Remote Photoplethysmography Based on a Skin Reflection Exponential Model

Jiachen Li[1], **Longzhen Tang**[1], **Shisheng Guo** [1,2*], **Guolong Cui**[1,2]
[1]University of Electronic Science and Technology of China
[2] Yangtze Delta Region Institute, University of Electronic Science and Technology of China, Quzhou
{ljc05265477@gmail.com, lzt97060@gmail.com, ssguo@uestc.edu.cn,
cuiguolong@uestc.edu.cn}

Abstract

Facial video–based remote photoplethysmography (rPPG) estimates physiological signals from subtle spatiotemporal variations in facial videos, but remains vulnerable to head motion, illumination changes, and environmental perturbations. Existing methods mainly distinguish rPPG-related variations from motion and lighting artifacts at the signal or feature level, while largely neglecting the optical reflection properties of human skin. In this work, we formulate a skin reflection exponential model to describe the nonlinear relationship among skin reflectance, illumination, motion-induced variations, and blood-volume changes. Guided by this model, we propose a physically inspired rPPG estimation framework that explicitly incorporates skin reflectance priors into signal extraction. Given a facial video, facial keypoints are first detected and encoded to represent motion dynamics, while illumination-related features and candidate rPPG representations are jointly modeled under the proposed exponential formulation. By constraining rPPG estimation with skin-reflection physics, the framework improves separation of physiological components from motion and illumination disturbances. Experiments demonstrate improved robustness under challenging conditions.

1 Introduction

Physiological indicators such as heart rate (HR), heart rate variability (HRV), and respiratory rate (RR) reflect the physiological state of the body and provide valuable information for clinical assessment [33]. Remote photoplethysmography (rPPG) aims to extract physiological signals from facial videos, enabling non-contact vital sign monitoring. With the widespread availability of camera-equipped mobile devices, rPPG has shown great potential in disease diagnosis, remote health monitoring, and driver fatigue detection [15]. The hemodynamic activity in facial blood vessels fluctuates periodically with cardiac contraction and relaxation, causing subtle changes in skin optical absorption. Although these changes are imperceptible to the naked eye, cameras can capture the resulting variations in incident, scattered, and reflected light within superficial skin tissues. Accordingly, rPPG estimates physiological signals by analyzing inter-frame color variations in facial videos [35, 34].

Early rPPG methods relied on handcrafted features for HR estimation [24, 6]. They typically performed color space correction based on prior assumptions, followed by blind source separation or skin reflectance modeling to extract rPPG signals from noise [35, 25]. However, assumptions such as uniform normalized skin tone under white light often fail in real-world scenarios, leading to degraded performance. With the development of deep learning, data-driven methods have become dominant for rPPG estimation [28, 14]. These methods learn latent periodic representations from facial videos

*corresponding author

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

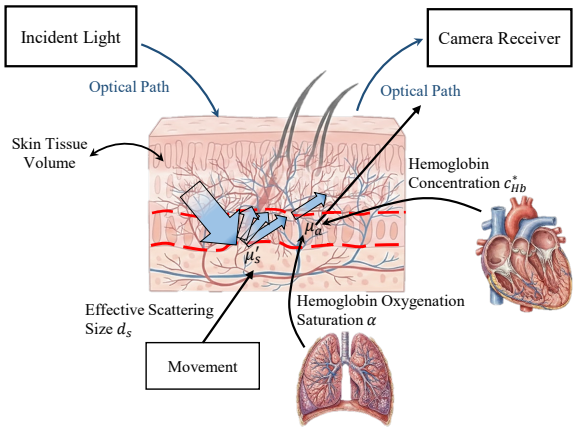


Figure 1: Schematic illustration of the visible light propagation path. Incident visible light reaches the skin surface and subsequently undergoes a series of optical interactions, including transmission, scattering, and absorption within the skin layers, before being reflected toward the camera. The intensity of the reflected light is determined by hemoglobin oxygenation saturation, hemoglobin concentration, the effective scattering size and so on.

and directly infer rPPG signals [3, 37], aiming to suppress illumination, motion, and environmental interference.

Most existing methods model lighting- and motion-induced skin color variations as additive specular components within dichromatic reflection models [35]. However, illumination changes and head motion affect not only the specular reflection component but also the diffuse reflection component. Therefore, suppressing only specular interference may leave residual noise in the diffuse component, reducing the accuracy and robustness of rPPG estimation.

In this paper, we propose a skin reflection exponential model that explicitly incorporates illumination variations and motion-induced perturbations, as shown in Figure 1. Unlike traditional reflection models that decompose skin color into specular and diffuse components, our model further describes light attenuation within skin tissues. Inspired by the reduced scattering and absorption coefficient model of human skin [40], we formulate skin reflectance through an exponential attenuation mechanism, where blood-volume changes, illumination variations, and motion disturbances jointly modulate facial appearance. This provides a physically meaningful basis for separating rPPG-related variations from non-physiological interference.

Inspired by the volumetric accumulation mechanism of NeRF [22], we develop a skin-reflection-guided framework that approximates light propagation and attenuation through a NeRF-inspired feature integration strategy. Guided by the proposed skin reflection exponential model, the framework jointly incorporates motion, illumination, and physiological representations to improve rPPG estimation. Experiments on five public datasets demonstrate its effectiveness and robustness. Our main contribution lies in introducing a physically inspired skin reflection model and integrating it into rPPG estimation. Nevertheless, the physiological factors modeled in this framework serve as physics-guided latent representations rather than direct measurements of physiological parameters. Accurate estimation of such parameters requires more explicit wavelength-dependent information, device calibration, and appropriate physiological supervision.

2 Related Work

2.1 Remote photoplethysmography

In recent years, many signal-processing methods based on the dichromatic reflection model have been developed for remote HR estimation [36, 6, 5]. By combining skin reflectance characteristics with prior knowledge of the skin color space, these methods normalize the skin color distribution to facilitate preliminary HR estimation [6, 35]. Some approaches further use blind source separation within the skin reflectance framework to disentangle specular and diffuse reflection components,

thereby decoupling rPPG signals from background interference [25]. With the development of deep learning, skin reflectance models have been increasingly integrated with neural networks to decouple illumination variations and motion information for more accurate rPPG reconstruction [5, 18]. Under the assumption that illumination and motion mainly affect the additive specular component, STmap-based methods guide models to focus on subtle chromatic variations in facial skin [18, 27]. Moreover, the periodic and frequency-domain properties of rPPG signals, together with facial semantic information, have been exploited to improve signal reconstruction [30, 41]. However, illumination and motion variations affect not only the specular component but also the diffuse reflection from skin tissues [40]. Motivated by this observation, we propose an improved skin reflection model to more comprehensively characterize the effects of illumination changes, head motion, and other perturbations on skin color.

2.2 NeRF-like Volumetric Rendering

Neural Radiance Fields (NeRF) represent scenes as continuous volumetric functions and synthesize novel views via differentiable volume rendering [22]. Along each camera ray, NeRF accumulates radiance contributions predicted at sampled points, weighted by opacity and transmittance, thereby providing a physically inspired formulation for modeling light propagation, attenuation, occlusion, and appearance formation [22]. NeRF and related neural rendering techniques have been widely extended to facial video synthesis and dynamic appearance modeling, including facial style transfer [16], talking face generation [1], adaptive 3D scene editing [9], 3D object segmentation [19], multi-view perceptual restoration [23], and high-fidelity talking portrait synthesis [13]. These works demonstrate the effectiveness of NeRF-like volumetric representations for modeling complex appearance variations and view-dependent effects. In rPPG estimation, facial color formation can also be viewed as a light transport process, where incident light interacts with layered skin tissues and is jointly affected by absorption, reduced scattering, illumination variations, and motion-induced perturbations. This process is structurally related to NeRF volume rendering, since both involve layer-wise contribution accumulation modulated by attenuation terms. Therefore, NeRF-like transmittance-weighted accumulation provides a useful analogy for modeling the exponential attenuation behavior of skin reflection, motivating rPPG estimation under an improved skin reflection model with NeRF-like volumetric modeling.

3 Method

3.1 Improved Skin Reflection Model

We first establish the relationship between blood flow, hemoglobin, melanin, effective scattering area, reduced scattering coefficient, and absorption coefficient proposed by [40]. Its mathematical expression can be written as

$$\mu_a(\lambda,t) = c^*_{Hb}(t)(\alpha(t)\varepsilon_{HbO_2}(\lambda(t),t) + (1-\alpha(t))\varepsilon_{Hb}(\lambda(t),t)) + C(\lambda,t) \tag{1}$$

$$\mu_s^{'}(\lambda,t) = (1 - f(\lambda(t),t))\mu_s^{'}(\lambda_{min}) \tag{2}$$

where $\mu_a(\lambda,t)$ and $\mu_s^{'}(\lambda,t)$ denote the absorption coefficient and reduced scattering coefficient, respectively. hemoglobin concentration $c^*_{Hb}(t)$, hemoglobin oxygen saturation $\alpha(t)$, melanin concentration $C(\lambda,t) = c_{mel}\varepsilon_{mel}(\lambda,t)$ are included in equation (1) to reflect changes in absorption coefficient. And $f(\lambda(t),t) = \frac{d_0^{1/2}}{d_s^{1/2}}\frac{\lambda-\lambda_{min}}{\lambda_{max}-\lambda_{min}}$, where $d_s^{1/2}$ is effective scattering size to reflect the variation of the reduced scattering coefficient, and $d_0 = 0.0625\mu m$ is a constant. And λ is the light wavelength. Within a sufficiently short observation window, variations in the target's hemoglobin concentration and melanin concentration can be reasonably neglected. Meanwhile, the movement of the human head can cause changes in the effective scattering area. Therefore, the reduced scattering coefficient can be considered time dependent. Meanwhile, it is worth noting that respiration induces periodic variations in hemoglobin oxygen saturation $\alpha(t)$, whereas cardiac activity leads to periodic fluctuations in the concentrations of oxyhemoglobin ε_{HbO_2} and deoxyhemoglobin ε_{Hb}. In conjunction with the temporal variation of visible illumination, these physiological factors jointly contribute to the change of the absorption coefficient. Subsequently, to simplify the entire propagation model,

we assume that the light source is located on the deep skin of the human face, and combine it with the Beer-Lambert law [31] to provide the absorption and scattering of light by the human skin

$$I(s) = \int_0^s \mu_s^{'} I_{\mu_s^{'}}(t) e^{-\int_0^s (\mu_s^{'}(z)+\mu_a(z))dz} dt \tag{3}$$

The conversion of the above equation to Riemann summation form is

$$I(s) = \sum_{i=1}^{N} \frac{\mu_s^{'}}{\mu_s^{'}+\mu_a} c_i T_i (1 - e^{-(\mu_s^{'}+\mu_a)\delta_i}) \tag{4}$$

where $T_i = e^{-\sum_{j=0}^{i-1}(\mu_s^{'}+\mu_a)\delta_i}$. By ignoring camera thermal noise and background illumination, equation (4) represents the final form of the derived ideal skin transmittance model. Notably, this formulation exhibits a high degree of similarity to the volume rendering equation. Motivated by this correspondence, NeRF is employed to guide the estimation of rPPG signals.

