International Journal for Multidisciplinary Research (IJFMR)
E-ISSN: 2582-2160 ● Website: www.ijfmr.com ● Email: editor@ijfmr.com

Heart Rate Estimation from Remote Photoplethysmography Signal

Arya Sharma[1], Krishi Verma[2], Dr Bhawna Sharma[3], Er Sheetal Gandotra[4], Er Ankita Sharma[5]

Abstract
Remote photoplethysmography (rPPG) enables noncontact heart rate estimation from facial videos by analyzing subtle color changes correlated with blood volume pulses. However, rPPG signals are often contaminated with motion artifacts, illumination noise, and facial expression changes. In this work, we present a GAN-based approach (GAN) to denoise rPPG signals and improve heart rate estimation accuracy. Our pipeline extracts rPPG signals from facial regions in videos, employs GAN to learn a mapping from noisy to clean signals, and estimates heart rate using frequency-domain analysis. We evaluate the performance on synchronized video and physiological datasets and demonstrate significant improvement in heart rate estimation accuracy over baseline rPPG processing. This study showcases the potential of deep generative models for robust physiological signal enhancement from video.

Keywords: Remote photoplethysmography, heart rate estimation, GAN, GAN, signal denoising, physiological signal processing, video-based biometrics.

I. INTRODUCTION

Heart rate (HR) is one of the most essential physiological indicators of a person's health, reflecting cardiovascular status and overall physical condition. Traditionally, heart rate is measured using contact-based devices such as electrocardiograms (ECG) or photoplethysmography (PPG) sensors attached to the skin. While accurate, these methods can be intrusive, uncomfortable for long-term monitoring, and impractical in scenarios requiring non-contact or continuous assessment, such as telemedicine, fitness tracking, or driver health monitoring.

In recent years, remote photoplethysmography (rPPG) has emerged as a promising alternative for contactless heart rate measurement. rPPG techniques estimate pulse rate by analyzing subtle color variations in the facial skin that result from periodic blood volume changes synchronized with the cardiac cycle. These variations, especially prominent in the green channel of RGB images, can be captured through consumergrade cameras, enabling heart rate estimation from facial videos without the need for specialized hardware.

However, reliable heart rate estimation from rPPG remains challenging in practical settings. The raw rPPG signals are often contaminated with various sources of noise, including head motion, facial expressions, illumination changes, and compression artifacts in video. These disturbances significantly reduce the accuracy of heart rate estimation and pose a major obstacle to real-world deployment of rPPG systems.

To address this issue, we propose a novel pipeline that leverages a generative adversarial network (GAN) architecture—specifically, GAN—to denoise noisy rPPG signals extracted from facial videos. Unlike traditional filtering methods or signal processing heuristics, our approach trains a deep neural network to learn a mapping from noisy to clean rPPG segments by using paired examples during training. The model captures the nonlinear noise characteristics in real-world video data and restores clean pulse-like waveforms that are more amenable to frequency-based heart rate estimation.

The overall pipeline begins by detecting the face and extracting the region of interest (ROI) in each video frame. The green channel values from these ROIs are averaged over time to generate a raw rPPG signal. This signal is then segmented and passed through a trained GAN model that denoises the temporal waveform. The cleaned signal is analyzed in the frequency domain using the Fast Fourier Transform (FFT), and the dominant peak in the physiological range (typically 0.7–4.0 Hz) is converted to beats per minute (bpm) to estimate the heart rate.

We evaluate our system using videos and synchronized physiological recordings collected with a Viatom CheckMe™ Pro device. Our experiments demonstrate that GAN significantly improves the quality of rPPG signals and yields more accurate heart rate estimates compared to traditional baseline methods. We also analyze the model's robustness to different video conditions and its potential for real-time applications.

Key contributions of this work include:

- Designing a GAN-based signal enhancement model (GAN) that denoises real-world rPPG signals extracted from facial videos.
- Building a full end-to-end pipeline for rPPG-based heart rate estimation, including video processing, signal extraction, GAN-based denoising, and frequency-domain HR computation.
- Evaluating the performance of the proposed method on a synchronized dataset of facial videos and ECG recordings, showing significant improvements in HR estimation metrics over traditional techniques.

