Remote Photoplethysmography (rPPG) Data Generation: A Concise Review

Ahmad Hmedeh[1] Frédéric Bousefsaf[1] Choubeila Maaoui[1]

[1]Université de Lorraine, LCOMS, Metz, France

ahmad.hmedeh@univ-lorraine.fr, frederic.bousefsaf@univ-lorraine.fr, choubeila.maaoui@univ-lorraine.fr

Abstract

Remote photoplethysmography (rPPG) enables non-contact physiological monitoring from video by exploiting subtle color variations induced by blood volume changes in the skin. Despite recent progress in model architectures, rPPG performance remains fundamentally limited by dataset bias, including face-centric data collection, anatomical inconsistency between imaged regions and reference signals, and insufficient demographic and motion diversity. These limitations hinder generalization beyond controlled facial scenarios.

This paper presents a concise review of recent synthetic data generation approaches for rPPG and examines how they address structural shortcomings of existing datasets. We analyze representative strategies spanning signal-level synthesis, video-based generation and editing, motion-oriented augmentation, biophysically grounded rendering, and fairness-driven data augmentation. Across studies, synthetic data consistently improves robustness, reduces domain shift, and mitigates skin-tone bias.

Overall, we argue that synthetic data generation is becoming an increasingly essential component for developing scalable, anatomically informed, and clinically relevant rPPG systems.

Keywords: Remote photoplethysmography, synthetic data generation, physiological signal estimation, video-based health monitoring, dataset bias, fairness

1 Introduction

Remote photoplethysmography (rPPG) is a non-contact technique that estimates physiological signals from subtle temporal variations in skin color captured by RGB cameras. These variations originate from blood volume changes in superficial vessels and rely on the same optical principles as contact-based photoplethysmography (PPG). Unlike contact PPG, however, rPPG enables unobtrusive and scalable monitoring without physical sensors, opening new opportunities for everyday health applications [1].

The optical basis of rPPG is commonly described by the Dichromatic Reflection Model (DRM), which decomposes reflected light into specular and diffuse components [2]. Specular reflection mainly conveys environmental illumination, whereas diffuse reflection is modulated by hemoglobin absorption and carries the pulsatile information of interest. In practice, rPPG algorithms recover physiological signals by analyzing subtle temporal color variations across video frames and applying signal processing or learning-based techniques to suppress noise.

Methodologically, the field has evolved from handcrafted signal transformations to deep learning. Early approaches such as ICA [3], CHROM [4], and POS [5] are computationally efficient but highly sensitive to motion, illumination changes, and assumptions about skin reflectance. More recent deep learning models, including DeepPhys [6], PhysNet [7], and TS-CAN [8], improve robustness by learning spatiotemporal representations directly from video. Transformer-based architectures such as PhysFormer [9] further enhance temporal modeling, while self-supervised approaches aim to reduce dependence on labeled physiological signals.

Yet despite these architectural advances, rPPG performance remains strongly constrained by the data on which models are trained. The pulse signal is intrinsically weak and easily masked by non-physiological variations such as motion artifacts, illumination changes, and video compression. Skin tone introduces an additional and well-documented source of bias: differences in melanin concentration alter light absorption and reduce the signal-to-noise ratio, which translates into systematic performance disparities across populations [10]. Crucially, these limitations cannot be resolved through model design alone—they reflect the limited scale and diversity of existing public rPPG datasets.

This is precisely where data becomes the bottleneck. Collecting large, demographically balanced datasets with synchronized ground-truth physiological signals is costly and logistically restrictive, which in turn limits model generalization and reproducibility. Synthetic data generation has therefore emerged as a promising complement to real recordings: by enabling controlled variation of physiological signals, appearance, motion, and environmental conditions, it offers a practical path to mitigate data scarcity and bias. The present work reviews recent approaches for generating synthetic rPPG data and examines how they support the training of more robust and generalizable rPPG models.