3.2 NeR-rPPG

As shown in Figure 2, NeR-rPPG includes two Stem modules, one MovE module, one LigM module, one NerM module, one ResM module and one rPPM module. Assuming the input facial video is $X \in \mathbb{R}^{C\times T\times H\times W}$, it passes through two Stem modules to obtain $X_n \in \mathbb{R}^{C\times T\times \frac{H}{8}\times \frac{W}{8}}$ and $X_r \in \mathbb{R}^{C\times T\times \frac{H}{8}\times \frac{W}{8}}$, respectively. Subsequently, X_n is fed into the MovE module and the LigM module to estimate the motion information and illumination characteristics of the target video, respectively. The extracted motion information, illumination information, and positional encoding are then jointly utilized to predict the color attributes of the target and the density of the sampled points. Based on the estimated rPPG signals, the ResM models the relationship between hemoglobin oxygen saturation and hemoglobin concentration. The NerM module further generates the color and density representations of the target, which are subsequently combined with equation (4) to synthesize the reconstructed target video. In parallel, X_r is sent to the rPPM module and performs rPPG signal estimation based on equation (4).

3.2.1 MovE

The MovE module is designed to capture motion-induced skin reflectance perturbations by jointly leveraging facial keypoints and video features. Rather than explicitly modeling head motion geometry, MovE encodes motion-related appearance variations that affect the effective scattering size and, consequently, the reduced scattering coefficient. Specifically, Dlib[4] is used for facial 81 keypoints detection at first, and then the keypoints K are differentiated along the time dimension to obtain the dynamic change information of the target $\mathcal{K} = \frac{k_{i+1}-k_i}{\delta t}, k_i \in K$. Meanwhile, embedding the obtained differential keypoints $\mathcal{K}$ with X_n, we can obtain $\mathcal{K}_e \in \mathbb{R}^{T\times d\times 2}$, $X_e \in \mathbb{R}^{d\times T\times \frac{H}{32}\times \frac{W}{32}}$. Subsequently, the cross attention module is used to fuse spatial information between $\mathcal{K}_e$ and X_e to obtain X_f, while preserving the temporal dimension unchanged. Finally, X_f is subjected to fast Fourier transform (FFT) and FeedForward to capture global motion changes of the target X_m. After the above modules, the obtained target motion information is considered to effectively reflect the changes in the effective scattering size in the skin caused by the target motion. Therefore, we can obtain $\frac{d_0^{1/2}}{d_s^{1/2}} \sim Linear(X_m) = \mathcal{X}_m$.

3.2.2 LigM

The LigM module is designed to model illumination variations in facial videos by capturing temporal changes in light characteristics under natural illumination dynamics. Built upon a U-Net architecture [26], LigM uses convolutional operations to estimate illumination information from the input data X_n. Since pixels within each video frame can be regarded as time-dependent observations from a common distribution, spatial differential convolution is introduced to capture potential illumination changes by computing pixel-wise gradients in the horizontal and vertical directions. The U-Net structure further integrates global context with local details, enabling fine-grained illumination estimation for each pixel. Specifically, X_n first undergoes spatial differential convolution, we can get

$$X_d = \alpha Conv(\frac{X_{i,j+1,k} - X_{i,j,k}}{\delta x}) + \beta Conv(\frac{X_{i,j,k+1} - X_{i,j,k}}{\delta y}) \tag{5}$$

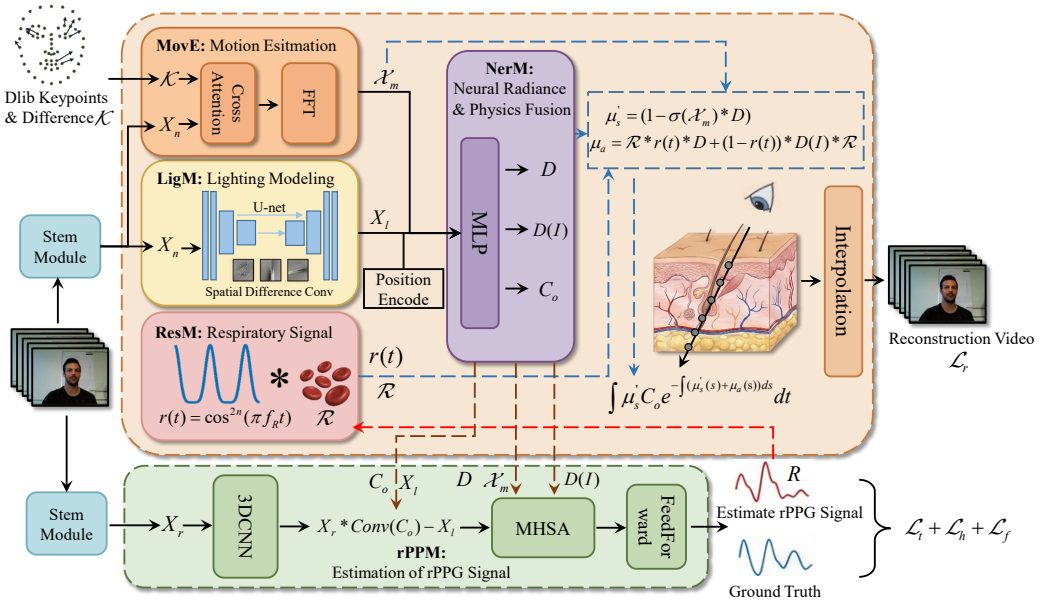


Figure 2: Overall framework of NeR-rPPG. **MovE** combines input video and keypoints to estimate the motion state of the target $\mathcal{X}_m \sim \frac{d_0^{1/2}}{d_s^{1/2}}$ in equation (2). **LigM** utilizes spatial difference convolution and U-Net to extract the illumination information from the input video X_l. **ResM** is used to generate potential respiratory signals of the target, thereby simulating changes in hemoglobin oxygen saturation $r(t) \sim \alpha(t)$ and changes in hemoglobin concentration $\mathcal{R} \sim c_{Hb}^*$. **NerM** is used to estimate target color $C_o \sim c_i$ in equation (4) and density $D \sim \varepsilon_{HbO_2}$, $D(I) \sim \varepsilon_{Hb}$, and reconstruct videos by combining with an improved reflection model. **rPPM** is used to estimate the rPPG signal of the target R.

where $Conv$ denotes the 3DCNN layer, α and β are learnable parameters to control gradient weights in different directions. i, j and k denote the i-th frame, j-th column and k-th row. Subsequently, the input X_d is processed through a normalization layer and activation function followed by a max-pooling. Multiple modules arranged in the aforementioned order are stacked to facilitate the extraction of high level feature representations. The upsampling stage is then implemented using the standard U-Net decoder to obtain light information X_l.

3.2.3 ResM

The ResM module generates the changes in hemoglobin oxygenation saturation caused by target breathing and the changes in hemoglobin concentration caused by heartbeat based on equation (1). In detail, assume the estimated rPPG signal is denoted as R, and the respiratory signal as r. The R is obtained from the rPPM module, while $r(t)$ can be derived from the following equation [7]

$$r(t) = cos^{2n}(\pi f_R t), n \in Z^+ \tag{6}$$

where f_R is the respiration frequency and n is a scalar parameter defining the waveform type. Equation 6 can effectively simulate the short period of stationary moment after lung deflation. f_R is the learnable parameter with MLP processing to estimate the respiratory frequency of target. To simplify the model, we equate $r(t)$ to $\alpha(t)$. It is worth noting that, due to the structure of human skin, including the stratum corneum and dermis, the absorption and scattering coefficients can be reasonably assumed to remain constant over time when visible light propagates through the stratum corneum. Based on this assumption, we simulate the corresponding optical behavior by introducing a constant to the sampling points along the visible light propagation path. Then, we can get $\mathcal{R} = C(R_i, a), i \leq N$, where R_i is the rPPG signal of i-th sample point, N represents the number of sample points and C is the concatenation function to concatenate along the sampling point dimension.

3.2.4 NerM

The NerM module is designed to render using the outputs of the aforementioned modules according to equations (1), (2) and (4). Firstly, we concatenate the outputs of the LigM and MovE modules with positional encodings, and then feed them into an MLP to generate density information for different substances as well as color information for the target. Its mathematical form can be expressed as

$$D, D_I, C_o = MLP(Pos, X_l, \mathcal{X}_m) \tag{7}$$

where D is the density of Oxyhemoglobin, D_I denote the density of Deoxyhemoglobin and C_o represents the color information of targets. Combine $\mathcal{X}_m$, D, D_I, C_o, $r(t)$ and $\mathcal{R}$, we have

$$\widetilde{\mu_a} = \mathcal{R} * r(t) * D + (1 - r(t)) * D_I * \mathcal{R} \tag{8}$$

$$\widetilde{\mu_s^{'}} = (1 - \sigma(\mathcal{X}_m) * D). \tag{9}$$

Subsequently, substitute equations (8) and (9) into equation (4), and replace c_i in equation (4) with C_o, we can successfully render the target facial videos $\mathcal{X}_o$ with interpolation.

