SkinMap: A Novel Weighted Full-Body Skin Segmentation for Robust Remote Photoplethysmography

Amirhossein Akbari*[1]
amirhoseinakbari@ee.sharif.edu

Zahra Maleki*[1]
zahra.maleki@ee.sharif.edu

Amirhossein Binesh[2]
ah.binesh@ce.sharif.edu

Babak Khalaj[1]
khalaj@sharif.edu

Department of {Electrical Engineering[1], Computer Engineering[2]}
Sharif University of Technology

Abstract

Remote photoplethysmography (rPPG) is an innovative method for monitoring heart rate and vital signs by recording a person with a simple camera, as long as any part of their skin is visible. This low-cost, contactless approach helps in remote patient monitoring, emotion analysis, smart vehicle utilization, and more. Over the years, various techniques have been proposed to improve the accuracy of this technology, especially given its sensitivity to lighting and movement. In the unsupervised pipeline, it is needed first to select skin regions from the video to extract the rPPG signal from the skin color change. We introduce a novel skin segmentation technique that is robust for real-world scenarios and prioritizes skin regions to enhance the quality of the extracted signal. It can detect areas of skin all over the body, making it more resistant to movement, while removing areas such as the mouth, eyes, and hair that may cause interference. Our model is evaluated on two public datasets, and we also present a new dataset called SYNC-rPPG to better represent real-world conditions. The results indicate that our model demonstrates a prior ability to capture heartbeats in challenging conditions, such as talking and head rotation, and maintain the mean absolute error (MAE) between predicted and actual heart rates, while other region of interest (ROI) selection methods fail to do so. In addition, it delivers comparable results in static scenarios and demonstrates high accuracy in detecting a diverse range of skin tones, making it a promising technique for real-world applications.

*Contributed equally to this work

1. Introduction

Remote photoplethysmography (rPPG) is an advanced non-contact technique that enables the measurement of vital physiological signals [49], such as heart rate (HR), respiration frequency (RF) and heart rate variability (HRV), by analyzing a video captured from any part of the skin surface. The light reaching the camera sensor has a periodic component that reflects variations in light absorption caused by changes in arterial blood volume [13, 17, 36]. This technology holds significant promise for applications in remote healthcare and emotion analysis [44], as it can capture data from any exposed area of skin without physical proximity. The extraction of the rPPG signal generally follows unsupervised methods that rely on a structured pipeline[26], where the area of the skin that is most likely to produce high-quality signals [18] is isolated using computer vision techniques [7, 18, 22, 37, 38, 46, 47, 51, 56]. Then conventional algorithms are applied to convert the RGB signal into the rPPG signal, such as LGI [35],POS [48], CHROM [11], PBV [14], PCA [21], OMIT [7], GREEN [29, 45], to extract the rPPG and estimate the heart rate. However, deep-learning based approaches have taken over many parts of processing. They either combine conventional techniques with deep learning models or provide end-to-end solutions [9, 12, 19, 25, 27, 32, 34, 40, 42, 53–55]. Unsupervised approaches tend to offer better generalization in different applications [24]. On the other hand, end-to-end supervised approaches predominantly learn to recognize facial noise patterns associated with the reference signal [10] and require dataset-specific training, which result in a lack of understanding of the underlying physiological mechanisms and are computationally expensive [56]. These limitations