2 Dataset Bias and the Need for Synthetic Data in rPPG

Most publicly available rPPG datasets are strongly face-centric and rely on fingertip photoplethysmography (PPG) signals as ground truth, implicitly assuming that pulse waveforms are anatomically invariant. Widely used benchmarks such as PURE [11], UBFC-RPPG [12], VIPL-HR [13], and VITAL [14] focus almost exclusively on facial video recordings while supervising learning with

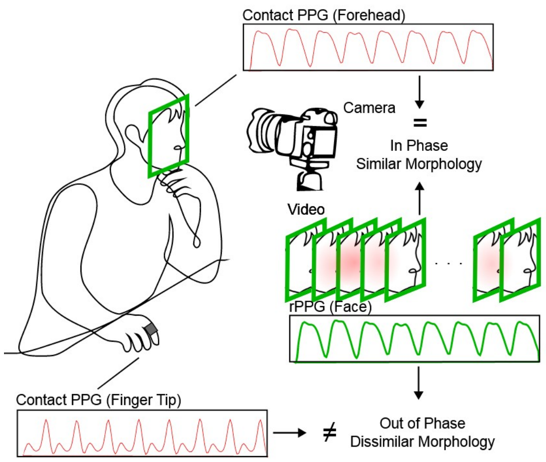


Figure 1: Difference in waveform shape and phase between contact-based PPG signals measured at the fingertip and at the face, highlighting the impact of anatomical location on pulse morphology. Extracted from [15].

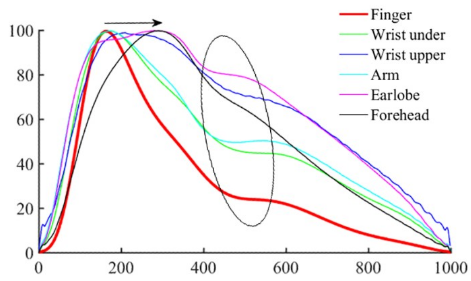


Figure 2: Normalized and averaged PPG waveforms from six anatomical sites under normal breathing conditions, illustrating site-dependent variability in pulse morphology. Extracted from [16].

distal reference measurements. Although these datasets incorporate variations in motion, illumination, and acquisition devices, they introduce a systematic physiological mismatch between the imaged region and the reference signal.

Physiological evidence directly contradicts this assumption. As illustrated in Fig. 1, McDuff et al. [15] demonstrate that PPG signals measured at the fingertip and at the face differ substantially in both phase and waveform morphology due to variations in pulse transit time and local vascular properties. These differences are not marginal artifacts but reflect fundamental cardiovascular dynamics. Consequently, rPPG models trained using anatomically consistent forehead PPG supervision achieve significantly lower estimation errors than those trained with fingertip signals when applied to facial videos, with reported mean squared error reductions of up to 40%.

Beyond the face–finger discrepancy, PPG morphology varies significantly across body regions. Fig. 2 reports normalized and averaged PPG waveforms from multiple anatomical sites, as shown by Hartmann et al. [16]. Systematic differences are observed in signal amplitude, pulse peak timing, and waveform shape across regions such as the finger, wrist, arm, leg, and forehead ($p < 0.001$). These findings confirm that PPG signals are inherently site-dependent and that models trained with reference signals from a single anatomical location are unlikely to generalize across regions.

An important and often overlooked consequence of this dataset design is the practical infeasibility of developing learning-based rPPG models for non-facial anatomical regions. Because the overwhelming majority of public datasets provide supervision exclusively for facial recordings, there is a lack of annotated data and evaluation benchmarks for regions such as the hands, arms, or torso. As a result, rPPG outside the face is not merely underexplored but structurally constrained by data availability. In the absence of region-specific supervision, model training and validation for non-facial rPPG become ill-defined, limiting progress beyond facial applications.

Only a limited number of datasets explicitly address this anatomical inconsistency. Notably, the MSPM dataset [17] provides full-body videos synchronized with multi-site PPG measurements, enabling the analysis of region-specific pulse morphology and pulse transit time. However, such datasets remain the exception rather than the norm.