3.2.5 rPPM

Based on Equations (11), (12), and (6), we constructed rPPM for rPPG signal estimation. Specifically, we first embedding the X_r, C_o, D, D_I, $\mathcal{X}_m$, and X_l. Subsequently, we use X_r and C_o to achieve preliminary decoupling $X_k = X_r * \mathcal{C}(C_o) - X_l$, where $\mathcal{C}$ denotes the 3DCNN layer with a GroupNorm and GELU activation function. Then, we utilize the X_k and 3DCNN to generate Query X_Q, Key X_K and Value X_V, and employ an improved MHSA for feature decoupling and extraction

$$X_s = softmax(\frac{X_Q X_K^T - X_Q \widetilde{\mu_s^{'}}}{\sqrt{d_k}}) X_V. \tag{10}$$

To further decouple X_s, we can have

$$X_o = X_s(r(t)(D - D_I) + D_I) \tag{11}$$

Then we feed X_o into the vanilla FeedForward module for nonlinearity. Meanwhil, we extract high dimensional features related to the rPPG signal from the input feature values by stacking the above modules. Eventually, we follow Yu et al[37] to use the same rPPG Generator module to generate the final rPPG signal R.

3.3 Loss Function

We follow [37] and [41] to use both time domain loss and frequency domain loss to constrain the estimated rPPG signal. For the time domain loss, we follow [20] and [39] in using the negative Pearson correlation coefficient as the loss function

$$\mathcal{L}_t = 1 - \frac{\sum_{i=1}^{l}(R_i - \bar{R})(r_i - \bar{r})}{\sqrt{\sum_{i=1}^{l}(R_i - \bar{R})^2}\sqrt{\sum_{i=1}^{l}(r_i - \bar{r})^2}} \tag{12}$$

where R_i and r_i are the i-th sample point of estimate rPPG signal and ground truth rPPG signal. $\bar{R}$ and $\bar{r}$ are the average value of estimated rPPG signal and ground truth rPPG signal. We follow [37] using the cross entropy loss to locate the heartbeat frequency component

$$\mathcal{L}_h = CE(maxIndex(PSD(R)), PSD(r)) \tag{13}$$

where PSD is Power Spectral density and $maxIndex$ denotes the index of the maximum value. Moreover, to further constrain the periodic characteristics of the estimated rPPG signal, we use negative Pearson coefficients in the frequency domain for additional periodic constraints $\mathcal{L}_f = \mathcal{L}_t(PSD(R), PSD(r))$. Meanwhile, due to the extremely weak rPPG signal, the video generated by NeR-rPPG rendering should closely resemble the original video. Therefore, we employ mean squared error loss to estimate the error between the rendered video and the original video.

$$\mathcal{L}_r = \| \mathcal{X}_o - X^{'} \|_2 . \tag{14}$$

The $X^{'}$ is the original video without any normalization. Then, the final loss can be written as

$$\mathcal{L}_{overall} = \mathcal{L}_r + \mathcal{L}_f + \mathcal{L}_h + \mathcal{L}_t \tag{15}$$

Table 1: Intra-dataset evaluation. Best results are marked in **bold** and second best in underline.

Methods	UBFC-rPPG			PURE			COHFACE			DEAP			MMPD		
	MAE	RMSE	ρ	MAE	RMSE	ρ	MAE	RMSE	ρ	MAE	RMSE	ρ	MAE	RMSE	ρ
GEERN	7.50	11.41	0.82	7.23	12.05	0.80	7.90	15.67	0.79	8.10	11.17	0.80	16.72	23.79	0.57
ICA	5.42	10.47	0.82	5.66	11.15	0.86	7.61	14.65	0.80	7.5	10.78	0.82	15.90	19.73	0.61
CHROM	5.20	9.92	0.83	5.82	10.80	0.86	7.80	12.45	0.80	7.47	10.31	0.82	15.63	19.11	0.62
DeepPhys	4.27	7.82	0.90	1.82	2.71	0.99	6.89	13.89	0.87	6.90	11.43	0.86	21.75	28.01	0.21
PhysNet	2.95	3.67	0.97	2.10	2.60	0.99	5.38	10.76	0.87	6.61	12.03	0.86	5.90	9.83	0.87
Meta-rPPG	2.54	4.98	0.98	2.52	4.63	0.98	3.46	7.71	0.95	5.16	9.51	0.89	-	-	-
Gideon et al.	1.85	4.28	0.98	2.23	2.97	0.99	3.6	5.6	0.95	5.13	8.16	0.89	-	-	-
Dual-GAN	0.44	**0.71**	0.99	0.82	1.31	0.99	3.26	7.29	0.96	4.28	8.11	0.91	-	-	-
Physformer	1.11	2.27	0.98	1.17	2.44	0.99	2.44	3.77	0.98	4.12	7.91	0.92	5.28	8.46	0.88
Contrast-Phys	0.64	1.00	0.99	1.00	1.40	0.99	-	-	-	-	-	-	-	-	-
Physformer++	0.69	1.03	0.99	0.70	1.22	0.99	0.74	2.34	0.99	3.03	3.96	0.96	-	-	-
TFA-PFE	0.76	1.62	0.99	1.44	2.50	0.98	1.31	3.92	0.98	-	-	-	-	-	-
MaskFusionNet	0.93	1.80	0.99	1.11	1.39	0.99	-	-	-	-	-	-	-	-	-
Rhythmformer	0.77	1.78	0.99	0.84	1.90	0.99	1.47	2.87	0.99	3.97	5.73	0.94	3.48	7.83	0.93
PhysMamba	0.52	1.13	0.99	0.86	1.38	0.99	0.95	1.81	0.99	3.65	5.02	0.95	4.71	8.34	0.90
RhythmMamba	0.62	1.37	0.99	0.69	0.97	0.99	0.87	1.73	0.99	3.32	4.87	0.95	3.73	7.77	0.92
Ours	**0.42**	0.98	0.99	**0.56**	**0.85**	0.99	**0.60**	**1.51**	0.99	**2.78**	**3.75**	0.96	**3.33**	**7.13**	0.93

4 Experiments

4.1 Datasets and Performance Metrics

The experiments of remote physiological measurement were conducted on five publicly available datasets: UBFC-rPPG[2], PURE[29], COHFACE[10], DEAP[11], MMPD[32]. **UBFC-rPPG** dataset contains 42 videos and corresponding recorded PPG signals. This dataset collected the heartbeat and videos of 42 health subjects playing a time sensitive mathematical game under varying amounts of ambient light. **PURE** dataset recorded 60 facial videos of 10 health subjects with six different activities. **COHFACE** dataset records facial videos and PPG signals of 40 health subjects under different natural lighting conditions. **DEAP** dataset consists of 874 facial videos associated with multi channel physiological signals. They were taken from 22 subjects by playing one minute musical excerpts to them. **MMPD** dataset consists of 660 one minute facial videos collected from 33 participants. The recordings encompass four activity scenarios (static, head rotation, speaking, and walking) and four illumination conditions (high-intensity LED, low-intensity LED, incandescent, and natural lighting). All videos were recorded at a resolution of 320 × 240 pixels and a frame rate of 30 fps. **Metrics**.This study uses the following metrics to evaluate the video level HR of model. Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Pearson Correlation Coefficient (ρ).

4.2 Implementation Details

For facial videos from five datasets, We divided the training set into validation set and test set based on 70%, 10%, and 20% of the subjects. Meanwhile, The extraction of Region of Interes (ROI) of Facial video processing is achieved by the utilization of the Dlib library. we scale the sliced video data to a resolution of 128×128. For the rPPG waveform, a second order Butterworth filter (cutoff frequencies:0.75 and 2.5Hz) was applied to filter the rPPG waveform, and power spectral density was computed by the Welch algorithm for further HR estimation. All experiments were conducted on an NVIDA RTX 4090. The AdamW optimizer is used and the learning rate is set as 1e-4 with OneCycle learning rate strategy. The weight decay is 5e-4 and the models are trained for 100 epochs on UBFC, PURE and COHFACE. While the model testing on DEAP is trained for 30 epochs. In the Intra-Dataset Testing, we compared our proposed method with the following methods: **GEERN**[24], **ICA**[34], **CHROM**[6], **DeepPhys**[5], **PhysNet**[36], **Meta-rPPG**[12], **Gideon et al.**[8], **PulseGan**[28], **Dual-Gan**[20], **Contrast-Phys**[30], **Physformer**[37], **TFA-PFE**[14], **Physformer++**[38], **MaskFusion-Net**[39], **Rhythmformer**[41], **PhysMamba**[21], **RhythmMamba**[42]. While in the Cross-Dataset Testing,we compared our proposed method with the following methods: **DeepPhys**[5], **PhysNet**[36], **Physformer**[37], **EfficientPhys**[17], **Rhythmformer**[41], **Meta-rPPG**[12]. All experiments were conducted under the same experimental conditions.

4.3 Intra-dataset Testing

To verify the effectiveness of the proposed method, we conducted validation on five datasets, and the results are presented in Table 1. The experimental results indicate that the proposed method achieves

competitive performance compared with existing state-of-the-art approaches. This performance gain can be largely attributed to the explicit consideration of target motion and illumination variations and their effects on skin color, which provides an implicit regularization for the model. It is also observed that the accuracy on the COHFACE and DEAP datasets is lower than that on the UBFC and PURE datasets. This performance difference is likely due to the more complex data acquisition environments and larger data scales of COHFACE and DEAP, which pose greater challenges for robust physiological signal estimation.

4.4 Cross-dataset Testing

We follow the approach [41] of conducting cross dataset testing. The model was trained on the PURE, UBFC rPPG, and COHFACE datasets, and then tested on the PURE, UBFC rPPG, and COHFACE datasets. The training dataset is divided into two parts, with 80% used for training and the remaining 20% used for validation. And we compare the obtained results with the SOTA method as shown in the Table 2 and Table 3. The results demonstrate that the proposed method achieves

Table 2: Cross-dataset testing results. P $\rightarrow$ U means traning on PURE and test on UBFC.