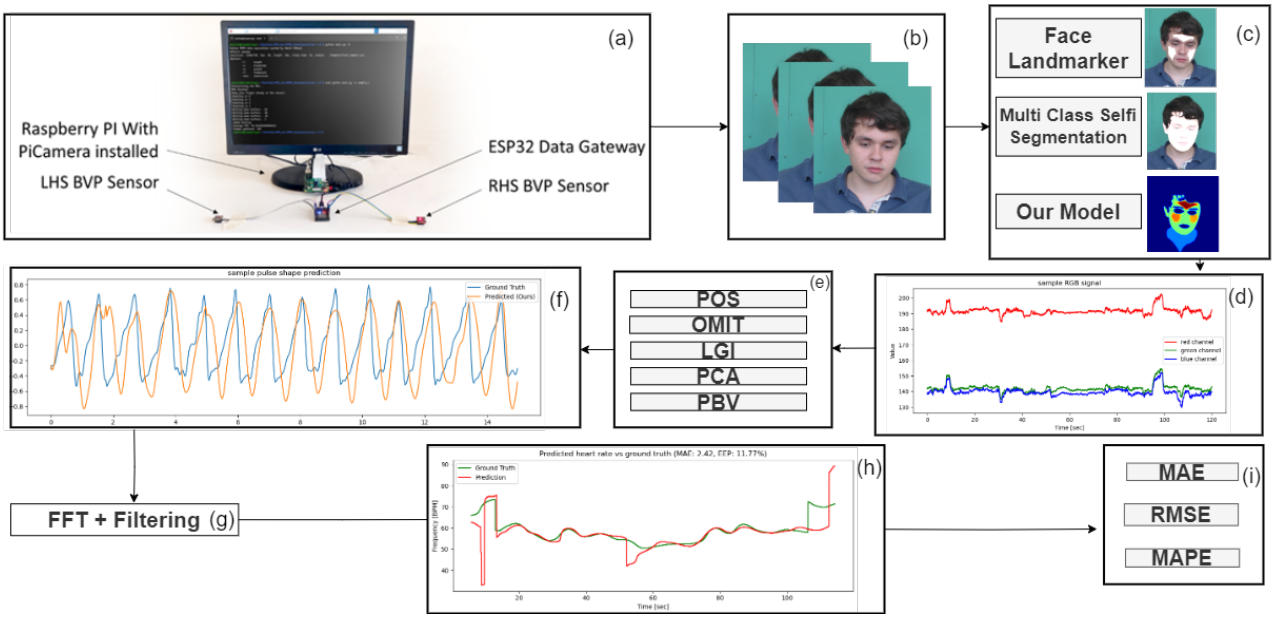

Figure 1. Unsupervised pipeline for heart rate estimation from video. (a) Data acquisition. (b) Video dataset collection synchronized with PPG signals. (c) Skin segmentation or ROI selection process. In this work, we compare two state-of-the-art models with our proposed segmentation approach. (d) RGB signal extraction by averaging skin pixels. (e) rPPG signal extraction methods are applied to the RGB signal. (f) Comparison of the extracted rPPG signal (orange) with the reference PPG pulse (blue). (g) Heart rate estimation. (h) Heart rate analysis over time. (i) Evaluation of our estimation using statistical metrics.

restrict their applicability in healthcare and real-time deployment on mobile devices.

Many similar studies proposing new unsupervised algorithms use face detection in combination with spatial averaging over the entire skin area as the region of interest (ROI). Color thresholding methods, such as YCbCr or HSV [20, 37], are also viable options when combined with face detection, although they are less effective in handling various skin tones and harsh lighting conditions. Face2PPG [7] has been introduced as a method to stabilize movement and facial expressions. However, this approach requires skin detection and geometric segmentation and remains limited to the face area and restricts the available skin regions. Many rPPG signal extraction algorithms rely on a well-defined, dynamic, weighted skin mask to improve rPPG signal reliability and robustness, and spatially-based techniques often prove to be effective [47]. There is a lack of research on dynamic approaches that utilize skin regions throughout the face and body, enabling a more robust signal extraction. This would reduce reliance on specific areas that can be blocked or compromised due to factors such as facial expressions, occlusions, or challenging lighting conditions [6, 33], ultimately offering a more versatile and reliable approach for the extraction of signals.

A key requirement for validating the robustness of rPPG methods is applying the pipeline to more realistic datasets. Although existing datasets provide video and ground truth signals [5, 30, 35, 39, 41], they often lack real-world complexity, such as significant head movements, dynamic facial expressions, varying lighting environments, and diverse physiological states. Some are limited to static scenarios with high-contrast backgrounds and require high-quality cameras for data acquisition, making them less accessible. In addition, the reference signal and the captured video are not synchronized in these datasets. Developing a dataset that incorporates diverse conditions while using affordable and widely available cameras would enhance the practical evaluation of rPPG signal extraction methods in more realistic cases.

Contributions

The contribution of this paper falls into two categories:

- We introduce novel skin segmentation model called SkinMap, capable of extracting both facial and body skin areas, generating a weighted mask that assigns higher weights to regions likely to produce higher-quality signals based on fundamental knowledge of rPPG signals, considering factors such as lighting and angle, without relying on any face detection or face landmark detection. The model is fine-tuned on a synthesized image dataset.
- We present a new dataset called SYNC-rPPG that captures data in four real-world scenarios. Data collection was done using an affordable camera and sensor with synchronized sampling rate.

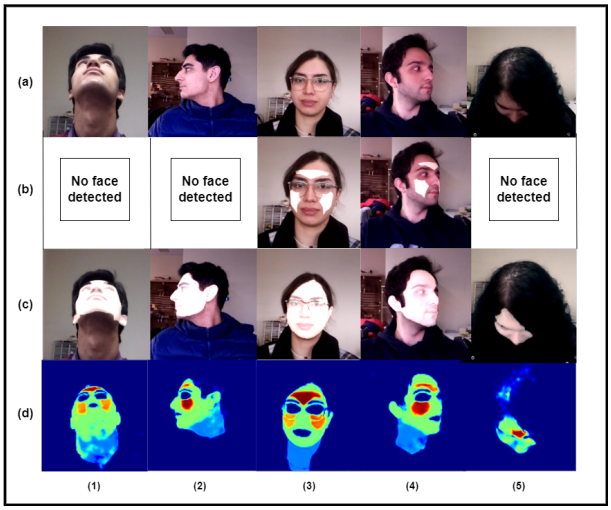

Figure 2. Illustration of our dataset and segmentation results. (a) Frame samples from the rotation task of our dataset. (b) Segmentation results using Face Landmark Detection, where white areas indicate detected ROIs. In some frames, the Landmarker failed to detect a face. (c) Segmentation results using the Multi-Class Selfie Segmentation, where white areas represent detected skin regions. (d) Heat-map visualization of the output of our segmentation model.

