❤️ Heartian: Physiology-Aware Relightable Gaussian Head Avatar[*]

Xiaoyue Fan
University College London
London, UK
merryxyfan@gmail.com

Jose Echevarria
Adobe Research
New York, NY, USA
echevarr@adobe.com

Akshay Paruchuri
Stanford University
Stanford, CA, USA
akshaypa@stanford.edu

Kaan Akşit
University College London
London, UK
kaanaksit@kaanaksit.com

Abstract

Gaussian head avatars typically model intrinsic facial appearance as temporally static, omitting subtle cardiac-induced skin-color variation. We propose ❤️ Heartian, a physiology-aware modulation framework that learns cardiac-cycle-dependent per-frame albedo modulation of facial skin-region Gaussians within a relightable head avatar to encode remote photoplethysmography (rPPG) signals. Using synchronized contact PPG supervision, ❤️ Heartian models the prescribed cardiac waveform as the sum of two Gaussian functions and learns per-frame spatial residuals via a lightweight MLP. Across 152 stationary recordings from UBFC-rPPG, PURE, and MMPD, attribute-space recovery of the supplied signal achieves a pooled recording-level heart-rate MAE of 0.29 bpm and MAPE of 0.38%. The signals remain detectable after rendering by benchmark rPPG methods, with the best tested configuration – a motion-augmented TS-CAN decoder pretrained on UBFC-rPPG – recovering heart rate from the rendered MMPD avatars at 0.97 bpm MAE and 1.21% MAPE. Meanwhile, ❤️ Heartian maintains reconstruction quality comparable to the baseline, with negligible average PSNR degradation of 0.005 dB. Overall, our work embeds recoverable rPPG signals as controllable material attributes to subject-specific Gaussian head avatars, while retaining reconstruction quality.

CCS Concepts

• **Computing methodologies → Computer vision**; **Reconstruction**; • **Applied computing → Life and medical sciences**.

Keywords

3D Gaussian Splatting, Gaussian Head Avatar, Remote Photoplethysmography, Relightable Physiology-Aware Avatars

ACM Reference Format:

Xiaoyue Fan, Jose Echevarria, Akshay Paruchuri, and Kaan Akşit. 2026. ❤️ Heartian: Physiology-Aware Relightable Gaussian Head Avatar. In *SIGGRAPH Asia 2026 Technical Communications (SA Technical Communications '26), December 01–04, 2026, Kuala Lumpur, Malaysia.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3829339.3847838

1 Introduction

Conventional videos of people reveal physiological signals through subtle, temporal skin-color variation, from which rPPG recovers vital signs such as heart rate. Prior works synthesize physiological facial appearance in images [McDuff and Nowara 2021] or videos [Wang et al. 2022], whereas [Lu et al. 2025] uses Gaussian representations to estimate unknown physiology from input video. However, 3D-consistent physiological appearance modeling remains underexplored in subject-specific, relightable head avatars.

Volumetric 3D human reconstruction has been significantly advanced by 3DGS [Kerbl et al. 2023]. The work by [Lu et al. 2025] has validated the utility of GS in rPPG tasks with dynamic motions, where explicit 3D representation disentangles the chromatic and geometric variations beyond purely surface-pixel-level analysis. Meanwhile, HRAvatar [Zhang et al. 2025] introduces a human head reconstruction method, leveraging Gaussian primitives upon an enhanced FLAME-based face model, supporting real-time relighting. However, it maintains a pose-independent static appearance, where per-Gaussian color properties are optimized globally rather than per frame. Consequently, while frame-wise lighting variations contribute to the final rendered appearance, inter-frame skin chromatic variations of physiological origin are inherently not captured.

