electronics

MDPI

Article

iBVP Dataset: RGB-Thermal rPPG Dataset with High Resolution Signal Quality Labels

Jitesh Joshi and Youngjun Cho *

Department of Computer Science, University College London, 169 Euston Road, London NW1 2AE, UK; jitesh.joshi.20@ucl.ac.uk
* Correspondence: youngjun.cho@ucl.ac.uk

Abstract: Remote photo-plethysmography (rPPG) has emerged as a non-intrusive and promising physiological sensing capability in human–computer interface (HCI) research, gradually extending its applications in health-monitoring and clinical care contexts. With advanced machine learning models, recent datasets collected in real-world conditions have gradually enhanced the performance of rPPG methods in recovering heart-rate and heart-rate-variability metrics. However, the signal quality of reference ground-truth PPG data in existing datasets is by and large neglected, while poor-quality references negatively influence models. Here, this work introduces a new imaging blood volume pulse (iBVP) dataset of synchronized RGB and thermal infrared videos with ground-truth PPG signals from ear with their high-resolution-signal-quality labels, for the first time. Participants perform rhythmic breathing, head-movement, and stress-inducing tasks, which help reflect real-world variations in psycho-physiological states. This work conducts dense (per sample) signal-quality assessment to discard noisy segments of ground-truth and corresponding video frames. We further present a novel end-to-end machine learning framework, iBVPNet, that features an efficient and effective spatio-temporal feature aggregation for the reliable estimation of BVP signals. Finally, this work examines the feasibility of extracting BVP signals from thermal video frames, which is under-explored. The iBVP dataset and source codes are publicly available for research use.

Keywords: remote PPG; RGB-thermal dataset; signal-quality labels

Citation: Joshi, J.; Cho, Y. iBVP Dataset: RGB-Thermal rPPG Dataset with High Resolution Signal Quality Labels. *Electronics* **2024**, *13*, 1334. https://doi.org/10.3390/electronics13071334

Academic Editor: Luca Mesin

Received: 7 February 2024
Revised: 9 March 2024
Accepted: 14 March 2024
Published: 2 April 2024

1. Introduction

The foundation of optical sensing for the blood volume pulse signal laid by [1] proved to be significantly useful in clinical care settings, and the past decade also witnessed the proliferation of health-tracking devices and smart watches that monitor heart-rate and heart-rate-variability metrics. Since Verkruysse [2]'s pioneering investigation on the feasibility of extracting photo-plethysmography signals from RGB cameras in a contactless manner, increasing attention has been given to a wide range of imaging-based physiological sensing methods and their promising applications and contexts where non-invasive and contactless measurement techniques are preferred, such as stress and mental workload recognition [3,4] and biometric authentication [5].

Several rPPG datasets have been made available for academic use, and this recent review, [6], overviews these datasets. Some of the widely used datasets include MANHOB-HCI [7], PURE [8], MMSE-HR [9], VIPL-HR [10], UBFC-rPPG [11], UBFC-Phys [12], V4V [13], and SCAMPS [14]. These datasets consist of the RGB videos with resolution ranging from 320 × 240 to 1920 × 1080, and frame rates from 20 frames per second (FPS) to 120 FPS. The ground-truth data often consist of the photo-plethysmography (PPG) signal and/or electro-cardiography (ECG) signal, along with their computed pulse rate (PR) or heart rate (HR) metrics. The majority of these datasets are acquired in laboratory settings with controlled lighting conditions [7,9,11,12], and varying head movement, poses, and emotional changes. PURE [8], ECG-Fitness [15], and VIPL-HR [10] datasets introduce

illumination changes for recording videos. VIPL-HR [10] and MANHOB-HCI [7] deploy multiple cameras to capture videos with different resolution as well as face poses. Unlike other datasets, the SCAMPS [14] dataset consists of synthetically generated videos with randomly sampled appearance attributes such as skin texture, hair, clothing, lighting, and environment, making it more suited to train supervised methods, rather than to evaluate rPPG methods. Most of the datasets focus on the RGB imaging modality.

As the ground-truth signals for rPPG datasets are collected using contact-based PPG or ECG sensors, it is essential to screen noise artifacts [16,17] present in such signals. Data collection scenarios including altered physiological states, varying ambient conditions, and head movement offer rich real-world representations, enabling robust training of supervised methods, as well as the realistic validation of both supervised and unsupervised rPPG extraction methods. However, these scenarios actively involving participants' movement result in noises in the reference PPG signals as well due to varying light issues on the contact points. Representative noise artifacts present in ground-truth PPG signals of the existing datasets can be observed in Figure 1.