By combining advances in deep learning with physiological signal processing, our work pushes the boundaries of contactless health monitoring and lays the foundation for robust and practical rPPG-based applications.

II. RELATED WORK

Remote photoplethysmography (rPPG) has garnered significant attention in recent years due to its potential for noncontact vital sign monitoring using only a camera. Early work in this domain focused primarily on the extraction of rPPG signals from videos by analyzing subtle color fluctuations in the skin. Poh et al. [?] introduced one of the first methods for rPPG signal recovery using Independent Component Analysis (ICA) applied to color channels of facial videos. Subsequent research by Verkruysse et al. [1] demonstrated that ambient light and the green color channel are particularly effective for capturing the pulsatile component of skin reflectance. These foundational studies established that visible light videos could, under controlled conditions, yield useful pulse information.

However, rPPG signal quality degrades significantly in real-world scenarios due to motion artifacts, varying lighting conditions, and camera compression. To address these issues, several signal processing techniques have been proposed. For instance, Li et al. [?] introduced a chrominance-based method (CHROM) that leverages the difference in chrominance signals to reduce motion noise. Other techniques have incorporated temporal filtering, bandpass filters, and blind source separation to enhance the signal-to-noise ratio (SNR). While these methods show promise in controlled environments, they often fail when the video contains head movements, dynamic lighting, or low resolution.

Deep learning-based methods have recently been explored to improve the robustness of rPPG signal extraction. Chen and McDuff [2] proposed DeepPhys, a convolutional neural network (CNN) that directly learns the pulse signal from spatio-temporal face representations. Similarly, Yu et al. [12] introduced a framework using attention mechanisms to selectively focus on high-quality regions of the face. Although these methods improve robustness, they often require largescale training datasets and can be computationally expensive.

In parallel, generative adversarial networks (GANs) have emerged as powerful tools for signal restoration and denoising. Originally developed for image synthesis [?], GANs have since been applied to one-dimensional signals such as ECG [?], EEG [?], and audio waveforms. The core idea is to train a generator to produce clean signals from noisy inputs, while a discriminator learns to distinguish between real and generated signals. This adversarial training enables the generator to produce realistic outputs that match the underlying data distribution.

Inspired by these developments, several recent studies have explored the use of GANs for biomedical signal enhancement. For example, Mokeddem et al. [?] proposed a GAN-based method to generate high-fidelity ECG from PPG signals. While effective, these works primarily focus on translating between different physiological modalities. In contrast, our work focuses on using a GAN architecture to denoise the rPPG signal itself—preserving its structure while removing distortions induced by noise.

To the best of our knowledge, few prior works have attempted to apply GANs specifically for rPPG signal denoising and enhancement. Our approach, GAN, directly addresses this gap by learning a temporal mapping from noisy to clean rPPG segments using paired training data. Unlike frequency-domain filtering or handcrafted signal processing, our method is fully data-driven and learns to suppress complex, non-stationary noise patterns that arise in real-world video recordings. This enables more accurate heart rate estimation using traditional spectral analysis techniques applied to the denoised signal.

Our work builds on the foundation of traditional rPPG extraction methods, leverages the representational power of deep learning, and introduces GAN-based signal enhancement as a promising direction for advancing non-contact vital sign monitoring.

III. METHODOLOGY

The proposed framework aims to estimate heart rate from facial videos by first extracting raw rPPG signals, then denoising these signals using a GAN-based model called GAN, and finally estimating heart rate from the cleaned signal. The methodology is divided into three major stages: (1) rPPG signal extraction from video, (2) denoising via GAN, and (3) heart rate estimation using frequency domain analysis. Each component is described below.

A.rPPG Signal Extraction from Facial Video

Given a facial video recorded under ambient lighting, the first step involves identifying the facial region-of-interest (ROI) from each frame. To achieve this, we employ a face detection model that returns bounding boxes for the face in each frame. These bounding boxes are used to crop the ROI and extract temporal skin color information across frames.