We propose ❤️ Heartian, a physiology-aware modulation framework that embeds prescribed rPPG signals into relightable Gaussian head avatars. Building on HRAvatar reconstruction, we optimize per-frame spatial modulation factors acting on the albedo of facial skin-region Gaussians. Specifically, the modulation comprises two components: a fundamental waveform modeled as the sum of two Gaussian functions, following [Tang et al. 2020], and a lightweight multilayer perceptron (MLP) that learns per-frame spatial residuals. The rPPG signal is extracted as the mean green channel intensity after modulation across the skin-region Gaussians, supervised against the ground truth PPG waveform. In summary, our method enables the recovery of heart rate from embedded signals with a mean MAE of 0.29 bpm and MAPE of 0.38 % against ground truth PPG measurements while preserving comparable reconstruction quality after modulation, with a marginal cost of 0.005 dB, 0.00003, and 0.0003 in average PSNR, SSIM, and LPIPS, respectively. Meanwhile, embedded rPPG signals are vastly preserved in the rendered videos, evaluated with the benchmark methods by rPPG-Toolbox [Liu et al. 2023], exhibiting particularly strong performance on the MMPD dataset [Tang et al. 2023]. Unlike methods that use a generalizable 4D Gaussian representation as an intermediate to disentangle feature variation for rPPG estimation [Lu et al. 2025], ❤️ Heartian embeds prescribed rPPG signals via attribute modulation into subject-specific Gaussian head avatars as a renderable property. Furthermore, while retaining the relighting capability of HRAvatar, our pipeline enables rPPG avatar synthesis under novel illumination environments without compromising recoverable signal quality. Our strategy also enables prescribed waveform and heart-rate control within physiology-aware Gaussian representations, potentially supporting controlled physiological training

[*] ❤️: Twemoji © Twitter, Inc. and other contributors, 2019.

ACM ISBN 979-8-4007-2841-9/2026/12
https://doi.org/10.1145/3829339.3847838

augmentation and informing future physiology-aware real-time avatar models for applications such as telemedicine. [1]

2 Methods

Our method embeds synchronized rPPG signals into a relightable Gaussian head avatar reconstructed from monocular video, building on HRAvatar [Zhang et al. 2025] as our baseline. ❤️ Heartian is an offline, per-recording authoring method rather than a real-time physiology-inference system. Monocular videos undergo standard preprocessing and facial tracking to extract per-frame FLAME parameters, which are subsequently fed into HRAvatar to reconstruct a high-fidelity, subject-specific, and relightable head avatar represented as a set of physically parameterized 3D Gaussians $\mathcal{G}$.

PPG Signals Modeling. Cardiac pulsation results in skin color variations, imperceptible to the naked eye but detectable to cameras. We identify a dedicated set of Gaussians $\mathcal{G}^s$ from $\mathcal{G}$ and modulate them during rendering, embedding the target rPPG signal without altering the underlying Gaussian parameters. To identify $\mathcal{G}^s$, we extract a facial skin-region mask from the well-trained $\mathcal{G}$ representation. Each $\mathcal{G}_i^s$ is examined by its Euclidean distance to the nearest vertex on the FLAME mesh. As depicted in Figure 1, a selected $\mathcal{G}_i$ is designated as $\mathcal{G}_i^s$ if this distance falls below a threshold (e.g., 0.02) and its normal points in the front-facing direction, excluding Gaussians on the sides of the head where hair occlusion is likely.

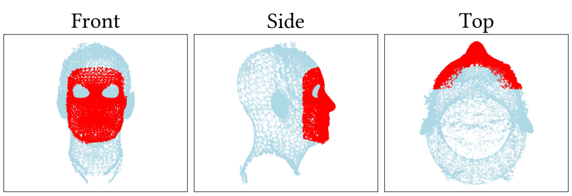

Figure 1: Skin Gaussians $\mathcal{G}^s$ (red) overlaid on $\mathcal{G}$ (blue).