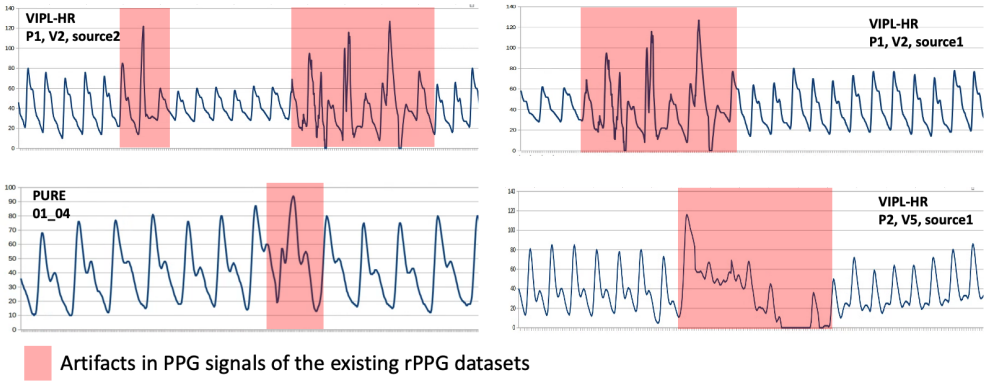

Figure 1. Noise artifacts present in the illustrative samples of PPG signals from existing rPPG datasets.

While researchers have proposed de-noising algorithms to remove artifacts from ECG as well as PPG signals [18,19], such methods are often limited to the signals that are not severely corrupted and are therefore reparable, while insufficient to denoise signals with substantial artifacts. Despite the availability of several rPPG datasets, the signal quality of the ground-truth signals is often neglected. Besides misleading the training of supervised methods, poor-quality ground-truth PPG signals can lead to the inappropriate evaluation of rPPG methods.

In this work, our collected RGB-thermal rPPG dataset (which we call iBVP dataset) is first assessed for signal quality of the ground-truth PPG signals. Signal-quality assessment is conducted both manually through the visual inspection of the PPG signals as well as using an automated approach. For the latter, we deploy SQA-PhysMD, adapted from a recent work [20], in which the CNN-based decoder module is replaced with a more efficient matrix-decomposition-based module [21] and the inference is made per sample, making it a dense 1D segmentation task. SQA-PhysMD shows significantly improved performance against the existing SOTA methods for an automated PPG-signal-quality assessment [22,23] on the iBVP dataset (preliminary results presented in Table A4 in Appendix C). The noisy segments can be flexibly removed from the ground-truth PPG signals as well as the corresponding video frames for training as well as evaluating rPPG methods. We make the original dataset available with high-resolution-signal-quality labels generated using SQA-PhysMD as well as the ones marked manually.

There have been several rPPG methods proposed over the decade, which can be categorized as unsupervised or model-based, feature-based supervised, and end-to-end supervised methods. Some of the notable foundational work with unsupervised methods includes Green [2], ICA [24], CHROM [25], PBV [26], POS [27], and LGI [28]. The research focus then progressed to *feature-based supervised* methods, which include HR-CNN [15],

DeepPhys [29], MTTS-CAN [30], RhythmNet [31], NAS-HR [32], PulseGAN [33], and Dual-GAN [34]. More recent rPPG development focuses on *end-to-end supervised* methods, which can be sub-categorized into the methods using convolution networks such as PhysNet [35], 3DCNN [36], SAM-rPPGNet [37], and RTrPPG [38], and transformer-network-based methods such as PhysFormer [39], PhysFormer++ [40], and EfficientPhys [41]. In addition to RGB frames, researchers have also shown the benefits of combining RGB with additional imaging modalities such as near-infrared (NIR) and long-wave infrared (thermal IR). One recent work in this direction proposed an Information-enhanced Network (IENet) [42], highlighting the superior performance of the model that combines RGB and NIR frames. To address the challenges in training end-to-end approaches as well as to optimize the usage of computing resources, a recent work proposed a multi-stage network, MSDN [43], which implements three stages that can be trained independently.