Method	P $\rightarrow$ U		U $\rightarrow$ P	
	MAE$\downarrow$	RMSE$\downarrow$	MAE$\downarrow$	RMSE$\downarrow$
DeepPhys	3.21	6.09	5.95	17.29
PhysNet	2.79	5.32	8.81	19.85
Physformer	1.87	3.68	12.92	24.36
EfficientPhys	4.07	7.45	6.92	17.93
Rhythmformer	1.89	3.92	3.84	10.24
Meta-rPPG	2.54	4.57	7.89	18.72
ours	**1.69**	**3.64**	**3.76**	**9.39**

competitive performance, suggesting its ability to effectively capture target rPPG-related features while mitigating the influence of irrelevant factors. Furthermore, it is observed that the cross-dataset evaluation performance between PURE and COHFACE is inferior to that between PURE and UBFC. This performance can be attributed to the significant discrepancy in data distributions across the datasets, as COHFACE contains more complex acquisition conditions, including both natural and non-natural illumination.

Table 3: Cross-dataset testing results. P $\rightarrow$ C means traning on PURE and test on COHFACE.

Method	P $\rightarrow$ C		C $\rightarrow$ P	
	MAE$\downarrow$	RMSE$\downarrow$	MAE$\downarrow$	RMSE$\downarrow$
DeepPhys	11.09	21.77	13.02	19.59
PhysNet	10.15	18.23	12.88	18.52
Physformer	9.87	16.72	10.63	17.12
EfficientPhys	16.43	24.23	13.92	21.46
Rhythmformer	10.48	17.76	12.84	17.24
Meta-rPPG	13.18	19.28	14.08	19.70
ours	**9.33**	**15.46**	**9.12**	**14.39**

4.5 Ablation Study

4.5.1 Key Module Impact

We conducted ablation experiments on the modules in NeR-rPPG. Specifically, by replacing certain modules in NeR-rPPG with a linear layer, we effectively eliminated those modules. The result can be seen in Table 4. The results in the table show that when all modules are eliminated, the performance of the NeR rPPG model decreases. It is worth noting that the performance degradation of NeR rPPG is minimal when the ResM module is eliminated. This is because ResM is mainly used to simulate

Table 4: Ablation experiment results of Key Modules. "w/o" means "without".

Key Modules	PURE		COHFACE	
	MAE	RMSE	MAE	RMSE
w/o MovE	1.29	3.22	1.46	3.74
w/o LigM	1.23	3.35	1.52	3.77
w/o ResM	0.68	1.81	0.89	2.53
w/o NerM	1.17	2.63	1.82	3.95

changes in blood oxygen concentration caused by respiration, and compared to rPPG signal changes, the frequency is lower, so the impact will be smaller.

4.5.2 Loss Function Impact

We conducted ablation experiments on the proposed loss function, and the results are presented in the Table 5. As shown in the table, removing the $\mathcal{L}_t$ loss leads to the most pronounced performance

Table 5: Ablation experiment results of loss function

$\mathcal{L}_t$	$\mathcal{L}_h$	$\mathcal{L}_f$	$\mathcal{L}_r$	PURE		COHFACE	
				MAE	RMSE	MAE	RMSE
×	✓	✓	✓	1.56	3.84	1.79	3.83
✓	×	✓	✓	1.34	3.56	1.67	3.11
✓	✓	×	✓	0.62	1.89	0.72	2.10
✓	✓	✓	×	1.42	3.77	1.52	3.24

degradation in NeR-rPPG. This behavior can be attributed to the sparsity of periodic signals in the frequency domain, which makes it challenging for the model to converge to an optimal solution in the absence of explicit time domain constraints. When the $\mathcal{L}_r$ loss is eliminated, NeR-rPPG still exhibits performance degradation, which is due to the fact that the $\mathcal{L}_r$ loss can guide the rendering of NeR-rPPG, thereby constraining the direction of gradient descent.

4.5.3 Impact of the Input of rPPM

To evaluate the effectiveness of rPPM input in reconstructing rPPG signals, we conducted ablation experiments on rPPM input and the results are shown in Table 6. The result shows that when the

Table 6: Ablation experiment results of the inputs of rPPM.

Inputs	PURE		COHFACE	
	MAE	RMSE	MAE	RMSE
w/o C_o	1.32	2.13	1.34	2.74
w/o X_l	0.89	1.79	1.12	2.63
w/o D and $\mathcal{X}_m$	2.03	3.18	2.65	4.58
w/o D and $D(I)$	2.14	3.63	2.04	4.12

input of rPPM is eliminated, the performance of NeR-rPPG decreases to varying degrees. This is because the input of the aforementioned variables can provide prior information for rPPM.

4.6 Complexity Analysis

To evaluate the computational cost of the proposed physics-guided framework, we compare NeR-rPPG with representative rPPG methods in terms of model parameters, FLOPs, inference time, and GPU memory consumption. The results are summarized in Table 7.

As shown in Table 7, NeR-rPPG achieves competitive estimation performance with moderate computational complexity and GPU memory consumption, despite having more parameters. Although its

Table 7: Computational complexity comparison of different rPPG methods.

Method	Parameters (M)	FLOPs (G)	Inference Time (ms)	GPU Memory (MB)
DeepPhys	1.98	168.01	77.82	37.28
PhysNet	0.77	98.17	61.77	11.43
PhysFormer	7.38	75.91	51.06	28.63
EfficientPhys	1.91	84.97	45.09	26.68
RhythmFormer	3.25	57.71	35.73	29.06
NeR-rPPG	11.26	82.11	64.92	23.05

inference time is relatively higher, our primary contribution lies in introducing a novel physics-guided framework for rPPG estimation rather than optimizing computational efficiency.

4.7 Additional Robustness Experiments under Illumination Variations

To further evaluate robustness against illumination variations, we conduct test time perturbation experiments without modifying the trained models. Two types of illumination perturbations are introduced. **Temporal Color Drift**, which applies gradually varying channel-wise RGB gains to simulate dynamic white-balance and color temperature changes; and **Spatial Gradient Illumination**, which introduces spatially varying brightness to simulate non-uniform facial illumination. NeR-rPPG and the baseline methods are evaluated under identical perturbation settings.

Table 8: Robustness comparison under different illumination perturbations on the UBFC-rPPG and PURE datasets.

Perturbation	Method	UBFC-rPPG		PURE	
		MAE ↓	RMSE ↓	MAE ↓	RMSE ↓
Spatial Gradient Illumination	PhysMamba	**1.99**	**5.78**	1.74	5.94
	RhythmMamba	3.02	7.37	2.39	7.82
	NeR-rPPG	2.07	6.12	**1.39**	**3.75**
Temporal Color Drift	PhysMamba	17.56	21.80	11.63	15.90
	RhythmMamba	20.73	26.31	18.19	26.51
	NeR-rPPG	**12.83**	**18.25**	**7.48**	**12.21**

As shown in Table 8, NeR-rPPG demonstrates competitive robustness under both illumination perturbations. Under spatial gradient illumination, NeR-rPPG achieves the best performance on PURE and comparable results to PhysMamba on UBFC-rPPG. Notably, under temporal color drift, NeR-rPPG consistently outperforms both baselines across all metrics. These results suggest that the proposed skin reflection model improves robustness to illumination variations, particularly under dynamic color changes.

5 Conclusion

In this study, we propose NeR-rPPG, a skin-reflection-exponential-model-guided framework for rPPG estimation. Inspired by NeRF-like volumetric rendering, NeR-rPPG uses transmittance-weighted accumulation to approximate light propagation and attenuation in layered skin tissues, while jointly modeling motion, illumination, and physiological representations for more robust rPPG extraction. Extensive experiments demonstrate its effectiveness. Nevertheless, the framework still has several limitations. The assumptions regarding respiration and melanin are relatively simplified, which may limit the accurate modeling of individual skin properties. In addition, the need to render video sequences introduces considerable computational overhead.

6 Acknowledgements

This work was supported by the Municipal Government of Quzhou (2024D050, 2024D047).

References

[1] Chongke Bi, Xiaoxing Liu, and Zhilei Liu. Nerf-ad: Neural radiance field with attention-based disentanglement for talking face synthesis. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 3490–3494. IEEE, 2024.

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

[3] Deivid Botina-Monsalve, Yannick Benezeth, and Johel Miteran. Rtrppg: An ultra light 3dcnn for real-time remote photoplethysmography. In *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, pages 2146–2154, 2022.

[4] Nataliya Boyko, Oleg Basystiuk, and Nataliya Shakhovska. Performance evaluation and comparison of software for face recognition, based on dlib and opencv library. In *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, pages 478–482. IEEE, 2018.

[5] Weixuan Chen and Daniel McDuff. Deepphys: Video-based physiological measurement using convolutional attention networks. In *Proc. Eur. Conf. Comput. Vis.*, pages 349–365, 2018.

[6] Gerard De Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Trans. Biomed. Eng.*, 60(10):2878–2886, Jun. 2013.

[7] Shuqin Dong, Yuchen Li, Jingyun Lu, Zhi Zhang, Changzhan Gu, and Junfa Mao. Accurate detection of doppler cardiograms with a parameterized respiratory filter technique using a k-band radar sensor. *IEEE Transactions on Microwave Theory and Techniques*, 71(1):71–82, 2022.

[8] John Gideon and Simon Stent. The way to my heart is through contrastive learning: Remote photoplethysmography from unlabelled video. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3995–4004, 2021.

[9] Runze He, Shaofei Huang, Xuecheng Nie, Tianrui Hui, Luoqi Liu, Jiao Dai, Jizhong Han, Guanbin Li, and Si Liu. Customize your nerf: Adaptive source driven 3d scene editing via local-global iterative training. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6966–6975, 2024.

[10] Guillaume Heusch, André Anjos, and Sébastien Marcel. A reproducible study on remote heart rate measurement. *arXiv preprint arXiv:1709.00962*, 2017.

[11] Sander Koelstra, Christian Muhl, Mohammad Soleymani, Jong-Seok Lee, Ashkan Yazdani, Touradj Ebrahimi, Thierry Pun, Anton Nijholt, and Ioannis Patras. Deap: A database for emotion analysis; using physiological signals. *IEEE transactions on affective computing*, 3(1): 18–31, 2011.

[12] Eugene Lee, Evan Chen, and Chen-Yi Lee. Meta-rppg: Remote heart rate estimation using a transductive meta-learner. In *European conference on computer vision*, pages 392–409. Springer, 2020.