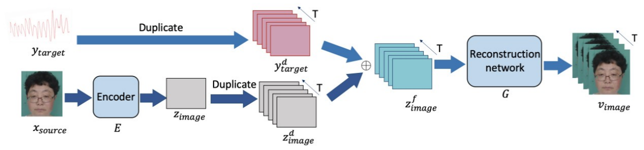
As a consequence, several dominant failure modes of rPPG—limited anatomical generalization, sensitivity to motion artifacts, and demographic bias—cannot be resolved through additional real data collection alone. Most datasets lack region-consistent physiological supervision and realistic motion diversity, while demographic imbalances further exacerbate performance disparities. In this context, data augmentation and synthetic data generation should be viewed as increasingly essential modeling tools rather than optional enhancements. By enabling controlled variation of anatomical regions, motion dynamics, and skin tone while preserving physiological consistency, synthetic data provides a practical mechanism to overcome structural limitations inherent to existing rPPG datasets.

3 Synthetic Data Generation for rPPG

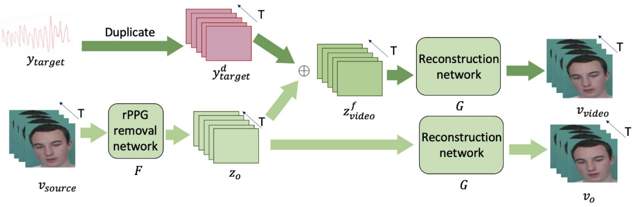
The structural limitations of existing rPPG datasets—namely anatomical bias, sensitivity to motion, and demographic imbalance—have motivated a growing body of work on synthetic data generation. Rather than forming a single methodological paradigm, existing approaches differ substantially in the level at which data is synthesized and in the specific failure modes they aim to address. This section reviews recent and representative methods that directly support the training and generalization of learning-based rPPG models, with a focus on signal-level control, video-level generation, robustness to motion, biophysical realism, and demographic fairness.

3.1 Signal-Level PPG Synthesis for Physiological Control

Early synthetic data generation approaches focused on one-dimensional photoplethysmography (PPG) signal synthesis to model pulse morphology and heart rate variability under controlled conditions. Tang et al. [18] proposed a parameterized model in which each cardiac cycle is represented as the sum of Gaussian functions corresponding to systolic and diastolic components, enabling controllable and physiologically plausible wave-



(a) Image-to-Video Network.



(b) Video-to-Video Network.

Figure 3: Two architectures from Tsou et al. [22]: (a) Image-to-Video Network and (b) Video-to-Video Network.

form generation. Related Gaussian-based and stochastic formulations have been explored in prior work [19, 20].

Beyond waveform simulation, signal-level synthesis has also been leveraged for representation learning. Song et al. [21] used synthetic PPG signals derived from ECG or BVP recordings to pretrain convolutional neural networks for heart rate estimation, showing improved cross-dataset generalization after fine-tuning on real rPPG data. While these approaches provide precise control over physiological parameters and are useful for pretraining or benchmarking, they do not address video-level challenges such as motion, illumination variation, or skin reflectance. As a result, their impact on end-to-end rPPG estimation remains limited.

3.2 Video-Level rPPG Embedding and Editing

To bridge the gap between synthetic physiological signals and video-based learning, several works have proposed generating or modifying facial videos to embed controlled rPPG signals. Tsou et al. [22] introduced a multi-task learning framework that jointly performs rPPG estimation and video-based data augmentation. Their approach relies on two complementary generative pipelines, illustrated in Fig. 3. The Image-to-Video network synthesizes a facial video from a single static image and a target rPPG waveform, enabling large-scale data generation without requiring video input. In contrast, the Video-to-Video network first removes the existing physiological signal from a real video before embedding a new target rPPG signal, allowing controlled physiological manipulation while preserving appearance and motion.

By coupling rPPG estimation with both generation paradigms, this framework enables end-to-end learning that enforces physiological consistency in the synthesized videos. Experimental results on COHFACE, PURE, and UBFC-RPPG demonstrate measurable reductions in estimation error across all three datasets when training with synthetic data. Nevertheless, the realism of the generated videos remains constrained by limited skin texture detail and the absence of fine-grained micro-expressions.