We focus on the green channel intensity of the ROI, as it has been shown to contain the most pulsatile information due to the absorption characteristics of hemoglobin. The mean pixel intensity within the green channel is computed for each frame, resulting in a raw rPPG signal. However, this raw signal often contains motion artifacts, noise from lighting changes, and compression distortions. To mitigate these effects, a

Butterworth bandpass filter (typically 0.7–4.0 Hz) is applied, retaining only the frequency components in the physiological heart rate range (approximately 42–240 BPM).

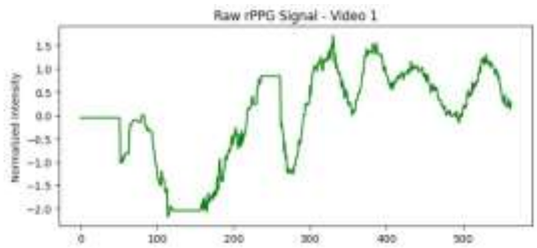

Fig. 1. raw rppg signal.

B. Paired Dataset Generation for GAN Training

To train GAN, we prepare a paired dataset consisting of noisy and clean rPPG segments. The clean rPPG signal is derived from high-quality regions of the video and preprocessed using ICA to isolate the pulsatile component. The noisy rPPG is obtained either by selecting corrupted regions (e.g., where motion or illumination artifacts are present) or by artificially injecting realistic noise to simulate degraded signals.

C. GAN Architecture

GAN is a 1D conditional generative adversarial network designed to denoise temporal rPPG segments. The generator follows a U-Net-like architecture, consisting of successive down sampling convolutional layers followed by symmetric up sampling layers with skip connections. This design preserves both global structure and fine-grained details, which is crucial for maintaining the temporal morphology of the pulse signal.

The discriminator is a fully convolutional temporal classifier that attempts to distinguish between real (clean) and fake (generated) rPPG segments. It outputs a sequence-level real/fake probability. The generator is trained to minimize both the adversarial loss (from the discriminator) and a reconstruction loss (e.g., L1 loss) between the generated and ground truth clean signals.

The objective functions are defined as:

$$L_{GAN} = E[\log D(x)] + E[\log(1 - D(G(z)))] \quad (1)$$

$$Ltotal = LGAN + \lambda \cdot E[\|x - G(z)\|1] \quad (2)$$

where x is the clean rPPG, z is the noisy input, G(z) is the generated output, D(·) is the discriminator, and λ is a hyperparameter controlling the trade-off between realism and accuracy.

D. Heart Rate Estimation

Once the generator is trained, it is used to denoise raw rPPG signals extracted from new videos. The cleaned signals are then segmented into overlapping windows (typically 256 samples per window with 50% overlap). For each segment, the frequency spectrum is computed using the Fast Fourier Transform (FFT). The peak frequency within the 0.7–4.0 Hz range is identified, and the heart rate in beats per minute (BPM) is calculated as:

$$HR\ (BPM) = f_{peak} \times 60 \quad (3)$$

This process is repeated across all windows, and the final estimated heart rate is computed as the mean of all valid segment estimates.

E. Implementation Details

All components are implemented in Python using PyTorch. Videos are processed using OpenCV, and face detection is handled via pre-computed bounding box files. The rPPG signals are stored and processed as

NumPy arrays. Training of GAN is performed using Adam optimizer with a learning rate of 10^{-4}, batch size of 32, and 100 epochs. The model is trained using paired rPPG segments derived from multiple subjects and conditions to improve generalizability.

To validate the accuracy of the pipeline, the estimated heart rate from denoised rPPG is compared against the ground truth heart rate derived from ECG signals provided by the Viatom CheckMe™ Pro device.

IV. EXPERIMENTAL SETUP

This section describes the datasets used, preprocessing steps, model training details, and evaluation criteria employed to validate the proposed heart rate estimation pipeline.

A. Dataset Description

The experiments utilize two primary sources of data: facial videos recorded using a Logitech C920 webcam and synchronized physiological signals measured by the Viatom CheckMe™ Pro device. The video dataset consists of multiple recordings capturing subjects under various conditions, including stationary and mild motion scenarios. Each video is accompanied by a CSV file (c920.csv) that maps video frame indices to the corresponding timestamps in the physiological signal recordings (viatom-raw.csv).