[13] Jiahe Li, Jiawei Zhang, Xiao Bai, Jun Zhou, and Lin Gu. Efficient region-aware neural radiance fields for high-fidelity talking portrait synthesis. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7568–7578, 2023.

[14] Jianwei Li, Zitong Yu, and Jingang Shi. Learning motion-robust remote photoplethysmography through arbitrary resolution videos. In *Proc. AAAI Conf. Artif. Intell.*, volume 37, pages 1334–1342, 2023.

[15] Wan-Hua Lin, Dan Wu, Chunyue Li, Heye Zhang, and Yuan-Ting Zhang. Comparison of heart rate variability from ppg with that from ecg. In *The International Conference on Health Informatics: ICHI 2013, Vilamoura, Portugal on 7-9 November, 2013*, pages 213–215. Springer, 2014.

[16] Kunhao Liu, Fangneng Zhan, Yiwen Chen, Jiahui Zhang, Yingchen Yu, Abdulmotaleb El Saddik, Shijian Lu, and Eric P Xing. Stylerf: Zero-shot 3d style transfer of neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 8338–8348, 2023.

[17] Xin Liu, Brian Hill, Ziheng Jiang, Shwetak Patel, and Daniel McDuff. Efficientphys: Enabling simple, fast and accurate camera-based cardiac measurement. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 5008–5017, 2023.

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

[19] Yichen Liu, Benran Hu, Chi-Keung Tang, and Yu-Wing Tai. Sanerf-hq: Segment anything for nerf in high quality. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3216–3226, 2024.

[20] Hao Lu, Hu Han, and S Kevin Zhou. Dual-gan: Joint bvp and noise modeling for remote physiological measurement. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 12404–12413, 2021.

[21] Chaoqi Luo, Yiping Xie, and Zitong Yu. Physmamba: Efficient remote physiological measurement with slowfast temporal difference mamba. In *Chinese Conference on Biometric Recognition*, pages 248–259. Springer, 2024.

[22] Ben Mildenhall, Pratul P Srinivasan, Matthew Tancik, Jonathan T Barron, Ravi Ramamoorthi, and Ren Ng. Nerf: Representing scenes as neural radiance fields for view synthesis. *Communications of the ACM*, 65(1):99–106, 2021.

[23] Ashkan Mirzaei, Tristan Aumentado-Armstrong, Konstantinos G Derpanis, Jonathan Kelly, Marcus A Brubaker, Igor Gilitschenski, and Alex Levinshtein. Spin-nerf: Multiview segmentation and perceptual inpainting with neural radiance fields. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20669–20679, 2023.

[24] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Exp.*, 18(10): 10762–10774, May 2010.

[25] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Trans. Biomed. Eng.*, 58(1): 7–11, Oct. 2010.

[26] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[27] Hang Shao, Lei Luo, Jianjun Qian, Mengkai Yan, Shuo Chen, and Jian Yang. Remote photoplethysmography in real-world and extreme lighting scenarios. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 10858–10867, 2025.

[28] Rencheng Song, Huan Chen, Juan Cheng, Chang Li, Yu Liu, and Xun Chen. Pulsegan: Learning to generate realistic pulse waveforms in remote photoplethysmography. *IEEE J. Biomed. Health Inform.*, 25(5):1373–1384, Jan. 2021.

[29] Ronny Stricker, Steffen Müller, and Horst-Michael Gross. Non-contact video-based pulse rate measurement on a mobile service robot. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 1056–1062. IEEE, 2014.

[30] Zhaodong Sun and Xiaobai Li. Contrast-phys: Unsupervised video-based remote physiological measurement via spatiotemporal contrast. In *European Conference on Computer Vision*, pages 492–510. Springer, 2022.

[31] Donald F Swinehart. The beer-lambert law. *Journal of chemical education*, 39(7):333, 1962.

[32] Jiankai Tang, Kequan Chen, Yuntao Wang, Yuanchun Shi, Shwetak Patel, Daniel McDuff, and Xin Liu. Mmpd: Multi-domain mobile video physiology dataset. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, pages 1–5. IEEE, 2023.

[33] Andriy Temko. Accurate heart rate monitoring during physical exercises using ppg. *IEEE Trans. Biomed. Eng.*, 64(9):2016–2024, Mar. 2017.

[34] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Opt. Express*, 16(26):21434–21445, Dec 2008.

[35] Wenjin Wang, Albertus C Den Brinker, Sander Stuijk, and Gerard De Haan. Algorithmic principles of remote ppg. *IEEE Trans. Biomed. Eng.*, 64(7):1479–1491, Sep. 2016.

[36] Zitong Yu, Xiaobai Li, and Guoying Zhao. Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks. *arXiv preprint arXiv:1905.02419*, 2019.

[37] Zitong Yu, Yuming Shen, Jingang Shi, Hengshuang Zhao, Philip HS Torr, and Guoying Zhao. Physformer: Facial video-based physiological measurement with temporal difference transformer. In *Proc. IEEE/CVF Cof. Comput. Vis Pattern Recognit.*, pages 4186–4196, 2022.

[38] Zitong Yu, Yuming Shen, Jingang Shi, Hengshuang Zhao, Yawen Cui, Jiehua Zhang, Philip Torr, and Guoying Zhao. Physformer++: Facial video-based physiological measurement with slowfast temporal difference transformer. *International Journal of Computer Vision*, 131(6): 1307–1330, 2023.

[39] Yizhu Zhang, Jingang Shi, Jiayin Wang, Yuan Zong, Wenming Zheng, and Guoying Zhao. Maskfusionnet: A dual-stream fusion model with masked pre-training mechanism for rppg measurement. *IEEE Transactions on Circuits and Systems for Video Technology*, 34(11): 11521–11534, 2024.

[40] George Zonios and Aikaterini Dimou. Modeling diffuse reflectance from semi-infinite turbid media: application to the study of skin optical properties. *Optics express*, 14(19):8661–8674, 2006.

[41] Bochao Zou, Zizheng Guo, Jiansheng Chen, and Huimin Ma. Rhythmformer: Extracting rppg signals based on hierarchical temporal periodic transformer. *CoRR*, 2024.

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

A Mathematical Derivations

A.1 Proof of Equation (3)

Assuming that the cross-sectional area of each particle in the skin is equal to A, and the particles are arranged closely to each other, with a particle density of ρ per unit volume. At this time, when a beam of light with intensity $I(s)$ is incident and passes through a skin layer with thickness dz, we can obtain the attenuated light intensity

$$I(z + dz) - I(z) = -\rho A I(z) dz \tag{16}$$

Given that dz is extremely small, then $I(z + dz) - I(z) \approx dI$. It is worth noting that the attenuation of light is caused by particle absorption, so the above equation can be rewritten as

$$\frac{dI}{dz} = -\mu_a(z) I(z) \tag{17}$$

In addition, considering the presence of both in-scattering and out-scattering among particles, the illumination variation equation within an infinitesimal thickness dz can be derived following the

formulation of equation (17)

$$\begin{aligned} \frac{dI}{dz} &= -\mu_a(z)I(z) - \mu_s(z)I(z) + \mu_s(z)I_z(z) \\ &= \mu_s(z)I_z(z) - I(z)(\mu_a(z) + \mu_s(z)) \\ \Rightarrow & I^{'}(z) + I(z)(\mu_a(z) + \mu_s(z)) = \mu_s(z)I_s(z) \end{aligned} \tag{18}$$

where $\mu_s = \mu_s^{'}$. The equation (18) is a linear ordinary differential equation. We first assume $(\mu I)^{'} = y$, then we have

$$\begin{aligned} \int (\mu I)^{'} dx &= \int y dx \\ \Rightarrow \mu I &= \int y dx + C \\ \Rightarrow I &= \frac{\int y dx + C}{\mu} \end{aligned} \tag{19}$$

while consider $(\mu I)^{'} = y$, we have

$$\begin{aligned} (\mu I)^{'} &= y \\ \Rightarrow \mu I^{'} + \mu^{'} I &= y \\ \Rightarrow I' + \frac{\mu^{'}}{\mu} I &= \frac{y}{\mu} \end{aligned} \tag{20}$$

Set $(\mu_a(z) + \mu_s(z)) = \frac{\mu^{'}}{\mu}$ and $\mu_s(z)I_s(z) = \frac{y}{\mu}$. Then, we have $\mu = e^{\int (\mu_s(z)+\mu_a(z))}$ and $y = \mu_s(z)I_s(z)e^{\int (\mu_a(z)+\mu_s(z))}$. Substitute these two equations into equation (19), we have

$$\begin{aligned} I(z) &= \frac{\int^z \mu_s(z)I_s(z)e^{\int^t (\mu_a(t)+\mu_s(t))dt} dz + C}{e^{\int^z (\mu_a(z)+\mu_s(z))dz}} \\ &= \int^z \mu_s(z)I_s(z)e^{\int^t (\mu_a(t)+\mu_s(t))dt} dz \cdot e^{-\int^z (\mu_a(z)+\mu_s(z))dz} + Ce^{-\int^z (\mu_a(z)+\mu_s(z))dz} \\ &= \int^z \mu_s(z)I_s(z)e^{-\int_t^z (\mu_a(t)+\mu_s(t))dt} dz + Ce^{-\int^z (\mu_a(z)+\mu_s(z))dz}. \end{aligned} \tag{21}$$

The $Ce^{-\int^z (\mu_a(z)+\mu_s(z))dz}$ can be regarded as background light, and since it is assumed that the light source is located deep within the skin, this term can be neglected. Finally, we can obtain the equation (3)

A.2 Proof of Equation (4)

We divide the optical path into N equally spaced intervals. Meanwhile, for ease of computation, we assume μ_s,μ_a, $(z)I_s(z)$ that is equal everywhere throughout the single optical path interval, thus obtaining

$$\begin{aligned} I(z_n \to z_{n+1}) &= \int_{z_n}^{z_{n+1}} \mu_s(z)I_s(z)e^{-\int_0^z (\mu_a(t)+\mu_s(t))dt} dz \\ &= \mu_s I_s \int_{z_n}^{z_{n+1}} e^{-\int_0^z (\mu_a(t)+\mu_s(t))dt} dz \\ &= \mu_s I_s \int_{z_n}^{z_{n+1}} e^{-(\int_0^{z_n} (\mu_a(t)+\mu_s(t))dt + \int_{z_n}^z (\mu_a(t)+\mu_s(t))dt)} dz \\ &= \mu_s I_s e^{-\int_0^{z_n} (\mu_a(t)+\mu_s(t))dt} \int_{z_n}^{z_{n+1}} e^{-(\int_{z_n}^z (\mu_a(t)+\mu_s(t))dt)} dz \\ &= \mu_s I_s e^{-\int_0^{z_n} (\mu_a(t)+\mu_s(t))dt} \int_{z_n}^{z} e^{-(\mu_a+\mu_s)(z-z_n)} dz \\ &= \frac{\mu_s}{\mu_a+\mu_s} I_s e^{-\int_0^{z_n} (\mu_a(t)+\mu_s(t))dt} (1 - e^{(\mu_a+\mu_s)(z_{n+1}-z_n)}) \end{aligned} \tag{22}$$

Accumulate the above equation (22), and make $\delta_n = z_{n+1} - zn$. Then, we have

$$
\begin{aligned}
T_n &= e^{-\sum_{j=0}^{n-1}(\mu_s+\mu_a)\delta_n} \\
I(n) &= \sum_{n=1}^{N} \frac{\mu_s}{\mu_s+\mu_a} I_s T_n (1 - e^{-(\mu_s+\mu_a)*\delta_n})
\end{aligned}
\tag{23}
$$

Finally, we can obtain the equation (4).