3.3 Motion Transfer and Robustness-Oriented Augmentation

While the previous approaches focus on embedding physiological signals into existing or synthesized videos, a complementary line of work targets the dominant source of error that affects rPPG in unconstrained environments: motion. Neural motion transfer has therefore emerged as an effective augmentation strategy to improve robustness. Paruchuri et al. [23] applied Face-Vid2Vid to transfer both rigid head motion and non-rigid facial expressions between videos while preserving the underlying physiological signal. Extensive cross-dataset evaluations showed error reductions of up to 78.9% in mean absolute error, highlighting the effectiveness of motion-centric augmentation for improving generalization.

Benezeth et al. [24] extended this paradigm to challenging scenarios such as near-infrared imaging and fitness activities. By animating static facial images using real motion trajectories and embedding synthetic rPPG signals, their approach significantly improved performance under low-light and high-motion conditions. While motion transfer effectively addresses robustness to movement, it does not explicitly model blood volume dynamics or waveform morphology, and therefore primarily benefits heart rate estimation rather than more detailed physiological inference.

3.4 Biophysical and Physically-Based Rendering Approaches

Biophysical modeling provides a principled direction for synthetic rPPG generation by explicitly simulating light–skin interaction and blood flow dynamics. McDuff et al. [25] introduced a physically based avatar rendering pipeline, illustrated in Fig. 4, that synthesizes facial videos with dynamic blood flow signals grounded in the Dichromatic Reflection Model.

As shown in Fig. 4, the pipeline modulates subsurface scattering properties using an input pulse signal while jointly controlling facial appearance, illumination, background, and motion. Training rPPG models with a combination of real and synthetic avatar data improved cross-dataset generalization and substantially reduced estimation error for darker skin tones. Subsequent work scaled this paradigm through fully synthetic datasets such as SCAMPS [26]. Wang et al. [27] further proposed a scalable framework combining 3D morphable face models and UV-space blood volume modulation to generate realistic rPPG videos from arbitrary input images, demonstrating consistent accuracy gains across demographic groups. Despite their advantages, biophysical approaches remain computationally expensive and subject to a simulation-to-real gap.

3.5 Fairness-Oriented and Sim-to-Real Augmentation

Demographic bias, particularly related to skin tone, remains a critical challenge in rPPG research. Ba et al. [28] proposed a bio-realistic skin tone translation framework that explicitly preserves pulsatile blood volume variations through joint optimization, as illustrated in Fig. 5.

The framework alternates between a generation phase, where skin tone is modified while preserving physiological

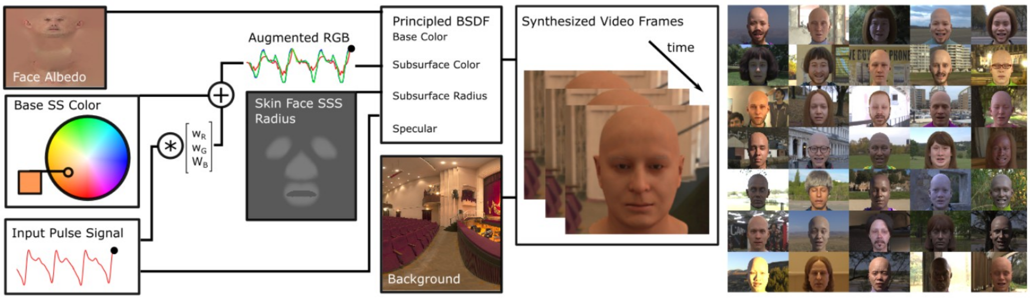


Figure 4: Overview of the physically based avatar rendering pipeline proposed by McDuff et al. [25]. Subsurface scattering parameters are modulated by an input pulse signal while facial albedo, skin tone, lighting, background, and head motion are controlled through principled BSDF-based rendering.