For each $\mathcal{G}_i^s$, a spatially varied transient offset is applied to the green channel albedo $\mathbf{c}$ on a per-frame basis, while all other attributes remain unchanged. Specifically, the modulation at frame t follows:

$$\mathbf{c}_i^{\mathrm{m}}(t) = \mathbf{c}_i^{\mathrm{base}}(t) + A \cdot m_i(t), \tag{1}$$

where A is a learnable scaling factor and $m_i(t)$ denotes the per-frame modulation scalar. The PPG waveform within each cardiac cycle characteristically exhibits two distinct peaks, a systolic and a diastolic wave. Their structure can be embedded in phase space by mapping the temporal signal to a unit circle parameterized by the cumulative cardiac phase $\theta(t)$ [Tang et al. 2020], as illustrated in Figure 2. Therefore, we model the $m_i(t)$ comprising the physiologically motivated fundamental waveform and the learned residual:

$$m_i(t) = \underbrace{\sum_{k=1}^{2} A_k \exp\left(-\frac{d(\theta(t),\mu_k)^2}{2\sigma_k^2}\right)}_{\text{fundamental}} + \underbrace{B \cdot f_{\mathrm{MLP}}(\theta(t), b(t), p_i)}_{\text{residual}}, \tag{2}$$

where $d(\theta,\mu) = \arctan 2(\sin(\theta-\mu), \cos(\theta-\mu))$ denotes the shortest angular distance on the unit circle. The fundamental component models the characteristic PPG waveform morphology as a sum of two Gaussian functions in phase space [Tang et al. 2020], with learnable centers μ_k, widths σ_k, and amplitudes A_k. $\theta(t)$ is modeled as a learnable initial phase ϕ_0 accumulated with per-frame increments $\{\delta_k\}$, which are passed through a softplus activation to ensure positivity, and projected onto the unit circle as:

$$\begin{aligned} \phi(t) &= \phi_0 + \sum_{k=0}^{t} \delta_k, \\ \theta(t) &= \arctan 2(\sin(\phi(t)), \cos(\phi(t))). \end{aligned} \tag{3}$$

The residual term is a lightweight f_{MLP} conditioned on $\theta(t)$, a beat index embedding $b(t)$, and the normalized spatial position p_i of each $\mathcal{G}_i^s$. While $\theta(t)$ encodes the continuous phase position within a cardiac circle, $b(t)$ provides a discrete beat identity that enables f_{MLP} to capture inter-beat variability. The spatial conditioning on p_i further allows the modulation to vary across the face, reflecting the spatially heterogeneous nature of the skin perfusion. Specifically, the phase embeddings follow the Fourier-based positional embedding approach [Mildenhall et al. 2020] with 10 harmonics:

$$e_\phi(\theta) = [\sin(k\theta), \cos(k\theta)]_{k=1}^{10} \in \mathbb{R}^{20}. \tag{4}$$

$b(t)$ is encoded via a learnable embedding $e_b(t) \in \mathbb{R}^8$, while p_i is embedded by a two-layer encoder into $e_s(p_i) \in \mathbb{R}^8$. The global features $[e_\phi(\theta),\ e_b(t)]$ are expanded and concatenated with the per-Gaussian $e_s(p_i)$, forming $\boldsymbol{x} = [e_\phi(\theta),\ e_b(t),\ e_s(p_i)] \in \mathbb{R}^{36}$, projected to dimension 128, passed through 2 residual blocks with tanh activations, and linearly projected to a per-Gaussian scalar residual.

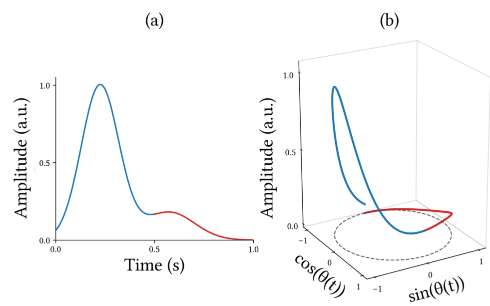

Figure 2: A typical PPG waveform with systolic (blue) and diastolic waves (red) in the time (a) and phase domains (b).