This work strives to improve accuracy while ensuring robustness, simplicity, and accessibility, making it applicable in real-world scenarios.

2. Methodology

As shown in Fig. 1, the unsupervised pipeline for extracting rPPG signals typically involves the following steps:

1. Dataset Collection and Pre-processing: This step involves collecting video data synchronized with a reference signal and providing the necessities to read and manage the available or collected dataset.
2. Video Processing: In this step, a skin segmentation or ROI selection technique is applied, followed by calculating the average or weighted average of the pixel values within the skin region to obtain the RGB signal throughout the video.
3. RGB to rPPG conversion: This step transforms skin color variations into physiological signals using algorithms that combine the RGB channels, signal processing, band-pass filtering, and de-noising to extract the rPPG signal.
4. Heart Rate Estimation: The heart rate is estimated by performing frequency analysis on the rPPG signal and comparing it with the reference signal.
5. Evaluation of results: The extracted rPPG signal is evaluated based on various metrics and statistical analysis to assess its accuracy and robustness.

Figure 3. Model output on a random sample from the COCO [23] dataset, showcasing its reliability in real-world applications.

We achieve the weighted average of skin areas by adopting a variant of the well-established DeepLabV3 architecture with a ResNet-50 backbone, leveraging its proven performance in semantic segmentation [8]. We expect a weighted average of skin areas to obtain the RGB signal. We replace the final layer of the default and auxiliary classifier of DeeplabV3 with a single channel convolutional layer. A sigmoid activation function is attached to the final layer to confine the output values between 0 and 1. After fine-tuning the model on a large and suitable dataset, we expect it to effectively segment all available skin areas while assigning the best possible weights based on the subject's position and lighting conditions in each frame of the video dataset. In the process of training and evaluating our model, we use two state-of-the-art MediaPipe skin segmentation methods: MediaPipe Face Landmark Detection [16], a real-time model that predicts 468 3D facial landmarks, and MediaPipe Multi-Class Selfie Segmentation, a Vision Transformer-based model designed for real-time segmentation of human subjects. It outputs segmentation masks that include the classification of background, hair, body skin, face skin, clothing, and accessories. However, it does not explicitly differentiate non-skin facial areas, such as the eyes, mouth, or glasses [28]. A comparison of the results from MediaPipe Land-marker, Multi-Class Selfie Segmentation, and our trained model is illustrated in Fig. 2. As shown, the Landmarker failed to detect the face at harsh angles and when it was not fully visible.

Training our segmentation model requires a diverse dataset of human images under various environmental conditions. Although there are some public skin segmentation datasets [1, 15, 52], they contain a very limited number of precise samples and are not suitable for our training [31]. To address this, we created a custom dataset by extracting human images from the COCO dataset [23], which offers a wide variety of real-world scenes with diverse backgrounds and skin tones. In the process of signal extraction, the eye region is prone to excessive movement, which introduces noise into the signal [18, 20]. Meanwhile, the cheeks and forehead exhibit the highest amplitude of the pulse signal [18]. Therefore, prioritizing the segmentation of these regions over other skin areas is essential for accurate signal extraction. We assign a weight to each region based on its

angle relative to the camera and its biological priority. By combining the MediaPipe models, we generate this comprehensive skin mask that includes all available skin regions relevant to pulse extraction. The proposed methodology uses MediaPipe models only to generate synthetic training data with weighted skin segmentation, which no existing photo dataset provides.

In the first step of generating our training dataset, we extract images containing humans with fully visible faces from COCO dataset using the MediaPipe Face Landmark Detector to identify and select relevant images. As described in Methodology, prioritizing the segmentation of some regions over other skin areas is essential for accurate signal extraction. To assign importance to different facial regions, we classify them into three priority levels: Priority 1 (forehead and cheeks, providing the highest-quality pulse signals), Priority 2 (other facial regions, excluding the eyes, eyebrows, and lips), and Priority 3 (other skin surfaces on the body). Regions with higher priority are given greater weight in the final skin mask. We introduce a weighting mechanism based on the angular orientation of the skin relative to the camera and the priority level. We assign a weight to each region based on its angle relative to the vertical axis. For Priority 1 regions, the weight varies between 4 and 2 depending on the angle, while Priority 2 and 3 regions are assigned fixed weights of 2 and 1, respectively. When the camera is perpendicular (90°) to the face, specular reflections are minimized [50]. we use facial landmarks to estimate the orientation of the face. The weighting function for Priority 1 regions is defined in equation Eq. (1). To explain Eq. (1) for the weighting of priority 1 regions in our synthesized photo dataset, we assign a weight P_i to each region based on its angle θ_i relative to the vertical axis. A lower angle between the surface's normal vector and the camera improves signal quality. We adjusted the ROI weighting so that smaller angles increase the weight, and if the angle exceeds the threshold, the weight matches the surrounding area. This results in a weight curve ranging from 2 (facial areas) to 4, with a 45 degree angle increasing to 3 [50]. The cosine function, used for the effective area, smooths the curve and minimizes noise in challenging poses.