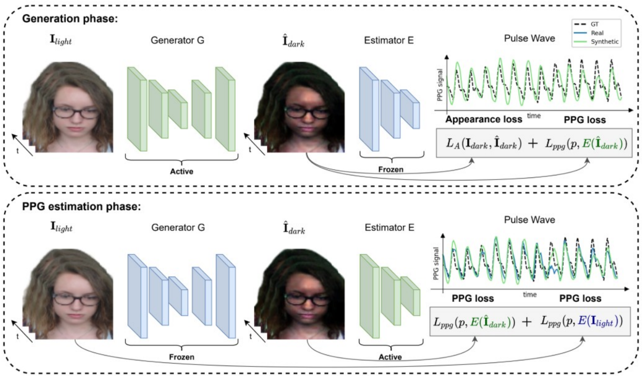


Figure 5: Joint optimization framework proposed by Ba et al. [28]. The generator translates facial videos to darker skin tones while the rPPG estimator is alternately frozen and optimized, ensuring that appearance changes do not suppress physiological signals.

cues, and an estimation phase, where the rPPG model is optimized on both original and tone-translated videos. This strategy significantly improves rPPG estimation accuracy across Fitzpatrick skin tone groups without requiring additional data collection. Building on this idea, Comas et al. [29] introduced PhysFlow, which leverages conditional normalizing flows to disentangle skin tone from facial video content in a continuous color space, further improving robustness on darker skin tones.

To mitigate the simulation-to-real gap inherent in fully synthetic data, Ou et al. [30] proposed realism-oriented augmentation strategies that inject compression artifacts, motion noise, and signal perturbations into synthetic videos. When applied to models trained on synthetic data, these augmentations yielded substantial error reductions across multiple real-world datasets, underscoring the importance of domain-aware augmentation for effective sim-to-real transfer.

4 Discussion and Perspectives

Despite recent progress in synthetic data generation for remote photoplethysmography (rPPG), current approaches remain constrained by structural limitations that restrict their scalability and applicability. A central issue is the strong anatomical bias toward facial signals. Nearly all existing datasets and augmentation pipelines focus exclusively on the face due to its favorable optical properties and ease of acquisition. While this has enabled significant advances in heart rate estimation, it inherently limits the extension of rPPG methods to other anatomical regions.

This limitation is particularly critical for applications requiring localized physiological assessment, such as peripheral perfusion monitoring or vascular health evaluation. PPG signals vary substantially across body regions due to differences in skin structure, perfusion depth, and motion characteristics. However, the absence of region-specific datasets and the lack of synthetic generation frameworks that explicitly model these variations constitute a structural barrier. Without anatomically consistent supervision or simulation tools, developing robust rPPG models beyond the face remains ill-defined.

A second major constraint concerns the scalability of physically based simulation approaches. Rendering pipelines that explicitly model light–skin interaction and blood flow dynamics [25] provide high-fidelity synthetic data but incur substantial computational cost, making large-scale dataset generation impractical. This trade-off between realism and scalability limits the use of purely physics-driven methods for training modern data-intensive rPPG models.

In parallel, although modern generative models have demonstrated strong potential in computer vision, their adoption for rPPG data generation remains limited in scope. Most existing efforts focus on facial augmentation or simplified waveform synthesis, without fully exploiting recent advances in video generation. In particular, diffusion-based and temporally coherent video generative models remain largely unexplored for rPPG, despite their potential to provide improved realism, temporal stability, and controllable synthesis of subtle physiological signals.

Looking forward, a key research direction lies in the development of region-aware synthetic data generation frameworks. Future approaches should condition generation not only on pulse characteristics but also on anatomical and perfusion-related parameters. Hybrid pipelines that combine lightweight generative video models with approximate biophysical constraints may offer a practical balance between realism and efficiency. Overall, advancing synthetic data generation for rPPG will require a shift from face-centric augmentation toward scalable, anatomically informed, and context-aware generation strategies.

5 Conclusion

This paper reviewed recent advances in synthetic data generation for remote photoplethysmography (rPPG), with a focus on their role in addressing the limitations of existing datasets. Although rPPG has shown strong potential for non-contact physiological monitoring, its progress remains constrained by limited data diversity in terms of anatomical regions, motion conditions, and demographic representation. Our analysis highlights that most available datasets and generation strategies remain heavily face-centric, despite well-established evidence that rPPG signal characteristics vary across body regions.