Heart Rate Estimation Protocol. Following the rPPG principle, heart rate is estimated from the temporal variation of the mean green albedo $\bar{c}$ across $\mathcal{G}^s$. Unlike RGB-rendered values, which are susceptible to environmental illumination variations and rendering noise, albedo provides a more stable and cleaner basis for extracting subtle signals. Functions of rPPG-Toolbox [Liu et al. 2023] are adapted for signal detrending, bandpass filtering, and frequency analysis via Fast Fourier Transform (FFT). The raw signal is detrended to remove low-frequency drift via a smoothness prior, followed by a first-order Butterworth bandpass filter to isolate cardiac frequencies (f_L = 0.75 Hz, f_H = 3.0 Hz) using zero-phase forward-backward filtering to prevent phase distortion. Heart rate is taken as the dominant spectral peak within the filter range via FFT.

[1] Our approach could be layered atop existing avatar infrastructure, such as teleconferencing pipelines (e.g., *Google Beam*) or anatomically complete parametric head models (e.g., *GNM*) to embed prescribed physiological signals into avatar reconstruction. Institutions such as *Universität Hamburg* have explored integrating vital-sign monitoring into telemedicine workflows, further suggesting that physiology-aware representations may eventually be incorporated into broader real-time avatar pipelines.

Optimization. Our training follows a two-stage strategy. In the first stage, the HRAvatar baseline [Zhang et al. 2025] is trained to optimize the full Gaussian set $\mathcal{G}$ and FLAME parameters under $\mathcal{L}_{\text{base}}$, comprising the photometric loss. In the second stage, all baseline parameters are frozen to preserve well-learned geometry and static appearance, while only the per-frame-varying Remote Photoplethysmography (rPPG) signal modulation parameters of $\mathcal{G}^s$ are optimized. The rendered images pass through the full $\mathcal{L}_{\text{base}}$ to ensure the modulated albedo remains visually consistent. In addition, we incorporate rPPG-specific losses, targeting both frequency accuracy and waveform fidelity. The phase loss regularizes the cardiac rhythm by measuring the mean circular distance between $\theta(t)$ and the beat-aligned reference phase θ_{gt}, constructed by detecting systolic peaks in the ground truth PPG signal and linearly interpolating between consecutive peaks. This normalization ensures each cardiac cycle occupies the same phase interval, allowing the global μ_k and σ_k to learn a consistent morphological representation across the full sequence without drifting with heart rate variability:

$$\mathcal{L}_{\text{phase}} = \mathbb{E}\left[\left(\text{atan2}\left(\sin(\theta - \theta_{\text{gt}}),\ \cos(\theta - \theta_{\text{gt}})\right)\right)^2\right]. \tag{5}$$

Besides, we apply a weak phase regularization via instantaneous phase estimates θ_{POS} derived from raw frames using the Plane-Orthogonal-to-Skin (POS) algorithm [Wang et al. 2016] followed by a Hilbert transform. Unlike the supervised $\mathcal{L}_{\text{phase}}$ enforcing morphological consistency, POS-generated signals offer stable periodicity as a soft prior to discourage arbitrary phase drift while preserving the model's capacity to learn morphological variation:

$$\mathcal{L}_{\text{pos}} = \mathbb{E}\left[\left(\text{atan2}\left(\sin(\theta - \theta_{\text{pos}}),\ \cos(\theta - \theta_{\text{pos}})\right)\right)^2\right]. \tag{6}$$