$$P_i = \begin{cases} 3 + \left(2\cos\left(\frac{3}{2}\theta_i\right) - 1\right), & \text{if } |\theta_i| < \frac{\pi}{3} \\ 2, & \text{otherwise} \end{cases} \quad (1)$$

After that, the assigned weights are normalized between 0 and 1 to maintain consistency with the network output scale. We apply this process to selected human images from COCO, creating a suitable dataset for training that consists of 8,000 images with reliable ground-truth masks for skin segmentation and weighting. This dataset is used to train our model for accurate skin segmentation.

Fig. 3 presents the results of the trained model on randomly selected images from the COCO dataset [23]. The results demonstrate robustness to skin tone variations and model capability to produce a normalized full-body skin mask based on the proposed priority-based weighting. We trained our DeepLabV3-ResNet50 model for 30 epochs, using 90 percent of the data for training and 10 percentage for validation. The final RGB signal of each video is processed using five commonly used rPPG algorithms, following their implementations in [26]. The extracted rPPG signal is then used to estimate heart rate. Heart rate is determined using the Fourier transform (FFT), and band-pass filtering, which extracts frequency components within the physiological heart rate range (0.5 to 3.2 Hz). The strongest frequency in this range is identified as the heart rate in beats per minute (BPM).

3. Experiments

We apply the pipeline illustrated in Fig. 1 to the three introduced segmentation models, including SkinMap, MediaPipe Selfie Segmentation (MCSS) and Face Landmark Detection. The comparison is conducted on our dataset (SYNC-rPPG), UBFC-rPPG [5], and UBFC-PHYS [30]. Data were collected from 20 individuals, each video lasting 30 seconds. Each subject participated in four different scenarios. In the first, the subject remains calm with no movement. In the second, the subject is asked to read something emotional out-load or talk about an important memory of them. In the third, the subject performs rapid head rotations. In the fourth, the recording takes place after three minutes of exercise. As shown in [7], the PPG signals of fingertip in public datasets can flip due to movement or disconnections, causing heart rate errors. Our dataset uses two sensors, averaging their signals for reliability. If one sensor disconnects, we discard its data to avoid affecting the results. In Tab. 1 in supplementary material, we compare SYNC-rPPG with existing datasets. We should note that age is less critical for rPPG extraction than the BPM range [2, 3, 43]. We offer a wider variation of BPM by including the post-exercise recovery task. SYNC-rPPG includes luminosity variations to account for changes in lighting conditions.

To evaluate the extracted heart rate, we use the mean absolute error (MAE), the root mean square error (RMSE), and the mean absolute percentage error (MAPE) [20]. Since the PPG signal at fingertip has a natural delay compared to the face and neck rPPG signals [4], the Pearson Correlation Coefficient (PCC) may not be meaningful. Therefore, we compute PCC across all time shifts within one second and define MPCC as the maximum value obtained.

One way to evaluate models is by measuring the average frames where they fail to adjust a mask. This issue is more common with ROI-based models during head rotation, as seen in Fig. 2. In our dataset, Face Landmark Detection misses the average of 0.75 frames in talking and 118 frames in rotation tasks in each video, while SkinMap and MCSS

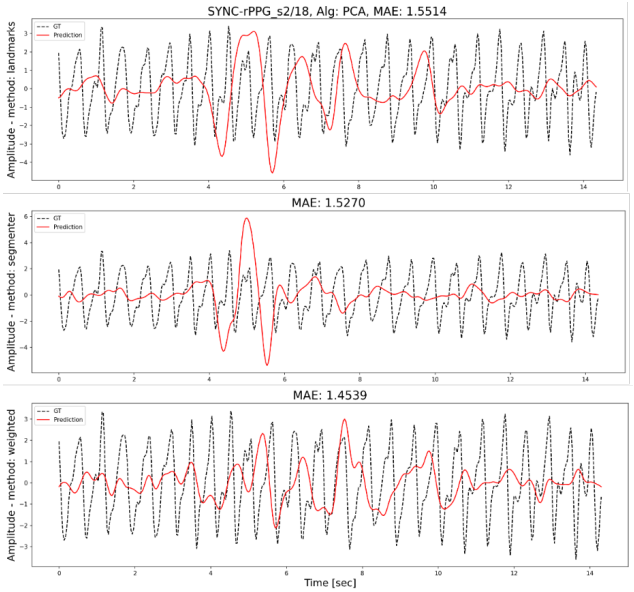

Figure 4. Extracted signals using models, up: Face Landmark Detection, middle: Multi-Class Selfie Segmentation, down: SkinMap. Black line: ground truth. Red line: prediction

perform flawlessly. As illustrated in Fig. 4, extracted signals from one of the samples in our dataset are shown. For the rotation task, it is evident that our model could reconstruct the true shape and peaks of the signal much better, whereas in several cases the Face Landmark failed to detect the face as shown in Fig. 2.