In this work, we first present an evaluation of the introduced iBVP dataset, and then show preliminary results of a novel 3D-CNN-based end-to-end learning approach, iBVPNet, for estimating PPG signals. iBVPNet effectively captures blood volume pulse (BVP)-related spatial and temporal features from video frames. To evaluate iBVPNet and existing state-of-the-art (SOTA) models, we leverage maximum amplitude of cross-correlation (MACC) [44] as an additional metric while reporting commonly used metrics. In spite of MACC being highly relevant in evaluating rPPG estimation, it has not been discussed sufficiently in the literature. In summary, we make the following contributions:

- Introducing the iBVP dataset comprising RGB and thermal facial video data with signal-quality-assessed ground-truth PPG signals.
- Presenting and validating a new rPPG framework, iBVPNet, for estimating the BVP signal from RGB as well as thermal video frames.
- Discovering MACC [44] as an effective evaluation metric to assess rPPG methods.

2. iBVP Dataset

2.1. Data Collection Protocol

The data acquisition was conducted with an objective of inducing variations in physiological states, as well as head movement. Each participant experienced four conditions, including (a) rhythmic slow breathing and rest, (b) an easy math task, (c) a difficult math task, and (d) a guided head movement task, as depicted in Figure 2A. While a higher agreement between the rPPG and the ground-truth PPG can be achieved in the absence of these variances, the inclusion of the same in the data acquisition protocol enables simulating real-world physiological variations.

Cognitively challenging math tasks with varying degrees of difficulty levels were chosen, as these have been reported to alter the physiological responses [45–47]. The achieved distribution of heart rate computed from ground-truth PPG signals can be observed in Figure 2B. Furthermore, as wearable sensors are less reliable under significant motion conditions [48], we added an experimental condition that involved guided head movement. Each condition lasted for 3 min, with 1 min of rest after each condition. To randomize the sequence of conditions, we interchanged "A" with "D" and "B" with "C". The study protocol was approved by the University College London Interaction Centre ethics committee (ID Number: UCLIC/1920/006/Staff/Cho).

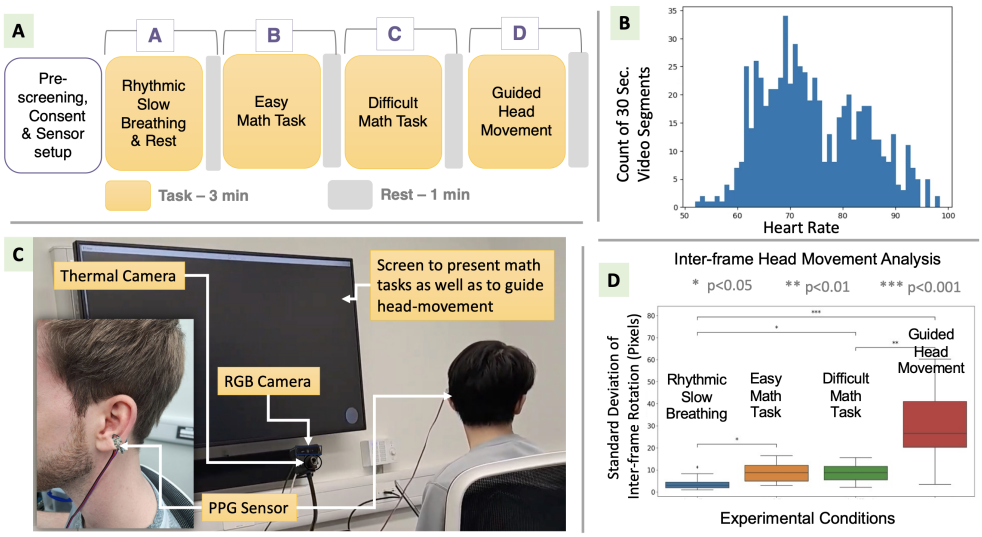

Figure 2. (**A**): Setup for acquiring iBVP dataset; (**B**): Analysis showing the magnitude of inter-frame head movement under different conditions involved in the data acquisition; (**C**): Data acquisition protocol; (**D**): Histogram showing the variations in heart rate.

2.2. Participants

PPG signals were collected from 33 participants (23 females) recruited through an online recruitment platform for research. Participants represented multiple ethnic groups and skin types with age ranging from 18 to 45 (27 ± 5.8). All participants reported having no known health conditions, provided informed consent ahead of the study, and were compensated for their time following the study. Detailed demographic information of study participants is provided in the Table A5 in Appendix D. After being welcomed and briefed, participants were asked to remove any bulky clothing (e.g., winter coats, jackets) and seated comfortably in front of a 65-by-37-inch screen, where they were fitted with PhysioKit [20] sensors. The PPG sensor was attached to participants' left or right ear with a metal clip. Of the 33 participants, one was excluded from the dataset due to insufficient consent for sharing the identifiable data with the research community. While the dataset comprises 32 participants, this work utilized data from the first 30 participants owing to logistic limitations.