The waveform losses are applied along two separate supervision paths using the same ground truth PPG signal. The MSE loss $\mathcal{L}_{\text{mse}}$ and Pearson correlation loss $\mathcal{L}_{\text{pearson}}$ supervise the fundamental signal directly, encouraging the Gaussian parameters to approximate the target waveform morphology. The same losses are also applied to the full modulation signal, providing gradients exclusively to f_{MLP} for correcting the remaining spatially varying discrepancy: where s denotes the detrended and normalized input signal for each respective path. To ensure physical plausibility, we additionally apply constraints on the fundamental waveform parameters following [Tang et al. 2020], enforcing $1 \geq A_1 > A_2 \geq 0$, $2 \geq \sigma_2 > \sigma_1 \geq 0$, and $\pi \geq \mu_2 > \mu_1 \geq -\pi$ via quadratic penalty terms:

$$\mathcal{L}_{\text{reg}} = \lambda_{\text{A}}\mathcal{L}_A + \lambda_\sigma\mathcal{L}_\sigma + \lambda_\mu\mathcal{L}_\mu. \tag{7}$$

The total loss in the second stage is:

$$\mathcal{L} = \mathcal{L}_{\text{base}} + \lambda_{\text{pha}}\mathcal{L}_{\text{phase}} + \lambda_{\text{pos}}\mathcal{L}_{\text{pos}} + \lambda_{\text{m}}\mathcal{L}_{\text{mse}} + \lambda_{\text{p}}\mathcal{L}_{\text{pearson}} + \mathcal{L}_{\text{reg}}, \tag{8}$$

where $\lambda_{\text{pha}} = \lambda_{\text{m}} = \lambda_{\text{p}} = 1$, $\lambda_{\text{A}} = \lambda_\sigma = 0.6$, $\lambda_{\text{pos}} = 0.3$ and $\lambda_\mu = 0.1$.

3 Implementation

Our pipeline. We base our pipeline on HRAvatar [Zhang et al. 2025] using PyTorch and optimize with Adam on an RTX 4060 GPU. Training proceeds in two stages: the first stage follows the same configuration to generate the baseline avatar, while the second stage introduces and trains the rPPG modulation parameters for 60 epochs. More details are provided in the supplementary material.

Table 1: Benchmark Evaluation Results. Performance of benchmark methods on our rendered 💓 Heartian videos, evaluated across datasets. Green highlights indicate metrics where embedded signals remain recoverable on par with the original benchmark results on the same selected subsets.

		Test Set					
		UBFC-rPPG		PURE		MMPD	
Method	**Train Set**	MAE↓	MAPE↓	MAE↓	MAPE↓	MAE↓	MAPE↓
POS	-	1.23 ± 0.77	1.01 ± 0.62	14.23 ± 6.78	29.34 ± 13.98	1.53 ± 0.63	2.66 ± 1.20
TS-CAN	UBFC-rPPG	-	-	10.01 ± 4.52	17.89 ± 8.68	2.59 ± 0.52	3.89 ± 0.94
	PURE	9.22 ± 7.47	7.41 ± 5.56	-	-	2.51 ± 0.57	3.87 ± 1.03
TS-CAN (MA)	UBFC-rPPG	-	-	4.74 ± 4.22	5.23 ± 4.67	0.97 ± 0.27	1.21 ± 0.31
PhysFormer	UBFC-rPPG	-	-	17.40 ± 6.55	34.29 ± 13.78	8.69 ± 1.31	13.96 ± 2.21
	PURE	5.53 ± 4.44	4.37 ± 3.32	-	-	3.12 ± 1.08	5.27 ± 1.95
FactorizePhys	UBFC-rPPG	-	-	14.58 ± 6.82	29.95 ± 14.13	2.88 ± 0.98	4.83 ± 1.70
	PURE	6.41 ± 5.09	5.05 ± 3.80	-	-	1.38 ± 0.50	2.40 ± 0.99
Ours vs Baseline							
POS	-	- 21.97	- 18.88	- 10.72	- 13.78	- 12.72	- 17.78
TS-CAN	PURE	- 26.98	- 24.72	-	-	- 12.87	- 16.28
TS-CAN (MA)	UBFC-rPPG	-	-	- 15.73	- 25.01	- 13.31	- 17.28

MAE = Mean Absolute Error in HR estimation (Beats/Min), MAPE = Mean Percentage Error (%).