We reviewed representative synthetic data approaches—spanning signal-level modeling, video-level generation, motion transfer, biophysical rendering, and realism-oriented augmentation—that have been proposed to mitigate specific failure modes such as motion sensitivity, domain shift, and skin-tone bias. Experimental results across multiple studies consistently demonstrate that incorporating synthetic data can significantly improve robustness and generalization of rPPG models. However, important challenges persist, including scalability constraints, computational cost, and limited support for anatomically diverse or region-aware data generation.

Overall, advancing rPPG toward broader and more clinically relevant applications will require a shift from face-centric, task-specific augmentation toward scalable, anatomically informed, and physiologically constrained synthetic data pipelines. Such a transition will be key to enabling robust rPPG systems that generalize beyond controlled settings and support the diverse use cases envisioned for camera-based physiological sensing.

References

[1] Philipp V. Rouast et al. "Remote Heart Rate Measurement Using Low-Cost RGB Face Video: A Technical Literature Review". In: *Frontiers of Computer Science* 12.5 (2018), pp. 858–872. DOI: 10.1007/s11704-016-6243-6.

[2] Chun-Hong Cheng et al. "Deep Learning Methods for Remote Heart Rate Measurement: A Review and Future Research Agenda". In: *Sensors* 21.18 (2021), p. 6296. DOI: 10.3390/s21186296.

[3] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam". In: *IEEE Transactions on Biomedical Engineering* 58.1 (2011), pp. 7–11. DOI: 10.1109/TBME.2010.2086456.

[4] Gerard de Haan and Vincent Jeanne. "Robust Pulse Rate From Chrominance-Based rPPG". In: *IEEE Transactions on Biomedical Engineering* 60.10 (2013), pp. 2878–2886. DOI: 10.1109/TBME.2013.2266196.

[5] Wenjin Wang et al. "Algorithmic Principles of Remote PPG". In: *IEEE Transactions on Biomedical Engineering* 64.7 (2017), pp. 1479–1491. DOI: 10.1109/TBME.2016.2609282.

[6] Weixuan Chen and Daniel McDuff. "DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018, pp. 356–373. DOI: 10.1007/978-3-030-01216-8_22.

[7] Zitong Yu, Xiaobai Li, and Guoying Zhao. "Remote Photoplethysmograph Signal Measurement from Facial Videos Using Spatio-Temporal Networks". In: *arXiv preprint arXiv:1905.02419* (2019). DOI: 10.48550/arXiv.1905.02419.

[8] Xin Liu et al. "Multi-Task Temporal Shift Attention Networks for On-Device Contactless Vitals Measurement". In: *Advances in Neural Information Processing Systems (NeurIPS)*. 2020, pp. 19400–19411.

[9] Zitong Yu et al. "PhysFormer: Facial Video-Based Physiological Measurement with Temporal Difference Transformer". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 2022, pp. 4176–4186. DOI: 10.1109/CVPR52688.2022.00415.

[10] Ping-Kwan Man et al. "Blood Pressure Measurement: From Cuff-Based to Contactless Monitoring". In: *Healthcare* 10.10 (2022), p. 2113. DOI: 10.3390/healthcare10102113.

[11] Ronny Stricker, Steffen Muller, and Horst-Michael Gross. "Non-Contact Video-Based Pulse Rate Measurement on a Mobile Service Robot". In: *IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. 2014, pp. 1056–1062. DOI: 10.1109/ROMAN.2014.6926392.

[12] Serge Bobbia et al. "Unsupervised Skin Tissue Segmentation for Remote Photoplethysmography". In: *Pattern Recognition Letters* 124 (2019), pp. 82–90. DOI: 10.1016/j.patrec.2017.10.017.

[13] Xuesong Niu et al. "VIPL-HR: A Multi-Modal Database for Pulse Estimation from Less-Constrained Face Video". In: *Proc. Asian Conf. on Computer Vision (ACCV)*. Springer, 2018, pp. 562–576. DOI: 10.1007/978-3-030-20873-8_36.

[14] Pradyumna Chari et al. *Diverse R-PPG: Camera-Based Heart Rate Estimation for Diverse Subject Skin Tones and Scenes*. arXiv:2010.12769. 2020. DOI: 10.48550/arXiv.2010.12769.