2.3. Data Acquisition

As depicted in Figure 2C, RGB and thermal cameras were positioned in front of the participant at around 1 m distance. Logitech BRIO 4K UHD webcam (Logitech International, Lausanne, Switzerland) was used to capture RGB video frames with 640×480 resolution, while thermal infrared frames were captured using thermal camera (A65SC, FLIR systems Inc., Wilsonville, OR, USA) having 640×512 resolution. Frame rate for both RGB and thermal video acquisition was set to 30 FPS. Customized software was developed using C++ programming language, which utilized the FLIR-provided Spinnaker library (https://www.flir.co.uk/products/spinnaker-sdk/, accessed on 7 November 2022) to acquire thermal infrared frames, while using OpenCV library functions to acquire RGB frames. The PhysioKit toolkit [20] was adapted to acquire the ground-truth PPG signals in synchronization with RGB and thermal frames. Implementation was adapted such that RGB frames, thermal frames, and PPG signal were acquired in separate and dedicated threads, while sharing a common onset trigger and a timer to stop the acquisition in a synchronized manner. PPG signals were acquired with a higher sampling rate of 250.

2.4. Morphology and Time Delay of PPG Signals

The majority of existing rPPG datasets have ground-truth PPG signals acquired using a finger probe or wrist watch, making it challenging to match the morphology as

well as phase [49] of ground-truth and extracted rPPG signals from facial video frames. The morphology of PPG waveform is site-dependent [50] and therefore it is crucial to acquire ground-truth signals for rPPG from a site that is closest to the face. In addition, a recent study highlights the significant delay, equivalent to a half pulse duration between the PPG signal acquired from the finger and the rPPG signals [49]. With these considerations for morphological resemblance between PPG and rPPG signals and minimum time delay or phase difference, we carefully chose the ear as the sensor site for acquiring ground-truth PPG signals and attached the sensor clip to the upper or lower lobe of the ear based on the best fit and comfort of the participants. This makes the iBVP dataset highly suitable for training as well as evaluating the deep learning-based models that estimate BVP signals. It can be argued that the models that can reliably estimate the BVP signals offer significant advantages over the models trained to directly estimate heart rate or other BVP-derived metrics.

2.5. Pre-Processing and Signal Quality Assessment

A band-pass filter (0.5–2.5 Hz) of the second order was applied to PPG signals, which were then re-sampled to a sampling rate of 30, to match it with the FPS of RGB and thermal video frames. The band-pass filter was further applied after re-sampling the signal to reduce the sampling artifacts. The cut-off for higher frequency was chosen as 2.5 Hz to preserve only the pulsating waveform with systolic peaks, while discarding the features related to the dicrotic notch and diastolic peak as rPPG signals may not contain these characteristic features.

While PPG signals acquired from the ear tend to have good signal quality [20,51], it is still prone to noise artifacts due to head movement. Figure 2D shows a comparison of head movement across different experimental conditions, computed as the inter-frame rotation of facial frames. It is not only important to assess the quality of ground-truth PPG signals, but also to measure skin perfusion for the validity of the PPG signals. Skin-perfusion values were computed as a ratio of pulsatile amplitude (AC) to non-pulsatile (DC) amplitude from raw PPG signals [52]. For signal quality, the ground-truth PPG signals of our iBVP dataset were first assessed manually through visual inspection by one author and the same were meticulously reviewed by another author. Manually marked signal-quality labels were added to the dataset. To prevent subjective bias in marking the signal quality of the ground-truth PPG signals, we investigated automated signal-quality assessment methods.

The conventional signal-quality assessment methods for PPG signals rely on extracting (i) frequency components to compute the signal-to-noise ratio [53], (ii) different measures of signal-quality indices (SQIs) [54] including relative power SQI [47], or (iii) analyzing morphological features and comparing with the template signal [55]. In several real-world settings, frequency-based SQI measures of signal quality can be misleading due to overlapping frequency components of noise artifacts [56]. Morphological-feature-based signal-quality assessment is challenging owing to several factors [55,57] that include the following: (i) it is required to accurately segment the pulses to match the template, (ii) as morphological features vary significantly between different individuals, a generalized pulse template cannot be used as a reference to match, and (iii) some noise artifacts resemble the pulse morphology, making it more challenging to discriminate between a good-quality PPG signal and noise artifacts.

