECCV). pp. 371–387 (2018)
21. Liu, S.Q., Yuen, P.C.: Robust remote photoplethysmography estimation with environmental noise disentanglement. IEEE Transactions on Image Processing **33**, 27–41 (2023)
22. Liu, X., Hill, B., Jiang, Z., Patel, S., McDuff, D.: Efficientphys: Enabling simple, fast and accurate camera-based cardiac measurement. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision. pp. 5008–5017 (2023)
23. Lu, H., Han, H., Zhou, S.K.: Dual-gan: Joint bvp and noise modeling for remote physiological measurement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 12404–12413 (2021)
24. Ma, L., Ma, T., Liu, R., Fan, X., Luo, Z.: Toward fast, flexible, and robust low-light image enhancement. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 5637–5646 (2022)
25. Narihira, T., Maire, M., Yu, S.X.: Direct intrinsics: Learning albedo-shading decomposition by convolutional regression. In: Proceedings of the IEEE international conference on computer vision. pp. 2992–2992 (2015)
26. Niu, X., Shan, S., Han, H., Chen, X.: Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation. IEEE Transactions on Image Processing **29**, 2409–2423 (2019)
27. Nowara, E.M., Marks, T.K., Mansour, H., Veeraraghavan, A.: Sparseppg: Towards driver monitoring using camera-based vital signs estimation in near-infrared. In: CVPRW (2018)
28. Nowara, E.M., Marks, T.K., Mansour, H., Veeraraghavan, A.: Near-infrared imaging photoplethysmography during driving. IEEE transactions on intelligent transportation systems **23**(4), 3589–3600 (2020)
29. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: International Conference on Medical image computing and computer-assisted intervention. pp. 234–241. Springer (2015)
30. Rother, C., Kiefel, M., Zhang, L., Schölkopf, B., Gehler, P.: Recovering intrinsic images with a global sparsity prior on reflectance. Advances in neural information processing systems **24** (2011)
31. Shafer, S.A.: Using color to separate reflection components. Color Research & Application **10**(4), 210–218 (1985)
32. Shao, H., Luo, L., Qian, J., Yan, M., Chen, S., Yang, J.: Remote photoplethysmography in real-world and extreme lighting scenarios. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 10858–10867 (2025)
33. Shen, L., Tan, P., Lin, S.: Intrinsic image decomposition with non-local texture cues. In: 2008 IEEE Conference on Computer Vision and Pattern Recognition. pp. 1–7. IEEE (2008)
34. Shen, L., Yeo, C.: Intrinsic images decomposition using a local and global sparse representation of reflectance. In: CVPR 2011. pp. 697–704. IEEE (2011)
35. Shi, J., Dong, Y., Su, H., Yu, S.X.: Learning non-lambertian object intrinsics across shapenet categories. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 1685–1694 (2017)

36. Stricker, R., Müller, S., Gross, H.M.: Non-contact video-based pulse rate measurement on a mobile service robot. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication. pp. 1056–1062. IEEE (2014)
37. Sun, Z., Li, X.: Contrast-phys: Unsupervised video-based remote physiological measurement via spatiotemporal contrast. In: European Conference on Computer Vision. pp. 492–510. Springer (2022)
38. Wang, W., Den Brinker, A.C., Stuijk, S., De Haan, G.: Algorithmic principles of remote ppg. IEEE Transactions on Biomedical Engineering **64**(7), 1479–1491 (2016)
39. Wei, C., Wang, W., Yang, W., Liu, J.: Deep retinex decomposition for low-light enhancement. In: British Machine Vision Conference (2018)
40. Yang, K., Long, N., Cao, K., Huang, B., Chen, Z., Zeng, Y., Tan, T., Shan, C., Jiang, J., Sun, Y.: Self-adaptive illumination enhancement for neonatal remote physiological measurement in nicus. IEEE Transactions on Consumer Electronics (2025)
41. Yu, Z., Shen, Y., Shi, J., Zhao, H., Torr, P.H., Zhao, G.: Physformer: Facial video-based physiological measurement with temporal difference transformer. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. pp. 4186–4196 (2022)
42. Zhang, K., Zhang, Z., Li, Z., Qiao, Y.: Joint face detection and alignment using multitask cascaded convolutional networks. IEEE signal processing letters **23**(10), 1499–1503 (2016)
43. Zou, B., Guo, Z., Chen, J., Zhuo, J., Huang, W., Ma, H.: Rhythmformer: Extracting patterned rppg signals based on periodic sparse attention. Pattern Recognition **164**, 111511 (2025)
44. Zou, B., Guo, Z., Hu, X., Ma, H.: Rhythmmamba: Fast, lightweight, and accurate remote physiological measurement. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 39, pp. 11077–11085 (2025)