[15] Daniel McDuff. *How Suboptimal is Training rPPG Models with Videos and Targets from Different Body Sites?* Mar. 15, 2024. URL: https://doi.org/10.48550/arXiv.2403.10582.

[16] Vera Hartmann et al. "Quantitative Comparison of Photoplethysmographic Waveform Characteristics: Effect of Measurement Site". In: *Frontiers in Physiology* 10 (2019), p. 198. DOI: 10.3389/fphys.2019.00198.

[17] Jeremy Speth et al. *MSPM: A Multi-Site Physiological Monitoring Dataset for Remote Pulse, Respiration, and Blood Pressure Estimation*. arXiv:2402.02224. 2024. DOI: 10.48550/arXiv.2402.02224.

[18] Qunfeng Tang et al. "Synthetic Photoplethysmogram Generation Using Two Gaussian Functions". In: *Scientific Reports* 10 (2020), p. 13883. DOI: 10.1038/s41598-020-69076-x.

[19] Diego Martín-Martínez et al. "Stochastic Modeling of the PPG Signal: A Synthesis-by-Analysis Approach with Applications". In: *IEEE Transactions on Biomedical Engineering* 60.9 (2013), pp. 2432–2441. DOI: 10.1109/TBME.2013.2257770.

[20] Lu Wang et al. "Multi-Gaussian Fitting for Pulse Waveform Using Weighted Least Squares and Multi-Criteria Decision Making". In: *Computers in Biology and Medicine* 43.11 (2013), pp. 1661–1672. DOI: 10.1016/j.compbiomed.2013.08.004.

[21] Rencheng Song et al. "Heart Rate Estimation From Facial Videos Using a Spatiotemporal Representation With Convolutional Neural Networks". In: *IEEE Transactions on Instrumentation and Measurement* 69.10 (2020), pp. 7411–7421. DOI: 10.1109/TIM.2020.2984168.

[22] Yun-Yun Tsou, Yi-An Lee, and Chiou-Ting Hsu. "Multi-Task Learning for Simultaneous Video Generation and Remote Photoplethysmography Estimation". In: *Computer Vision – ACCV 2020*. Springer, 2021, pp. 392–407. DOI: 10.1007/978-3-030-69541-5_24.

[23] Akshay Paruchuri et al. *Motion Matters: Neural Motion Transfer for Better Camera Physiological Measurement*. arXiv:2303.12059. 2023. DOI: 10.48550/arXiv.2303.12059.

[24] Yannick Benezeth et al. "Video-Based Heart Rate Estimation from Challenging Scenarios Using Synthetic Video Generation". In: *Biomedical Signal Processing and Control* 96 (2024), p. 106598. DOI: 10.1016/j.bspc.2024.106598.

[25] Daniel McDuff et al. *Advancing Non-Contact Vital Sign Measurement Using Synthetic Avatars*. arXiv:2010.12949. 2020. DOI: 10.48550/arXiv.2010.12949.

[26] Daniel McDuff et al. *SCAMPS: Synthetics for Camera Measurement of Physiological Signals*. arXiv:2206.04197. 2022. DOI: 10.48550/arXiv.2206.04197.

[27] Zhen Wang et al. "Synthetic Generation of Face Videos with Plethysmograph Physiology". In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, 2022, pp. 20555–20564. DOI: 10.1109/CVPR52688.2022.01993.

[28] Yunhao Ba et al. "Style Transfer with Bio-Realistic Appearance Manipulation for Skin-Tone Inclusive rPPG". In: *Proceedings of the IEEE International Conference on Computational Photography (ICCP)*. IEEE, 2022, pp. 1–12. DOI: 10.1109/ICCP54855.2022.9887649.

[29] Joaquim Comas et al. *PhysFlow: Skin Tone Transfer for Remote Heart Rate Estimation through Conditional Normalizing Flows*. arXiv:2407.21519. 2024. DOI: 10.48550/arXiv.2407.21519.

[30] Yuxuan Ou et al. *Training Robust Deep Physiological Measurement Models with Synthetic Video-Based Data*. arXiv:2311.05371. 2023. DOI: 10.48550/arXiv.2311.05371.