Machine learning- and deep learning-based methods for PPG-signal-quality assessment have recently attracted wider attention among researchers [22,23,57–59]. These developments have been captured in a recent survey [60] that reviews signal-quality assessment methods for contact-based as well as imaging-based PPG. While the majority of works focus on developing a classifier model [58,61–63] with binary inference for quality of the length of a PPG signal, a few recent works have proposed models that offer high-temporal-resolution-signal-quality assessment [20,22,23,64]. In this work, we first extend SQA-Phys, the signal-quality assessment method deployed in [20], such that the inference for signal quality is made per sample, resulting in a 1D dense segmentation task, and

secondly we replace the CNN-based decoder of the encoder–decoder architecture with a decoder based on matrix decomposition [21].

A recent work on 2D semantic segmentation [21] shows the efficacy of a matrix-decomposition-based decoder in capturing the global context. Inspired from the approach of low-rank discovery through matrix decomposition [21], this work adapts the hamburger module by implementing the Non-negative Matrix Factorization (NMF) for 1D features. Combining the 1D-CNN encoder module of SQA-Phys [20] and a matrix-decomposition-based decoder [21], we refer to this new architecture as *SQA-PhysMD*, as depicted in Figure 3. Datasets used for the training of the *SQA-PhysMD* model include the PPG DaLiA [65] training set, WESAD dataset [66], and TROIKA [17] dataset. The signal-quality labels for this training dataset are provided by the authors of [23] and are available as a part of their repository [67]. Training parameters and the model validation were conducted in line with the recent state-of-the-art works [22,23].

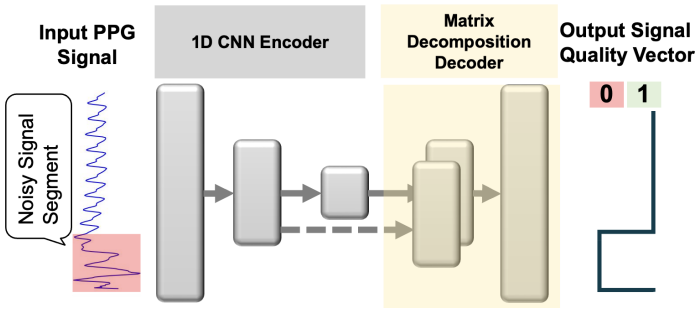

Figure 3. SQA-PhysMD: Signal-quality assessment module for PPG signals. Noisy PPG signal segments along with corresponding video frames are eliminated from the iBVP dataset.

SQA-PhysMD outperformed the SOTA models on the bench-marking PPG DaLiA testing set as well as showed very promising generalization on the PPG signals of the iBVP dataset, which can be observed in Appendix B. Signal-quality labels generated using the *SQA-PhysMD* model were therefore additionally added to the dataset to enable researchers to discard the noisy PPG signal segments along with their corresponding video frames. SQA-PhysMD can be further used with any existing rPPG dataset to clean the ground-truth signal and thereby eliminate the corresponding video frames.

As most of the rPPG methods deploy face detection as an initial step of the processing pipeline, we pre-process the dataset by first cropped facial regions. We use Python Facial Expression Analysis Toolbox (Py-Feat) [68] along with RetinaFace [69] to detect the facial frame in the RGB images. We then pick a cropping pixel dimension as 256 $\times$ 256 as it could contain the largest detected facial frame dimension with margins. This cropping reduces the image dimensions without incurring loss of information in temporal or spatial dimensions. Inspired by a recent work that explored different resolution of images for rPPG estimation using 3D CNN networks (RTrPPG) [38], we used 64 $\times$ 64 resolution for training and evaluation purposes. As thermal video frames were well acquired in alignment with the RGB video frames, the same cropping was used to pre-process the thermal video frames.