Dataset. We evaluate on three datasets: UBFC-rPPG [Bobbia et al. 2019], PURE [Stricker et al. 2014], and MMPD [Tang et al. 2023]. For PURE and UBFC-rPPG, we select 10 subjects from each, all recorded under stationary motion. For MMPD, the full 33 subjects are utilized, spanning Fitzpatrick skin types 3 to 6, consistent with stationary motion and retaining all four lighting conditions, yielding 132 videos from this dataset and 152 videos in total. All monocular videos are extracted into frames, cropped to a face-centered 512 × 512 px region, and foreground-matted across the full sequences.

4 Evaluation and Discussion

Evaluation. We evaluate our method by reorganizing the rendered avatar videos according to the requirements of each benchmark, and subsequently testing them through rPPG-Toolbox [Liu et al. 2023]. Our evaluation includes both the traditional unsupervised learning method POS [Wang et al. 2016] and three cutting-edge pre-trained supervised neural networks - TS-CAN [Liu et al. 2020], PhysFormer [Yu et al. 2022], and FactorizePhys [Joshi et al. 2024] - to assess the fidelity and recoverability of the prescribed rPPG signals embedded per recording. Since Gaussian-based avatar reconstructions inherently exhibit subtle frame-level jittering, we further evaluate 💓 Heartian with a motion-augmented variant of TS-CAN [Paruchuri et al. 2024] to investigate this impact. Table 1 shows 💓 Heartian preserves consistently recoverable rPPG signals after rendering across three datasets. Our method achieves particularly strong performance on MMPD [Tang et al. 2023], stably remaining recoverable by off-the-shelf rPPG estimators. In contrast, the best-performing method varies across the other two datasets, revealing gaps across datasets and benchmark methods. Notably, PURE [Stricker et al. 2014] is more sensitive to mesh jittering, whose performance is substantially improved with motion-augmented TS-CAN [Paruchuri et al. 2024], among others. Besides, PhysFormer [Yu et al. 2022] exhibits the largest deviation among supervised methods, as its transformer-based architecture generalizes less robustly across datasets with differing acquisition conditions [Liu et al. 2023]. We further composite the rendered avatar videos onto dynamic real backgrounds, showing no consistent degradation and confirming the robustness of our method beyond mere background removal. A cross-subject experiment on three subjects additionally demonstrates the controllability of our strategy, with the prescribed heart rates explicitly recovered. Beyond rPPG signals, we also assess the reconstruction quality of our modulated Heartian. A small number of videos exhibit visual artifacts due to initial transient spikes in the ground truth signal during data acquisition, which introduce large modulation magnitude, and lighting variation within a single recording, which heightens the inconsistent bases for albedo modulation strategy. However, these artifacts can be eliminated by adjusting the modulation scaling factor accordingly. Our method maintains reconstruction quality comparable to the baseline after albedo modulation with an average of 35.38 PSNR, 0.9697 SSIM, and 0.0374 LPIPS. Extensive results and visualizations are provided in the supplementary material.

Discussion. Since rPPG estimation relies on subtle skin color variations, signals are sensitive to degradation during video compression or transmission. Our approach of embedding rPPG signals directly into the Gaussian albedo offers attribute-level signal preservation. Table 2 shows the metrics of rPPG signals extracted from our modulated Heartian compared with the static HRAvatar baseline, whose extracted albedo is consistently a flat line, demonstrating its lack of inter-frame variation and thus heart rate information. Compared with results derived from the rendered videos in Table 1, the extracted signals show stronger fidelity, suggesting that rPPG information is numerically preserved within the avatar representation when it attenuates through rendering or post-processing.

Table 2: Attribute-level Evaluation. rPPG signal metrics extracted from the Gaussian albedo of baseline and Heartian.