B Model Details

B.1 Hyperparameter Settings

In this section, we provide a detailed explanation of the hyperparameter settings for each module in NeR-rPPG as shown in the Table 9

Table 9: Parameter settings for each module of NeR-rPPG.

Module	Hyperparameter	Value
MovE	Dim	64
	Head	8
	Head Dim	12
	Layer Number	4
LigM	Channel	64
	Layer Number	3
ResM	n	3
	Number of a	$N(a) \sim U(2, 24)$
	a	$a \sim \mathcal{N}(0, 1)$
NerM	Sample Points	32
	Dim	32
	Hidden Dim	128
	Position Encode Number	6
	Layer Number	6
rPPM	Dim	128
	Head	4
	Head Dim	32
	Layer Number	10

B.2 NeR-rPPG Algorithm Flow

The overall process of NeR-rPPG is shown in Algorithm 1.

Algorithm 1 NeR-rPPG algorithm flow

Input: X, K
Output: $R, \mathcal{X}_o$
$X_n = E_{stem0}(X), X_r = E_{stem1}(X)$
$\mathcal{K} = \frac{k_{i+1}-k_i}{\delta t}, k_i \in K$
$\mathcal{X}_m = MovE(\mathcal{K}, X_n)$
$X_L = LigM(X_n)$
$D, D_I, C_o = MLP(Pos, X_l, \mathcal{X}_m)$
$R = NerM(X_r, C_o, D, D_I, X_l, \mathcal{X}_m)$
Initialize $a \sim \mathcal{N}(0, 1), \mathcal{R} = C(R_i, a)$
$r(t) = cos^{2n}(\pi f_R t), n \in Z^+$
Obtain the $\mathcal{X}_o$ using equation (8), (9) and (4)

C Additional Experimental Results and Visualization

C.1 Discussion on Other Vital Signal Monitoring

Table 10: Estimation errors on four benchmark datasets.

UBFC-rPPG		PURE		COHFACE		DEAP	
MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
0.056	0.081	0.054	0.071	0.063	0.094	0.154	0.387

The primary objective of this work is to improve heart rate (HR) estimation through physics-guided skin reflectance modeling, rather than explicitly performing multitask physiological measurements. Nevertheless, the proposed framework incorporates physiological factors related to hemoglobin absorption, oxygenation, and respiration, providing a natural foundation for extending the model to multiple vital sign estimation tasks. To preliminarily explore this potential, we additionally evaluate respiratory rate (RR) estimation using the reconstructed rPPG signals, as reported in Table 10. However, the datasets employed in this study lack comprehensive ground-truth annotations for multiple physiological signals, limiting the development and systematic evaluation of multitask learning. Therefore, we focus on HR estimation in the present work, while extending the proposed physics-guided framework to joint physiological signal estimation remains a promising direction for future research.

C.2 Ablation of Physical Representation

To evaluate the effectiveness of the proposed NeRF-inspired modeling strategy, we conduct ablation experiments on four benchmark datasets, as reported in Table 11. We investigate the influence of decoder architectures by evaluating ViT-based and CNN-based alternatives. Although the CNN decoder generally outperforms the ViT decoder, both exhibit higher estimation errors than the complete NeR-rPPG framework. These results suggest that the performance improvements cannot be attributed solely to the choice of decoder architecture. Instead, the NeRF-inspired formulation provides an effective mechanism for integrating motion, illumination, and physiological representations under the proposed skin reflection model, thereby facilitating more accurate rPPG estimation.

As shown in Table 11, incorporating the NeRF-inspired module consistently improves estimation accuracy across all four datasets. Compared with the model without NeRF, the proposed framework achieves substantial reductions in both MAE and RMSE. Furthermore, although the CNN-based decoder outperforms the ViT-based decoder, neither achieves comparable performance to the complete NeR-rPPG framework. These results suggest that the proposed physics-guided modeling strategy provides more effective representations for rPPG estimation than conventional decoder architectures.

C.3 Hyperparameter Impact

Additional experiments were conducted on the hyperparameters listed in Table 9 to evaluate their impact on the performance of NeR-rPPG. As illustrated in Figure 3(a), a pronounced performance degradation is observed for NeR-rPPG when the number of layers in the rPPM is reduced. This can be attributed to the decreased nonlinear modeling capacity caused by the shallower rPPM, which in turn weakens the overall feature extraction capability of the model and leads to a decline in reconstruction accuracy. In contrast, reducing the number of layers in the other modules does not result in a significant performance drop. This is because these modules primarily serve as auxiliary components during the optimization process and impose constraints on the rPPM, whereas the rPPM functions as the core module responsible for modeling and reconstructing the rPPG signals.

As shown in Figure 3(c), increasing the value of n in ResM leads to a degradation in the performance of NerM. This behavior can be attributed to the fact that a larger n introduces stronger respiratory harmonics, which give rise to non-rPPG-related features. These spurious features interfere with the rPPM's ability to accurately reconstruct rPPG signals, thereby reducing the overall estimation accuracy.

Table 11: Ablation study of different network configurations on four datasets.

Method	UBFC-rPPG		PURE		COHFACE		DEAP	
	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE
w/o NeRF	0.88	1.45	1.31	2.78	1.67	3.41	4.03	7.62
w NeRF	**0.42**	**0.98**	**0.56**	**0.85**	**0.60**	**1.51**	**2.78**	**3.75**
Decoder(ViT)	1.63	2.56	1.73	2.91	1.71	3.50	4.13	7.91
Decoder(CNN)	0.94	1.70	1.35	2.84	1.52	2.97	3.89	6.92

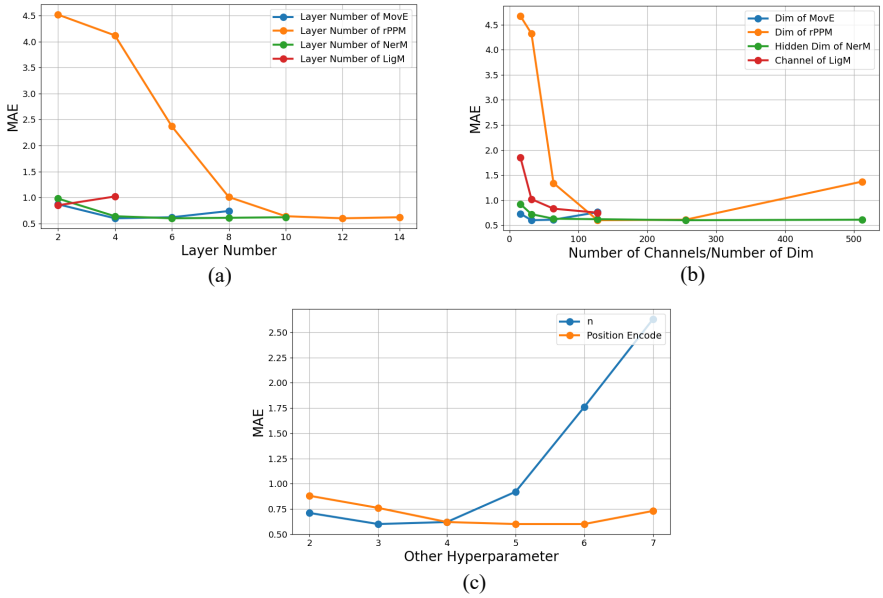


Figure 3: Experimental results for different hyperparameters in each module.

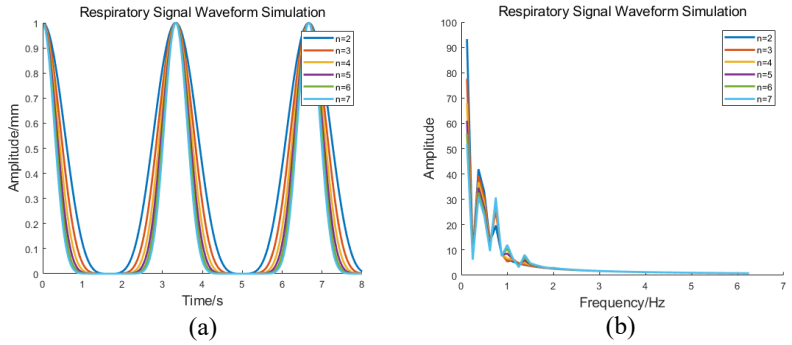


Figure 4: Experimental results for different hyperparameters in each module.

Furthermore, we visualize the respiratory signal waveforms and their corresponding frequency spectra generated by ResM in Figure 4. As shown, increasing the value of n leads to more pronounced and longer-duration plateaus between the peaks of the simulated respiratory signals. Correspondingly, the frequency spectra exhibit stronger respiratory harmonics. These effects indicate that larger n values introduce additional non-rPPG components, thereby posing greater challenges for rPPM in accurately estimating rPPG signals.