2.6. Comparison with Existing Datasets

Table 1 presents a comparison of different rPPG datasets (non-exhaustive), highlighting various aspects of each dataset. The advancements in rPPG research have been made owing to the availability of these datasets. A higher number of participants and varying scenarios including illumination conditions and tasks performed by the participants offer several advantages including the reliable validation of rPPG methods, as well as robust training of supervised rPPG algorithms. The key highlight of the iBVP dataset is its labels that are assessed for the signal quality, making it a highly reliable bench-marking dataset as well as a good candidate to train supervised models. Additionally, most of the existing datasets have captured ground-truth PPG signals from the finger or wrist, introducing not

just a phase difference but also morphological differences [50] with the rPPG signal that is extracted from facial regions. While the phase difference can be easily adjusted, when combined with morphological differences, it can not be optimally synchronized with the facial rPPG signals. The iBVP dataset therefore is more suitable for evaluating as well as training the models that estimate PPG signals in contrast to the models trained to estimate heart rate or related metrics with an end-to-end approach. For an exhaustive discourse and description of different rPPG datasets, it it recommended to refer to a recent review article [6].

Table 1. Comparison of different rPPG datasets.

Dataset	Modality	Subjects	Tasks	Duration (min)	Varying Illumination	SQ Labels	Resolution	Compression	FPS	Free Access
PURE [8]	RGB	10	S, M, T	60	Y	N	640 × 480	None	30	Yes
OBF * [70]	RGB, NIR	106	M	1000	N	N	640 × 480	None	30	No
MANHOB-HCI [7]	RGB	27	E	350	N	N	1040 × 1392	None	24	Yes
MMSE-HR [9]	RGB, 3D Thermal	40	E	935	N	N	RGB: 1040 × 1392; Thermal: 640 × 480	None	25	No
VIPL-HR [10]	RGB, NIR	107	S, M, T	1235	Y	N	Face-cropped	MJPG	25	Yes
UBFC-rPPG [11]	RGB	43	S, C	86	Y	N	640 × 480	None	30	Yes
UBFC-Phys [12]	RGB	56	S, C, T	504	N	N	1024 × 1024	JPEG	35	Yes
iBVP (Ours)	RGB, Thermal	32	B, C, M	381	N	Y	RGB: 640 × 480; Thermal: 640 × 512	None	30	Yes

B: Rhythmic breathing; E: Facial expression; M: Head movement; S: Stable; T: Talking; C: Controlled; SQ: Signal quality. * OBF dataset is temporarily unavailable at the time of submission.

3. Validation of iBVP Dataset

To evaluate the iBVP dataset, we chose the models that can be trained to infer BVP signals in an end-to-end manner. Among such models, 3D-CNN-architecture-based models including PhysNet3D [35] and RTrPPG [38] were found to be the most suitable to learn spatio-temporal features from facial video frames. The evaluation of the rPPG models trained with the iBVP dataset is performed with an objective to validate the iBVP dataset, supporting the use of the dataset as a bench-marking as well as training dataset.

We further introduce a novel 3D-CNN framework, iBVPNet, as illustrated in Figure 4. iBVPNet is a fully convolutional architecture designed with an objective of effectively learning spatio-temporal features for BVP estimation. It consists of three blocks, with each block distinctly learning spatial and temporal features. The first block aggregates spatial features, while encoding temporal features. The second block deploys large temporal kernels to encode the long-range temporal information. The final block further aggregates the spatial dimension while decoding the temporal features. Below, we describe the experiments and present the preliminary results, highlighting the efficacy of the proposed iBVPNet model.

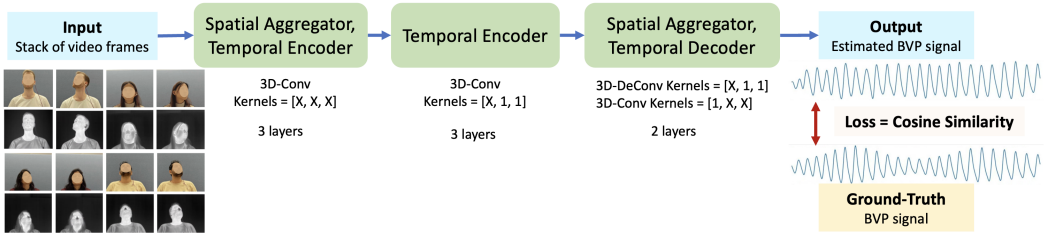

Figure 4. Overview of iBVPNet: The facial region is first cropped for every video frame, and then the frame is resized to a 64 × 64-pixel resolution. The architecture is fully convolutional and comprises three blocks, with the first and the last block aggregating the spatial features. Temporal encoding is achieved with the first 2 blocks, with the second block deploying higher-temporal kernels. The final block decodes the temporal signal while entirely reducing the spatial dimension.