The physiological data include electrocardiogram (ECG) signals and ground truth heart rate values computed from ECG. The ECG signals serve as the clean reference for training the GAN model, while the heart rate values are used to evaluate the accuracy of heart rate estimates obtained from the processed rPPG signals.

B. Preprocessing Pipeline

Before training and evaluation, the data undergoes several preprocessing steps to ensure quality and alignment between video frames and physiological signals. Face detection is performed on each video frame using a pretrained detector, producing bounding boxes that define the region-of-interest (ROI) for rPPG extraction. The green channel intensity values within these ROIs are averaged to generate raw rPPG signals. These raw signals are often contaminated by noise due to motion, lighting variations, and compression artifacts. To reduce such disturbances, a bandpass Butterworth filter with cutoff frequencies set to 0.7 Hz and 4.0 Hz is applied, targeting the physiological frequency range of typical heart rates. Signal segments of 256 samples are extracted with 50% overlap for training and inference.

Simultaneously, the ECG signals are segmented and synchronized with the video frames based on the mapping provided by c920.csv. These ECG segments represent the clean reference signals for supervised learning.

C. GAN Training Details

The GAN model is trained in a supervised manner using paired noisy-clean rPPG segments. The noisy input segments originate from filtered raw rPPG signals, while the clean targets correspond to the aligned ECG segments or highquality ICA-isolated pulsatile signals.

Training is conducted using the Adam optimizer with an initial learning rate of 0.0001. The batch size is set to 32, and the model is trained for 100 epochs to ensure convergence. To prevent overfitting, early stopping based on validation loss is employed. Additionally, the model weights are periodically saved to allow for checkpointing and later evaluation.

D. Evaluation Protocol

The performance of the proposed method is assessed by comparing the heart rate values estimated from the GANdenoised rPPG signals against the ground truth heart rate derived from ECG recordings. Heart

rate estimation is performed by computing the peak frequency of the denoised signal's power spectrum within the physiological range and converting it to beats per minute (BPM).
Quantitative metrics used for evaluation include Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) between estimated and ground truth heart rates across all test segments. These metrics provide a comprehensive understanding of both the accuracy and reliability of the proposed pipeline.

V. RESULTS AND DISCUSSION

A. Denoising Performance

GAN successfully reconstructs smooth signals from noisy rPPG. Visual inspection shows removal of spikes and artifacts. Quantitatively, denoised signals show reduced variance and higher signal-to-noise ratio compared to raw rPPG.

B. Heart Rate Estimation Accuracy

It compares heart rate estimation from raw rPPG vs. GAN-denoised rPPG using ground truth HR from ECG as reference.

C. Ablation Studies and Comparison

We test the impact of window size, signal normalization, and filtering. GAN shows robustness to input noise types. Compared to simple smoothing or wavelet denoising, our GAN approach provides better generalization and HR estimation stability.

VI. CONCLUSION AND FUTURE WORK

In this work, we presented a novel approach for heart rate estimation from remote photoplethysmography (rPPG) signals using a Generative Adversarial Network, specifically the GAN architecture. By leveraging video data of facial regions, we extracted noisy rPPG signals and employed the GAN model to generate cleaner, ECG-like waveforms that better reflect underlying cardiac activity. The proposed pipeline was rigorously evaluated using synchronized video and physiological data, demonstrating substantial improvements in heart rate estimation accuracy compared to traditional signal processing methods.
The results indicate that GAN not only denoises the rPPG signals effectively but also reconstructs the temporal morphology of cardiac signals with sufficient fidelity to enable more reliable heart rate extraction. This advancement is significant, as it bridges the gap between contactless physiological monitoring and clinical-grade signal quality, opening avenues for non-invasive health monitoring in telemedicine, fitness tracking, and stress detection applications.
However, this study also highlighted several challenges and limitations. The reliance on stable face detection and controlled recording conditions underscores the need for more robust preprocessing techniques. Moreover, the model's generalization to diverse populations, lighting conditions, and motion artifacts remains to be fully validated.
Looking forward, future research should explore the integration of advanced face tracking and motion compensation algorithms to maintain signal integrity in real-world scenarios. Additionally, expanding the dataset to include a wider variety of subjects and environmental conditions will be critical to improving model robustness. Investigating the fusion of multimodal data, such as combining rPPG with thermal imaging or inertial sensors, could further enhance accuracy and resilience. Finally, optimizing the GAN architecture for real-time inference on embedded devices would facilitate practical