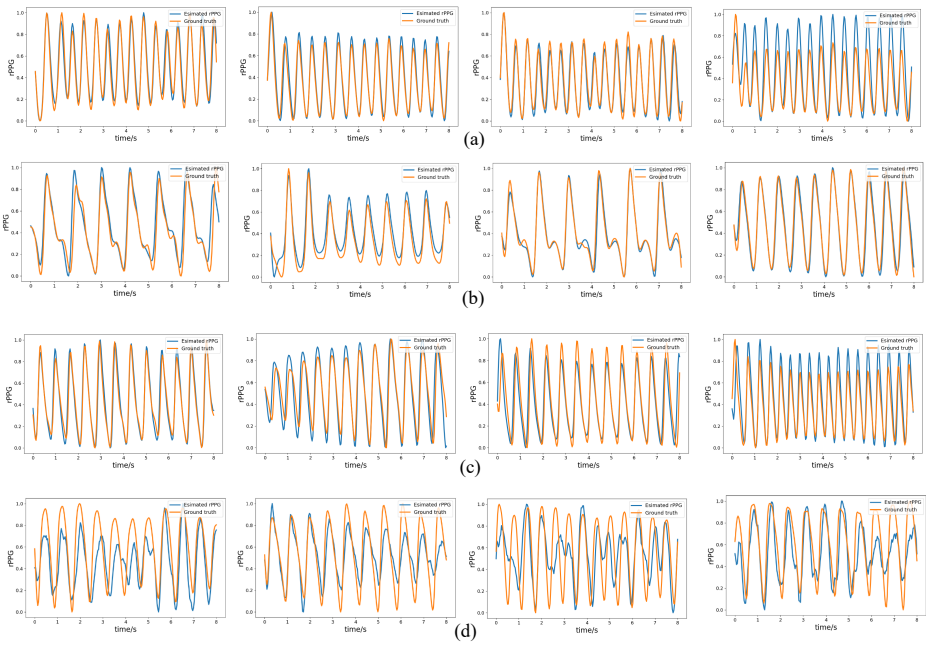


Figure 5: rPPG estimation results of NeR-rPPG on four datasets. (a) is the results on UBFC, (b) is the results on PURE, (c) is the results on COHFACE and (d) is the results on DEAP.

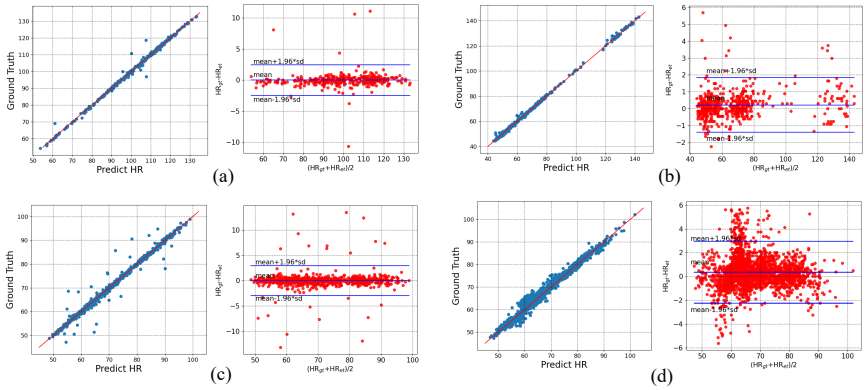


Figure 6: Bland–Altman plots of NeR-rPPG on four datasets. (a) is the results on UBFC, (b) is the results on PURE, (c) is the results on COHFACE and (d) is the results on DEAP.

C.4 Estimate rPPG Signal Visualization

To illustrate the rPPG signal estimation capability of NeR-rPPG, we visualize its estimation results in Figure 5. The results suggest that NeR-rPPG is able to capture key rPPG signal features and reconstruct temporal waveform morphology. This performance is likely facilitated by the proposed skin reflection model and the NeRF-guided framework, which together provide constraints on the rPPM module and support more accurate rPPG feature estimation. To further analyze the HR estimation capability of NeR-rPPG, we employed Bland–Altman plots to illustrate the differences between the estimated and ground-truth heart rates as shown in 6. As shown in the Figure 6, the heart rates estimated by NeR-rPPG exhibit a high level of agreement with the ground-truth values, with the majority of the samples falling within the limits of agreement. However, it is worth noting that the performance on the DEAP dataset is comparatively inferior. This can be attributed to the substantially larger dataset size of DEAP compared to the other three datasets, as well as the presence of challenging factors such as facial occlusions, which pose additional difficulties for accurate rPPG estimation.

C.5 Reconstruction Video Visualization

We visualize the reconstruction results of the rPPG signals as well as the rendering results of the videos as shown in Figure 7 As shown in Figure 7, NeR-rPPG is capable of accurately estimating

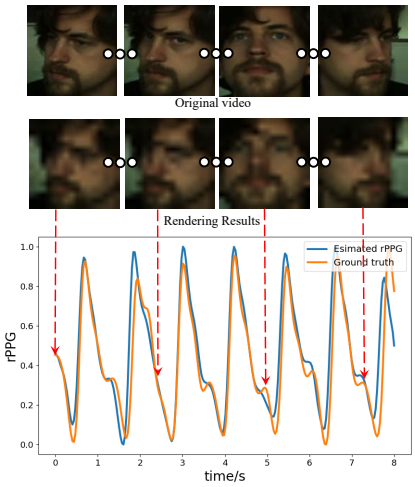


Figure 7: RPPG signal reconstruction results and video rendering results.

rPPG signals while simultaneously performing video rendering. It is observed that the rendered videos appear blurrier than the original inputs. This phenomenon can be attributed to the reduced spatial resolution of the features output by the Stem module, which downsamples the input to a spatial size of $\frac{H}{8} \times \frac{W}{8}$. Then, NerM performs rendering based on these low-resolution features and applies a reconstruction loss directly against the downsampled ground-truth video, thereby encouraging one-to-one pixel-level reconstruction at a reduced spatial resolution. When compared with the original full-resolution video, the rendered output must be upsampled, which inevitably leads to the loss of fine-grained details.

To provide a more detailed illustration of the video rendering performance of NeR-rPPG, we present the rendering results of NeR-rPPG on four datasets in the figures. It is worth noting that the rendered videos are generated at a spatial resolution of 16×16; consequently, after upsampling, certain fine-grained details cannot be fully recovered.

C.6 Pixel-wise Time–Frequency Spectrum Comparison Between the Rendered Video and the Downsampled Ground-Truth Video

To demonstrate whether the video rendered by NeR-rPPG can capture the genuine temporal information of the original video pixels, rather than merely fitting the global statistical distribution, we present below the pixel-wise time–frequency spectrum comparison between the rendered video and the downsampled ground-truth video, as well as the pixel-wise time–frequency spectrum comparison within the heart-rate frequency range.

As shown in Figure 14 and Figure15, the pixel-wise time–frequency spectra of the videos rendered by NeR-rPPG are highly consistent with those of the corresponding downsampled ground-truth videos. Moreover, Fig. 14 further indicates that the rendered videos preserve latent heart-rate-related components that are similar to those observed in the real videos. These results suggest that NeR-rPPG is able to capture meaningful temporal physiological information from facial videos, thereby supporting its effectiveness in extracting latent rPPG signals.



Figure 8: Rendering results on the UBFC dataset.

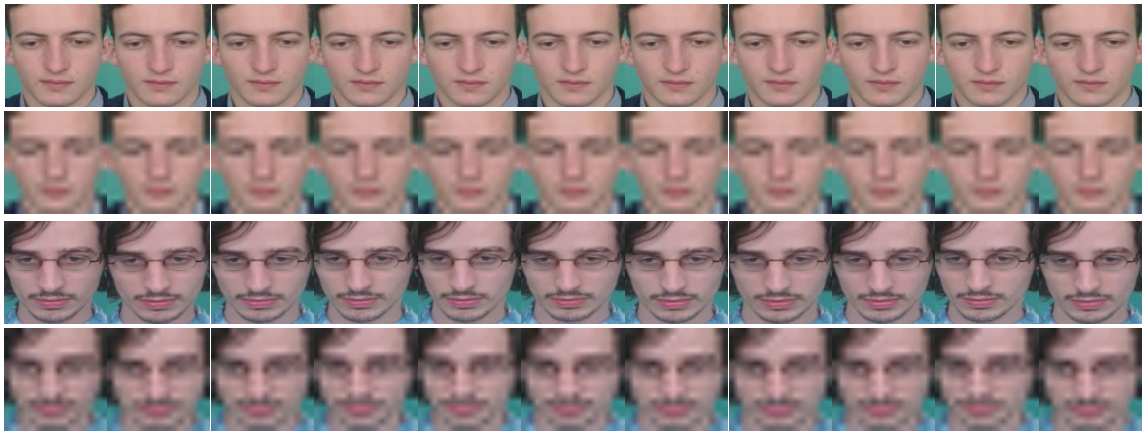


Figure 9: Rendering results on the PURE dataset.

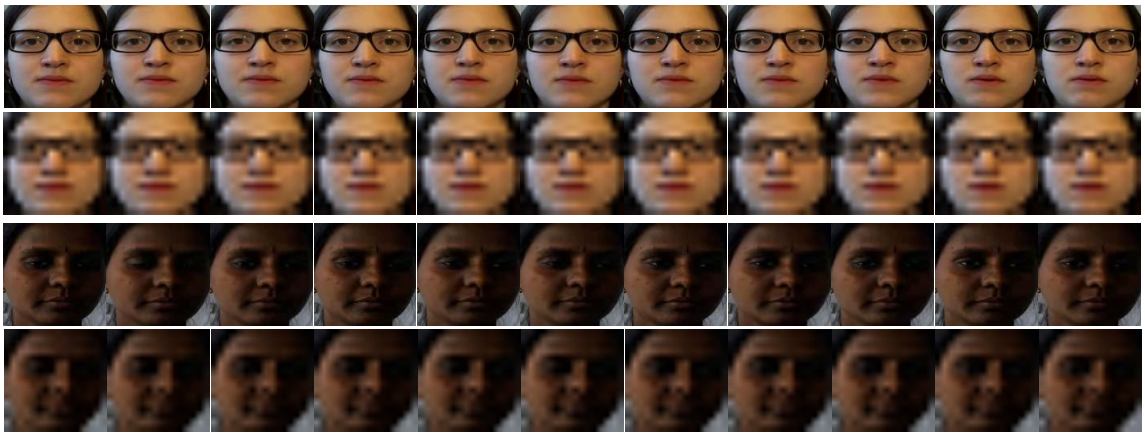


Figure 10: Rendering results on the COHFACE dataset.