3.1. Experiments

PhysNet3D [35], RTrPPG [38], and iBVPNet are trained using the iBVP dataset with a subject-wise 10-fold cross-validation approach. Models are separately trained and evaluated for RGB and thermal video frames. In each fold, data of 3 out of 30 participants are left out for validation, and the models are trained with the remaining data of 27 participants. A 20 s video segment and the corresponding ground-truth PPG signal are used for the training. A total of 600 face-cropped video frames are stacked and provided as input to the models, while the ground-truth PPG signals are resampled to 30 samples per second to match the count of video frames.

A batch size of 8 is used across all the experiments, and the learning rate is initialized to 1×10^{-4}, with a step-size of 2 iterations and gamma of 0.95. Models are trained for 100 iterations in each fold, with the cosine similarity (CS) loss function. Empirically, CS loss was found to achieve stable convergence in comparison with the negative Pearson correlation, which has been used in earlier works [35,38]. For augmenting the RGB video frames, video AugMix [71,72] was used to apply transforms that include changes related to contrast, equalization, rotation, shear, translation, and brightness. Thermal video frames are augmented using only rotation, shear, and translation transforms.

3.2. Evaluation Metrics

rPPG methods are commonly evaluated for HR measurement [6], whereas the methods aimed at estimating the BVP signals use the metrics that measure similarity between two time series signals. To evaluate the accuracy of HR measurement, widely used metrics include the Mean Absolute Error (MAE), Root Mean Square Error, and Pearson correlation coefficient [6]. Among the rPPG methods focused on BVP estimation, predominantly used metrics include the Template Match Correlation (TMC) [73] and signal-to-noise ratio (SNR) [35,38]. The performance of TMC can be affected by the accuracy of segmenting an individual pulse waveform from the PPG signals [54,73]. In this work, we propose using the metrics that align the two time-series signals without requiring to segment the waveform based on morphological features. Specifically, we compute cross-correlation between the ground-truth PPG signal and the estimated BVP signal at multiple time-lags, with an assumption that maximum amplitude of cross-correlation (MACC) [44] is achieved at the optimal alignment between the two signals. MACC computed for an individual pair of estimated and ground-truth PPG signals is further averaged across the testing dataset. We present the evaluation results using MACC and SNR metrics to assess the quality of estimated BVP signals. In addition, we also compute metrics based on HR measurement including root mean squared error (RMSE) and Pearson correlation between the HR values computed from ground-truth and estimated BVP signals.

3.3. Results

Evaluation metrics are first averaged for each fold out of 10-fold cross-validation and then further averaged across all the folds. Table 2 compares the averaged metrics for different end-to-end rPPG models trained with RGB video frames. For MACC, SNR, and RMSE (HR), the proposed iBVPNet shows superior performance, while the PhysNet3D shows the highest correlation for HR. Detailed fold-wise results for models trained with RGB video frames are presented in Table A1.

Table 2. Performance evaluation for rPPG estimation with RGB frames of iBVP dataset.

3D CNN Models	MACC (Avg)	SNR (Avg)	RMSE (HR)	Corr (HR)
PhysNet3D [35]	0.781	0.511	4.283	**0.848**
RTrPPG [38]	0.702	0.283	6.901	0.704
iBVPNet (ours)	**0.784**	**0.677**	**2.717**	0.813

The 33% increase in the SNR for the BVP signals estimated with trained iBVPNet models compared with the existing SOTA method is noteworthy. In Figure 5, we present estimated BVP waveforms for rhythmic breathing and head movement conditions to qualitatively compare the outcomes from our proposed models as well as the SOTA methods. It can be observed that head movement affects the quality of estimated rPPG signals both during rhythmic breathing as well as head movement conditions. Figure 5 further highlights that while the performance of RTrPPG [38] is not reliable, PhysNet3D [35] and the proposed iBVPNet show comparable estimated signals.

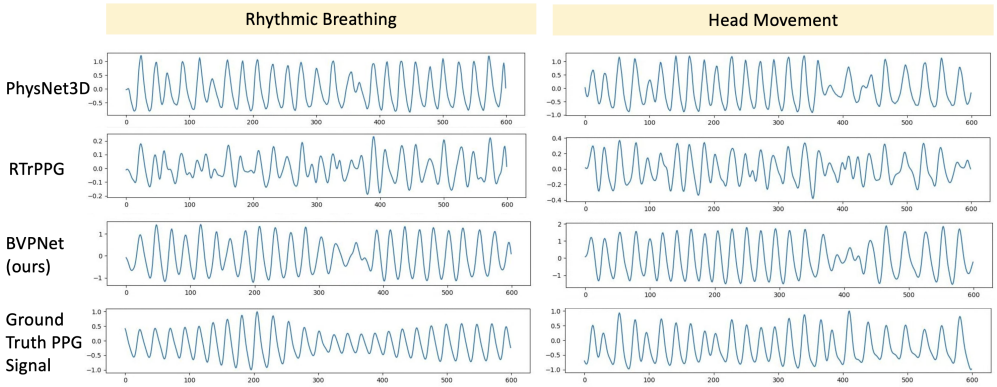

Figure 5. Qualitative comparison of the estimated BVP signals from RGB video frames.