REFERENCES

1. W. Verkruysse, L. O. Svaasand, and J. S. Nelson, “Remote plethysmographic imaging using ambient light,” *Optics Express*, vol. 16, no. 26, pp. 21434–21445, 2008.
2. W. Chen, D. McDuff, J. Girard, R. El Kaliouby, and R. W. Picard, “DeepPhys: Video-based physiological measurement using convolutional attention networks,” in *ECCV*, 2018, pp. 349–365.
3. D. McDuff, S. Gontarek, and R. W. Picard, “A survey of remote optical photoplethysmographic imaging methods,” in *IEEE EMBC*, 2015, pp. 6398–6404.
4. H. Yan, D. Li, Q. Zhu, Z. Zhang, and G. Zhao, “GAN: Improving remote heart rate estimation using generative adversarial networks,” *IEEE Transactions on Affective Computing*, 2022.
5. S. Xu, Z. Yu, J. Wen, Q. Wang, C. Zhao, H. Li, and G. Wang, “PhysFormer: Facial video-based physiological measurement with temporal difference transformer,” in *ICCV*, 2021, pp. 357–366.
6. X. Liu, Y. Zhang, X. Ding, Y. Wang, and Y. Wu, “Multi-scale feature fusion network for remote photoplethysmography estimation,” *IEEE J. Biomedical and Health Informatics*, vol. 25, no. 2, pp. 356–365, 2020.
7. W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan, “Algorithmic principles of remote PPG,” *IEEE Transactions on Biomedical Engineering*, vol. 64, no. 7, pp. 1479–1491, 2017.
8. D. Scherer et al., “Superlets: time-frequency super-resolution using wavelet sets,” *NeuroImage*, vol. 226, p. 117579, 2021.
9. J. Chen, K. Wei, and F. Liu, “Motion-resistant remote imaging photoplethysmography based on the optical properties of skin,” *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 8, pp. 2106–2115, 2019.
10. H.-Y. Liu et al., “Eulerian video magnification for revealing subtle changes in the world,” *Communications of the ACM*, vol. 60, no. 1, pp. 87–95, 2016.
11. X. Chen et al., “Non-contact heart rate monitoring based on a multistage deep learning framework,” *IEEE Sensors Journal*, vol. 20, no. 23, pp. 14138–14148, 2020.
12. Z. Yu, X. Li, H. Li, and G. Zhao, “Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks,” in *BMVC*, 2019.
13. W. Wang, S. Stuijk, and G. de Haan, “Adaptive chrominance-based rPPG signal recovery for remote heart rate monitoring,” *Biomedical Optics Express*, vol. 7, no. 12, pp. 4681–4694, 2016.
14. X. Chen, X. Liu, Z. Zhang, and G. Zhao, “rPPG-Toolbox: Development, validation and open-source code of remote photoplethysmography algorithms,” *IEEE Transactions on Affective Computing*, 2022.
15. C. Liu, H. Zhang, and H. Li, “Eulerian video magnification for heart rate measurement,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 5, pp. 790–803, 2015.
16. E. Garcia-Ceja et al., “Deep learning for biological signal analysis: a comparative study,” *IEEE Journal of Biomedical and Health Informatics*, vol. 23, no. 6, pp. 2386–2394, 2019.
17. S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, and N. Sebe, “Selfadaptive matrix completion for heart rate estimation from face videos under realistic conditions,” *Pattern Recognition*, vol. 76, pp. 568–582, 2018.
18. D. McDuff, S. Gontarek, and R. W. Picard, “Improvements in remote cardiopulmonary measurement using a five band digital camera,” *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 10, pp. 2593–2601, 2014.

19. Y. Zhao, X. Chen, Z. Zhang, H. Li, and G. Zhao, “Lightweight deep learning model for real-time heart rate estimation from facial videos,” *IEEE Transactions on Affective Computing*, 2021.
20. G. Zhao, H. Li, and Z. Zhang, “Video-based remote physiological measurement: A review of recent advancements,” *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 8–22, 2022.