SkinMap outperformed other methods in SYNC-rPPG dataset and UBFC-PHYS, which contain more real-life scenarios. Meanwhile, the Landmark Detection and Multi-Region [7] performed better in UBFC, which consists of more static cases. SkinMap maintains precision and low error margins in challenging scenarios, where MCSS (representing face skin without weight) and Landmarker (ROI selection) fail. In talking and rotation scenarios, our model outperforms others in both RMSE and MPCC. A higher MPCC value indicates similarity between the extracted pulse signal and ground truth. This evaluation suggests that, while simple ROI selection suffices for static conditions, it is insufficient for real-life applications. For practical use, models must use all available sources of information to be reliable. In addition, some approaches, including Multi-Region, integrate several components, including face detection, face alignment, and landmark detection, before ROI selection [7]. However, our model does not require any additional face detection or extensive pre-processing.

We analyze the segmentation accuracy of our model and its diversity with and without weights in different skin tones using [57], which provides annotations (light, dark, unsure, nan) for the COCO dataset. We used 10 % of our synthesized dataset for validation. The weighted mask achieves a mean accuracy of 0.97 and a mean F1 score of 0.924 for overall skin detection, as shown in Fig. 5.

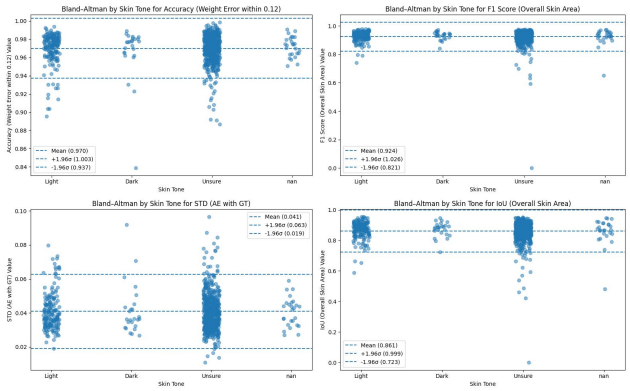

Figure 5. Evaluating skin segmentation by skin tone. Top left: accuracy (Weight Error within 0.12). Top right: F1 score (Overall Skin Area). Bottom left: standard deviation (AE with GT). Bottom right: IoU (Overall Skin Area)

4. Conclusions and Future Works

We present a full-body weighted skin segmentation model designed to utilize all available skin regions while intelligently and selectively adjusting weights for different areas suitable for unsupervised rPPG signal extraction pipelines. We collect a new video-PPG dataset, synchronized at the same sampling rate, containing four distinct real-world scenarios. Our model is trained on a synthesized dataset with image-mask samples. The results demonstrate the model's ability to accurately detect all available skin regions with strong generalization across a wide range of skin tones while distinguishing accessories and hair. Compared to previous methods, our approach achieves comparable results without requiring additional processing, outperforming existing methods in non-static scenarios. In the future, we will work to reduce the size of our full-body segmentation model to make it efficient for mobile devices. In addition, we will optimize the model for samples with lower resolutions to improve accessibility and reliability.

References

[1] Abdallah S. Abdallah, Mohamad Abou El-Nasr, and Amos Lynn Abbott. A new color image database for benchmarking of automatic face detection and human skin segmentation techniques. *World Academy of Science, Engineering and Technology, International Journal of Computer, Electrical, Automation, Control and Information Engineering*, 1:3769–3773, 2007. 3

[2] Bhargav Acharya, William Saakyan, Barbara Hammer, and Hanna Drimalla. Generalization of video-based heart rate

estimation methods to low illumination and elevated heart rates, 2025. 4

[3] Moussu A et al Allado E, Poussel M. Accurate and reliable assessment of heart rate in real-life clinical settings using an imaging photoplethysmography. *Journal of Clinical Medicine*, 2022. 4

[4] J. Allen and A. Murray. Effects of filtering on multisite photoplethysmography pulse waveform characteristics. In *Computers in Cardiology, 2004*, pages 485–488, 2004. 4

[5] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 92:35–42, 2017. 2, 4

[6] Mingyue Cao, Xu Cheng, Xingyu Liu, Yan Jiang, Hao Yu, and Jingang Shi. St-phys: Unsupervised spatio-temporal contrastive remote physiological measurement. *IEEE Journal of Biomedical and Health Informatics*, 28(8):4613–4624, 2024. 2

[7] Constantino Álvarez Casado and Miguel Bordallo López. Face2ppg: An unsupervised pipeline for blood volume pulse extraction from faces. *IEEE Journal of Biomedical and Health Informatics*, 27(11):5530–5541, 2023. 1, 2, 4, 5

[8] Liang-Chieh Chen, George Papandreou, Florian Schroff, and Hartwig Adam. Rethinking atrous convolution for semantic image segmentation. *CoRR*, abs/1706.05587, 2017. 3