	Baseline			Ours		
Dataset	MAE	MAPE	SNR	MAE↓	MAPE↓	SNR↑
UBFC-rPPG	59.00	53.81	-21.52	0.00	0.00	1.45
PURE	21.53	25.37	-7.13	0.17	0.27	3.47
MMPD	39.11	34.81	-11.47	0.33	0.42	4.04

Lighting condition and skin tone have been challenging factors for rPPG signal estimation. However, benefiting from our attribute-level embedding strategy, well-learned signals remain largely embedded and recoverable across predefined external conditions in MMPD [Tang et al. 2023], as shown in Table 3. Benchmark methods also demonstrate the same domain gaps in this comparison. The signals persist across skin tone 6 – typically the most challenging skin tone – as well as incandescent and natural lighting. Few biased cases predominantly occur for olive and mixed yellow-brown tones (i.e., skin tone 4 and 5). Since our direct modulation operates merely on the green channel, skin tones and lighting conditions that reduce green reflectance and thus the amplitude of the modulation base inherently weaken the modulation effects in our strategy. Hence, a similar effect is observed under warm and vibrant LED-high lighting. Furthermore, we conduct an ablation study on relighting effects by applying generated, evenly distributed white, cool, and warm lighting, as well as a real-world environment map captured in a hospital. Metrics remain largely consistent with those under original lighting with a mean ΔMAE of +0.20 bpm, demonstrating the robustness of our method under the tested relit conditions. More details and future directions are in the supplementary material.

Table 3: Lights and skin tones comparison within MMPD [Tang et al. 2023], tested with models pretrained on UBFC-rPPG [Bobbia et al. 2019].

	POS		TS-CAN (MA)		FactorizePhys	
	MAE ↓	MAPE ↓	MAE ↓	MAPE↓	MAE ↓	MAPE↓
Skin tone						
3	1.01 ± 0.82	1.80 ± 1.52	1.04 ± 0.44	1.35 ± 0.52	2.89 ± 1.48	4.66 ± 2.38
4	1.91 ± 1.44	2.33 ± 1.82	2.17 ± 1.11	2.35 ± 1.11	2.89 ± 1.88	3.61 ± 2.41
5	3.80 ± 2.34	7.44 ± 4.76	0.62 ± 0.34	0.87 ± 0.45	4.90 ± 2.89	9.50 ± 5.81
6	0.47 ± 0.21	0.53 ± 0.25	0.10 ± 0.10	0.12 ± 0.11	0.10 ± 0.07	0.12 ± 0.09
Light						
LED-low	1.41 ± 0.94	2.42 ± 1.84	0.63 ± 0.24	0.81 ± 0.32	4.34 ± 2.69	6.91 ± 4.31
LED-high	3.40 ± 2.16	6.52 ± 4.24	0.37 ± 0.14	0.54 ± 0.21	3.30 ± 2.01	6.26 ± 3.97
Incandescent	0.42 ± 0.26	0.58 ± 0.38	0.98 ± 0.31	1.34 ± 0.41	2.26 ± 1.51	3.75 ± 2.66
Nature	0.90 ± 0.76	1.12 ± 0.96	1.89 ± 1.01	2.13 ± 1.12	1.65 ± 1.36	2.41 ± 2.06

References

Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois. 2019. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters* (2019). doi:10.1016/j.patrec.2017.10.017

Jitesh Joshi, Sos S. Agaian, and Youngjun Cho. 2024. FactorizePhys: Matrix Factorization for Multidimensional Attention in Remote Physiological Sensing. In *Advances in Neural Information Processing Systems*. 96607–96639. doi:10.52202/079017-3063

Bernhard Kerbl, Georgios Kopanas, Thomas Leimkühler, and George Drettakis. 2023. 3D Gaussian Splatting for Real-Time Radiance Field Rendering. *ACM Transactions on Graphics* 42, 4 (July 2023). doi:10.48550/arXiv.2308.04079

Xin Liu, Josh Fromm, Shwetak Patel, and Daniel McDuff. 2020. Multi-Task Temporal Shift Attention Networks for On-Device Contactless Vitals Measurement. In *Advances in Neural Information Processing Systems*. 19400–19411.