NeurIPS Paper Checklist

1. **Claims**

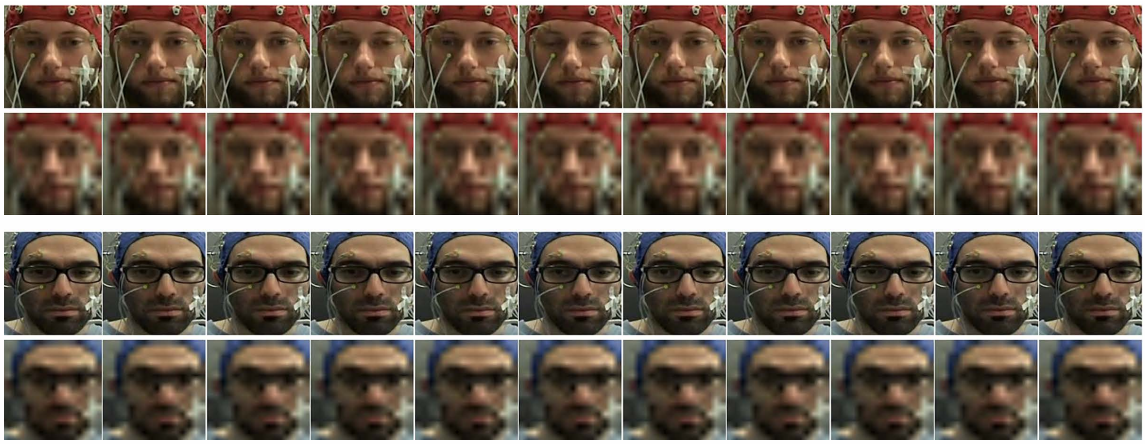


Figure 11: Rendering results on the DEAP dataset.

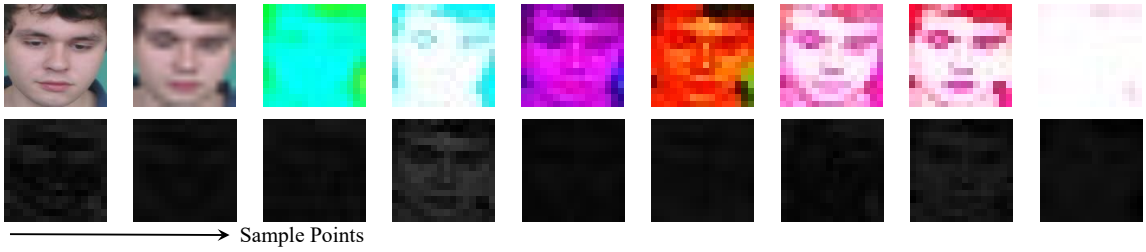


Figure 12: Weights of sampling points at different spatial locations and corresponding RGB results on the UBFC dataset.

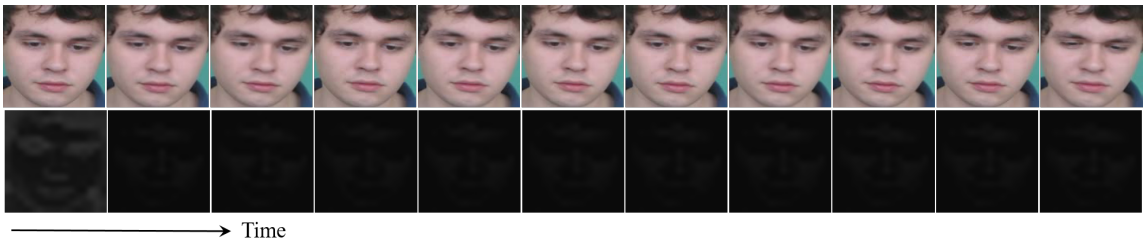


Figure 13: Temporal variation of the weight of the 0-th sampling point.

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

Answer: [Yes]

Justification: We have explicitly and clearly clarified the contributions and main content of our work in both the Abstract and the final paragraph of the Introduction.

Guidelines:

- The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
- The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.

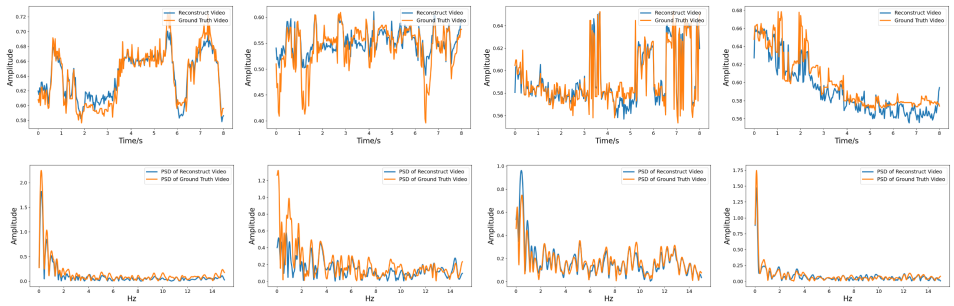


Figure 14: Pixel-wise time–frequency spectrum comparison between the rendered video and the downsampled ground-truth video.

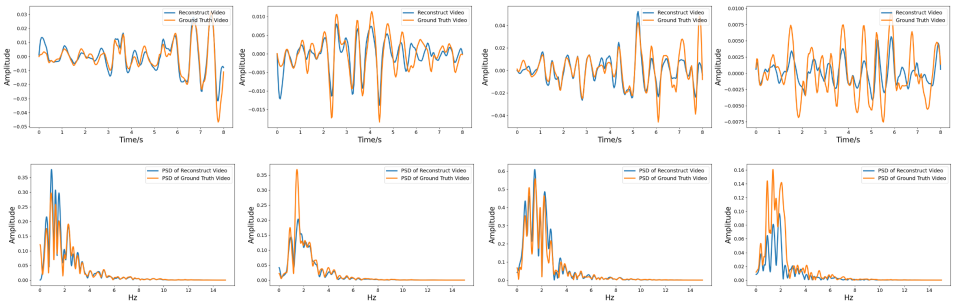


Figure 15: Pixel-wise time–frequency spectrum comparison within the heart-rate frequency range.

- The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
- It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

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

   Answer: [Yes]

   Justification: We have discussed the limitations of our method in the final paragraph of the Conclusion.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.

- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

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

   Answer: [Yes]

   Justification: We have provided the detailed derivation of the required equations and the associated assumptions in Appendix A.

   Guidelines:

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

4. **Experimental result reproducibility**

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

   Answer: [Yes]

   Justification: We disclose the key information required to reproduce the main experimental results, including the dataset settings, preprocessing pipeline, model configuration, training details, and evaluation protocol. Specifically, we describe the datasets and data splits, the facial video preprocessing procedure, the input and output settings, the main architectural components, the loss functions, and the evaluation metrics used for rPPG estimation. Additional implementation details and derivations are provided in the appendix.

   Guidelines:

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

- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

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

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

   Answer: [Yes]

   Justification:We provide the core implementation code. We also provide detailed information on how to obtain the datasets used in our experiments, including the corresponding access links or application procedures when required.

   Guidelines:

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

6. **Experimental setting/details**

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

   Answer: [Yes]

   Justification: We provide the detailed hyperparameters of the model in the appendix B.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.

- The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

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

   Answer: [Yes]

   Justification: We provide detailed statistical information on HR estimation in Appendix B.

   Guidelines:

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

8. **Experiments compute resources**

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

   Answer: [Yes]

   Justification: We provide the detailed experimental conditions for the computational analysis in the Implementation Details section.

   Guidelines:

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

9. **Code of ethics**

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

   Answer: [Yes]

   Justification: We conducted experimental validation using publicly available datasets, in compliance with the NeurIPS ethical guidelines.

   Guidelines:

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

10. **Broader impacts**

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

    Answer: [Yes]

    Justification: We discuss both the potential positive and negative societal impacts of this work. On the positive side, non-contact rPPG measurement may support remote health monitoring, telemedicine, driver fatigue detection, and low-cost vital-sign assessment using widely available cameras. On the negative side, camera-based physiological measurement may raise privacy concerns if used without consent, and its performance may vary across populations or challenging environments, which could lead to unreliable conclusions if deployed without proper validation. We therefore emphasize that the method should be used with informed consent, appropriate privacy protection, and careful evaluation across diverse populations and real-world conditions.

    Guidelines:

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

11. **Safeguards**

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

    Answer: [N/A]

    Justification: This work does not release new datasets or high-risk models. We conduct experiments on publicly available rPPG datasets and provide only the core implementation code for reproducibility.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.

- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: We properly cite the original papers and sources of all datasets, codebases, models, and tools used in this work. We also follow the corresponding licenses and terms of use, including dataset access requirements and usage restrictions. For datasets that require application or agreement, we obtain and use them according to their official access procedures.

    Guidelines:

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

13. **New assets**

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

    Answer: [Yes]

    Justification: We release part of the core implementation code and additional visualization results through an anonymized repository. We also specify the dataset access procedures, implementation details, and limitations in the appendix. No new dataset or model containing personal information is released.

    Guidelines:

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

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

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

Answer: [N/A]

Justification: This work does not involve crowdsourcing experiments or newly conducted human-subject studies. We conduct experiments using publicly available rPPG datasets and do not recruit participants or collect new human-subject data.

Guidelines:

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

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

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

Answer: [N/A]

Justification: This work does not involve crowdsourcing experiments or newly conducted human-subject studies. We conduct experiments using publicly available rPPG datasets and do not recruit participants or collect new human-subject data.

Guidelines:

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

16. **Declaration of LLM usage**

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

Answer: [N/A]

Justification: LLMs are not used as an important, original, or non-standard component of the core methodology. They are not involved in model design, data generation, labeling, training, evaluation, or result analysis.

Guidelines:

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