[9] Weixuan 'Vincent' Chen and Daniel J. McDuff. Deepphys: Video-based physiological measurement using convolutional attention networks. *ArXiv*, abs/1805.07888, 2018. 1

[10] Chun-Hong Cheng, Kwan-Long Wong, Jing-Wei Chin, Tsz-Tai Chan, and Richard H. Y. So. Deep learning methods for remote heart rate measurement: A review and future research agenda. *Sensors*, 21(18), 2021. 1

[11] Gerard de Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013. 1

[12] John Gideon and Simon Stent. The way to my heart is through contrastive learning: Remote photoplethysmography from unlabelled video. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 3975–3984, 2021. 1

[13] Amogh Gudi, Marian Bittner, and Jan van Gemert. Real-time webcam heart-rate and variability estimation with clean ground truth for evaluation. *Applied Sciences*, 10(23), 2020. 1

[14] G. Haan, de and A.J. Leest, van. Improved motion robustness of remote-ppg by using the blood volume pulse signature. *Physiological Measurement*, 35(9):1913–1926, 2014. 1

[15] Lei Huang, Tian Xia, Yongdong Zhang, and Shouxun Lin. Human skin detection in images by mser analysis. *2011 18th IEEE International Conference on Image Processing*, pages 1257–1260, 2011. 3

[16] Yury Kartynnik, Artsiom Ablavatski, Ivan Grishchenko, and Matthias Grundmann. Real-time facial surface geometry from monocular video on mobile gpus. *ArXiv*, abs/1907.06724, 2019. 3

[17] Fatema-Tuz-Zohra Khanam, Ali Abdulelah Al-Naji, and Javaan Chahl. Remote monitoring of vital signs in diverse non- clinical and clinical scenarios using computer vision systems: A review. *Applied Sciences*, 9:4474, 2019. 1

[18] Dae-Yeol Kim, Kwangkee Lee, and Chae-Bong Sohn. Assessment of roi selection for facial video-based rppg. *Sensors*, 21(23), 2021. 1, 3

[19] 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*, 2020. 1

[20] Kunyoung Lee, Jaemu Oh, Hojoon You, and Eui Chul Lee. Improving remote photoplethysmography performance through deep-learning-based real-time skin segmentation network. *Electronics*, 12:3729, 2023. 2, 3, 4

[21] Magdalena Lewandowska, Jacek Rumiński, Tomasz Kocejko, and Jedrzej Nowak. Measuring pulse rate with a webcam — a non-contact method for evaluating cardiac activity. In *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 405–410, 2011. 1

[22] Magdalena Lewandowska, Jacek Rumiński, Tomasz Kocejko, and Jedrzej Nowak. Measuring pulse rate with a webcam — a non-contact method for evaluating cardiac activity. In *2011 Federated Conference on Computer Science and Information Systems (FedCSIS)*, pages 405–410, 2011. 1

[23] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Doll'a r, and C. Lawrence Zitnick. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312, 2014. 3, 4

[24] Tianqi Liu, Hanguang Xiao, Yisha Sun, Yulin Li, Shiyi Zhao, Zhenyu Yi, and Aohui Zhao. Style-rppg: Exploration and analysis of style transfer in unsupervised remote physiological measurement. *Expert Systems with Applications*, 269: 126310, 2025. 1

[25] Xin Liu, Josh Fromm, Shwetak Patel, and Daniel McDuff. Multi-task temporal shift attention networks for on-device contactless vitals measurement, 2020. 1

[26] Xin Liu, Xiaoyu Zhang, Girish Narayanswamy, Yuzhe Zhang, Yuntao Wang, Shwetak Patel, and Daniel McDuff. Deep physiological sensing toolbox. *arXiv preprint arXiv:2210.00716*, 2022. 1, 4

[27] Xin Liu, Brian Hill, Ziheng Jiang, Shwetak Patel, and Daniel McDuff. Efficientphys: Enabling simple, fast and accurate camera-based cardiac measurement. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pages 4997–5006, 2023. 1

[28] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. Mediapipe: A framework for building perception pipelines, 2019. 3

[29] Luis Francisco Corral Martínez, Gonzalo Páez, and Marija Strojnik. Optimal wavelength selection for noncontact reflection photoplethysmography. In *The International Commission for Optics*, 2011. 1

[30] Rita Meziati, Yannick Benezeth, Pierre De Oliveira, Julien Chappé, and Fan Yang. Ubfc-phys, 2021. 2, 4

[31] Loris Nanni, Andrea Loreggia, Alessandra Lumini, and Alberto Dorizza. A standardized approach for skin detection:

Analysis of the literature and case studies. *Journal of Imaging*, 9(2), 2023. 3

[32] Girish Narayanswamy, Yujia Liu, Yuzhe Yang, Chengqian Ma, Xin Liu, Daniel McDuff, and Shwetak N. Patel. Bigsmall: Efficient multi-task learning for disparate spatial and temporal physiological measurements. In *IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2024, Waikoloa, HI, USA, January 3-8, 2024*, pages 7899–7909. IEEE, 2024. 1