Xin Liu, Girish Narayanswamy, Akshay Paruchuri, Xiaoyu Zhang, Jiankai Tang, Yuzhe Zhang, Soumyadip Sengupta, Shwetak Patel, Yuntao Wang, and Daniel McDuff. 2023. rPPG-Toolbox: Deep Remote PPG Toolbox. arXiv:2210.00716

Hao Lu, Yuting Zhang, Jiaqi Tang, Bowen Fu, Wenhang Ge, Wei Wei, Kaishun Wu, and Yingcong Chen. 2025. RhythmGuassian: Repurposing Generalizable Gaussian Model For Remote Physiological Measurement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*.

Daniel J. McDuff and Ewa Magdalena Nowara. 2021. "Warm Bodies": A Post-Processing Technique for Animating Dynamic Blood Flow on Photos and Avatars. *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (2021).

Ben Mildenhall, Pratul P. Srinivasan, Matthew Tancik, Jonathan T. Barron, Ravi Ramamoorthi, and Ren Ng. 2020. NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis. In *European Conference on Computer Vision (ECCV)*.

Akshay Paruchuri, Xin Liu, Yulu Pan, Shwetak Patel, Daniel McDuff, and Soumyadip Sengupta. 2024. Motion Matters: Neural Motion Transfer for Better Camera Physiological Measurement. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*. 5933–5942. doi:10.1109/WACV57701.2024.00583

Ronny Stricker, Steffen Müller, and Horst-Michael Gross. 2014. Non-contact Video-based Pulse Rate Measurement on a Mobile Service Robot. In *Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN 2014)*. 1056–1062. doi:10.1109/ROMAN.2014.6926392

Jiankai Tang, Kequan Chen, Yuntao Wang, Yuanchun Shi, Shwetak Patel, Daniel McDuff, and Xin Liu. 2023. MMPD: Multi-Domain Mobile Video Physiology Dataset. In *2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. 1–5. doi:10.1109/EMBC40787.2023.10340857

Qunfeng Tang, Zhencheng Chen, Rabab Ward, Carlo Menon, and Mohamed Elgendi. 2020. Synthetic photoplethysmogram generation using two Gaussian functions. *Scientific Reports* 10, 1 (2020), 13883. doi:10.1038/s41598-020-69076-x

Wenjin Wang, Albertus C. den Brinker, Sander Stuijk, and Gerard de Haan. 2016. Algorithmic Principles of Remote PPG. *IEEE Transactions on Biomedical Engineering* 64, 7 (2016), 1479–1491. doi:10.1109/TBME.2016.2609282

Zhen Wang, Yunhao Ba, Pradyumna Chari, Oyku Deniz Bozkurt, Gianna Brown, Parth Patwa, Niranjan Vaddi, Laleh Jalilian, and Achuta Kadambi. 2022. Synthetic Generation of Face Videos With Plethysmograph Physiology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 20587–20596. doi:10.1109/CVPR52688.2022.01993

Zitong Yu, Yuming Shen, Jingang Shi, Hengshuang Zhao, Philip Torr, and Guoying Zhao. 2022. PhysFormer: Facial Video-based Physiological Measurement with Temporal Difference Transformer. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*. doi:10.1109/CVPR52688.2022.00415

Dongbin Zhang, Yunfei Liu, Lijian Lin, Ye Zhu, Kangjie Chen, Minghan Qin, Yu Li, and Haoqian Wang. 2025. HRAvatar: High-Quality and Relightable Gaussian Head Avatar. In *Proceedings of the Computer Vision and Pattern Recognition Conference (CVPR)*. 26285–26296. doi:10.1109/CVPR52734.2025.02448