[33] Nhi Nguyen, Le Nguyen, Honghan Li, Miguel Bordallo López, and Constantino Álvarez Casado. Evaluation of video-based rppg in challenging environments: Artifact mitigation and network resilience. *Computers in Biology and Medicine*, 179:108873, 2024. 2

[34] Xuesong Niu, S. Shan, Hu Han, and Xilin Chen. Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation. *IEEE Transactions on Image Processing*, 29:2409–2423, 2019. 1

[35] Christian S. Pilz, Sebastian Zaunseder, Jarek Krajewski, and Vladimir Blazek. Local group invariance for heart rate estimation from face videos in the wild. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1335–13358, 2018. 1, 2

[36] Ming-Zher Poh, Daniel McDuff, and Rosalind Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE transactions on biomedical engineering*, 58:7–11, 2010. 1

[37] Matthieu Scherpf, Hannes Ernst, Leo Misera, Hagen Malberg, and Martin Schmidt. Skin segmentation for imaging photoplethysmography using a specialized deep learning approach. In *2021 Computing in Cardiology (CinC)*, pages 1–4, 2021. 1, 2

[38] Yi Sheng, Wu Zeng, Qiuyu Hu, Weihua Ou, Yuxuan Xie, and Jie Li. An improved approach to the performance of remote photoplethysmography. *Computers, Materials and Continua*, 73(2):2773–2783, 2022. 1

[39] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, and Maja Pantic. A multimodal database for affect recognition and implicit tagging. *IEEE Transactions on Affective Computing*, 3(1):42–55, 2012. 2

[40] Radim Spetlik, Jan Cech, Vojtěch Franc, and Jiri Matas. Visual heart rate estimation with convolutional neural network. 2018. 1

[41] R. Stricker, S. Müller, and H.-M. Gross. 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)*, pages 1056–1062, Edinburgh, Scotland, UK, 2014. IEEE. 2

[42] Zhaodong Sun and Xiaobai Li. Contrast-phys: Unsupervised video-based remote physiological measurement vianbsp;spatiotemporal contrast. page 492–510, Berlin, Heidelberg, 2022. Springer-Verlag. 1

[43] Debjyoti Talukdar, Luis Felipe de Deus, and Nikhil Sehgal. Evaluation of remote monitoring technology across different skin tone participants. *medRxiv*, 2023. 4

[44] Akito Tohma, Maho Nishikawa, Takuya Hashimoto, Yoichi Yamazaki, and Guanghao Sun. Evaluation of remote photoplethysmography measurement conditions toward telemedicine applications. *Sensors*, 21(24), 2021. 1

[45] Wim Verkruysse, Lars Svaasand, and John Nelson. Remote plethysmographic imaging using ambient light. *Optics Express*, 16:21434–21445, 2008. 1

[46] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Opt. Express*, 16(26):21434–21445, 2008. 1

[47] W. Wang, Sander Stuijk, and Gerard Haan. A novel algorithm for remote photoplethysmography: Spatial subspace rotation. *IEEE transactions on bio-medical engineering*, 0: 1, 2015. 1, 2

[48] Wenjin Wang, Albertus C. den Brinker, Sander Stuijk, and Gerard de Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2017. 1

[49] Wenjin Wang, Albertus C. den Brinker, and Gerard de Haan. Single-element remote-ppg. *IEEE Transactions on Biomedical Engineering*, 66(7):2032–2043, 2019. 1

[50] Kwan Long Wong, Jing Wei Chin, Tsz Tai Chan, Ismoil Odinaev, Kristian Suhartono, Kang Tianqu, and Richard Hau Yue So. Optimising rppg signal extraction by exploiting facial surface orientation. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2022, New Orleans, LA, USA, June 19-20, 2022*, pages 2164–2170. IEEE, 2022. 4

[51] Yuting Yang, Chenbin Liu, Hui Yu, Dangdang Shao, Francis Tsow, and Nongjian Tao. Motion robust remote photoplethysmography in cielab color space. *Journal of Biomedical Optics*, 21(11):117001, 2016. 1

[52] Hojoon You, Kunyoung Lee, Jaemu Oh, and Eui Chul Lee. Efficient and low color information dependency skin segmentation model. *Mathematics*, 11(9), 2023. 3

[53] Zitong Yu, Wei Peng, Xiaobai Li, Xiaopeng Hong, and Guoying Zhao. Remote heart rate measurement from highly compressed facial videos: An end-to-end deep learning solution with video enhancement. *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 151–160, 2019. 1

[54] Zitong Yu, Xiaobai Li, Xuesong Niu, Jingang Shi, and Guoying Zhao. Autohr: A strong end-to-end baseline for remote heart rate measurement with neural searching. *IEEE Signal Processing Letters*, PP:1–1, 2020.

[55] Zitong Yu, Yuming Shen, Jingang Shi, Hengshuang Zhao, Philip H. S. Torr, and Guoying Zhao. Physformer: Facial video-based physiological measurement with temporal difference transformer. *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4176–4186, 2021. 1

[56] Qi Zhan, Wenjin Wang, and Gerard Haan. Analysis of cnn-based remote-ppg to understand limitations and sensitivities. *Biomedical Optics Express*, 11, 2020. 1

[57] Dora Zhao, Angelina Wang, and Olga Russakovsky. Understanding and evaluating racial biases in image captioning. pages 14810–14820. Institute of Electrical and Electronics

Engineers Inc., 2021. Publisher Copyright: © 2021 IEEE.; 18th IEEE/CVF International Conference on Computer Vision, ICCV 2021 ; Conference date: 11-10-2021 Through 17-10-2021. 5

SkinMap: A Novel Weighted Full-Body Skin Segmentation for Robust Remote Photoplethysmography

Supplementary Material

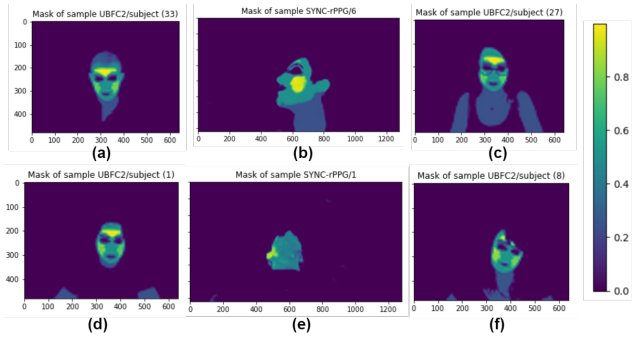

Figure 6. Extracted masks from samples of datasets.

Attribute	UBFC-rPPG	UBFC-PHYS	SYNC-rPPG
Sample count	50	168	80
Scenarios	rest	rest, talk, exercise	rest, talk, rotation, exercise
Video (FPS)	30	35	30
Sensor (Hz)	60	64	30
Resolution	640×480	1024×1024	1280×720
Age Range	18–25	not specified	18–25
Lighting	perfect	perfect	day-light + artificial
Sensor count	1	1	2

Table 1. Comparison of rPPG Datasets

4.1. Dataset

Our dataset (SYNC-rPPG) has been approved for public availability by each subject. Our institution does not require additional approval. For data acquisition, shown in Fig. 1 (a), we use a Raspberry Pi 4B with a 1.5 GHz processor, 8 GB of RAM, running Raspberry Pi OS and Python. For video capture, we employ the Raspberry Pi Camera V2. To capture heart pulse data, we integrate a MAX30102 sensor using I2C, with an ESP32 development board acting as a bridge to relay data to the Raspberry Pi and add a second MAX30102 sensor to capture pulse data from both hands, with each sensor connected to four ESP32 pins for simultaneous I2C connections. The sensors and camera are synchronized at 30 FPS to ensure accurate timing. A comparison of our dataset with existing ones that are used in this paper is reported in Tab. 1.

4.2. Mask illustration

Some of the more interesting masks generated by the model are illustrated in Fig. 6. (a) A bald subject where the entire head is correctly detected. (b) Even with only one side of the face visible, the mask is generated perfectly, with a higher weight on the cheeks than the forehead due to the angle. (c) The subject's hands and body skin are accurately extracted, while the glasses are excluded. (d) Both hands and the neck are included. (e) The subject is looking up, making the forehead and cheeks less visible, yet the model still performs well by detecting the neck. (f) A woman whose hair and necklace are successfully excluded.

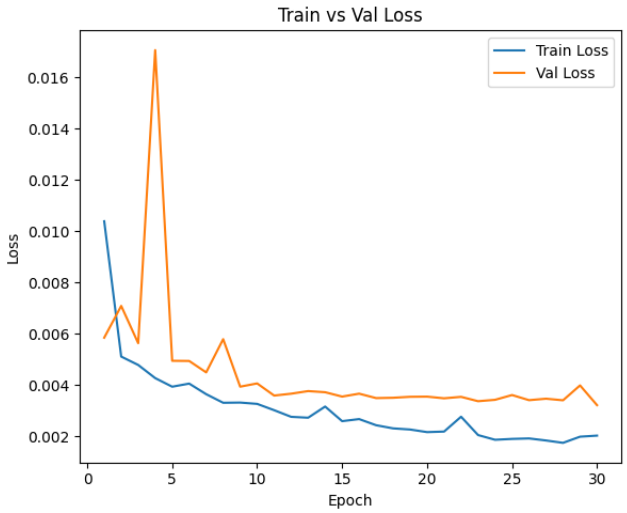

Figure 7. Train and validation loss of the model during training.

4.3. Training

The train and validation loss are visualized in Fig. 7. The training was conducted on an RTX 4090 GPU with 20GB of VRAM usage, supported by 198GB DDR5 RAM and an Intel i7-14700K CPU, running Python 3.10 with CUDA 12.4 on Linux Kernel 6.8. The training took approximately 4 hours. During training, the training loss steadily decreased and converged, and the validation loss, despite initial fluctuations, trended downward. We stopped at epoch 30, ensuring effective learning and generalization.