In Table 3, we further compare the performance of conventional and some of the recent SOTA models on the iBVP dataset. Pre-trained weights of the supervised models trained on the PURE dataset [8], as provided by rPPG-Toolbox [74], were used to run the inferences. The iBVPNet model was trained on the PURE dataset to compare its performance with the SOTA models on the iBVP dataset. It is noteworthy to observe that the unsupervised method POS [27] achieves state-of-the-art performance on the iBVP dataset, specifically when evaluated with the MAE (HR) metric. The iBVPNet model proposed in this work demonstrates the highest performance on most of the metrics including RMSE (HR), Corr (HR), SNR, and MACC. The MACC metric [44] suggested in this work for evaluating rPPG methods shows good agreement with the existing metrics.

Table 3. Performance evaluation of SOTA models on iBVP dataset.

Unsupervised Models	MAE (HR)	RMSE (HR)	Corr (HR)	SNR (BVP)	MACC (BVP)
ICA [24]	10.25	14.11	0.10	−9.10	0.25
GREEN [2]	11.84	15.85	0.15	−10.18	0.22
CHROM [25]	4.18	8.33	0.56	−4.94	0.46
POS [27]	**3.51**	**6.98**	**0.71**	**−4.57**	**0.47**
LGI [28]	6.08	11.23	0.41	−6.38	0.39
PBV [26]	9.94	14.47	0.23	−8.74	0.27

Supervised Models (Trained on PURE [8])	MAE (HR)	RMSE (HR)	Corr (HR)	SNR (BVP)	MACC (BVP)
DeepPhys [29]	7.93	11.96	0.40	−8.04	0.34
TS-CAN [30]	6.68	10.27	0.36	−7.70	0.36
PhysNet3D [35]	4.85	8.69	0.56	−4.59	0.44
PhysFormer [39]	8.05	13.07	−0.02	−8.32	0.34
EfficientPhys [41]	5.06	9.23	0.65	−6.15	0.44
IBVPNet (ours)	**3.60**	**6.94**	**0.71**	**−3.35**	**0.50**

To compare the intra-dataset performance of five SOTA models on the existing datasets, we compiled a Table A3 in Appendix B, based on the data from a recent review article [6]. Table A3 also presents a comparison of two SOTA methods together with our proposed method on the introduced iBVP dataset, validating the usefulness of iBVP dataset in extracting BVP signals and HR.

Lastly, we performed the evaluation on infrared thermal image frames, which the iBVP dataset offers with higher temporal resolution (30 FPS) than the existing datasets. It can be noted from Table 1 that MMSE-HR is the only other dataset that offers thermal infrared

image frames for an rPPG estimation task; however, the frame rate for the same is 25 FPS. Table 4 compares the metrics averaged across multiple folds, for different end-to-end rPPG models trained with thermal video frames. Although the iBVPNet showed superior performance across all evaluation metrics as compared with the SOTAs, the overall quality of BVP estimation was not found satisfactory. Detailed fold-wise results for models trained with thermal video frames are presented in Table A2. Figure 6 qualitatively compares the outcomes of different rPPG methods in estimating BVP waveform from thermal video frames for rhythmic breathing and head movement conditions. This highlights that the BVP information extracted from the thermal frames was not strong. Similar results have been reported in [10] from NIR-based BVP estimation. However, a very recent work, [42], has shown highly promising results with the combination of RGB and NIR, indicating the further need for investigating the combination of RGB-thermal modalities for improving the performance on rPPG estimation.

Table 4. Performance evaluation for rPPG estimation using thermal frames of iBVP Dataset.

3D CNN Models	MACC (Avg)	SNR (Avg)	RMSE (HR)	Corr (HR)
PhysNet3D [35]	0.360	−0.135	6.339	−**0.019**
iBVPNet (ours)	**0.413**	**0.159**	**5.400**	0.095

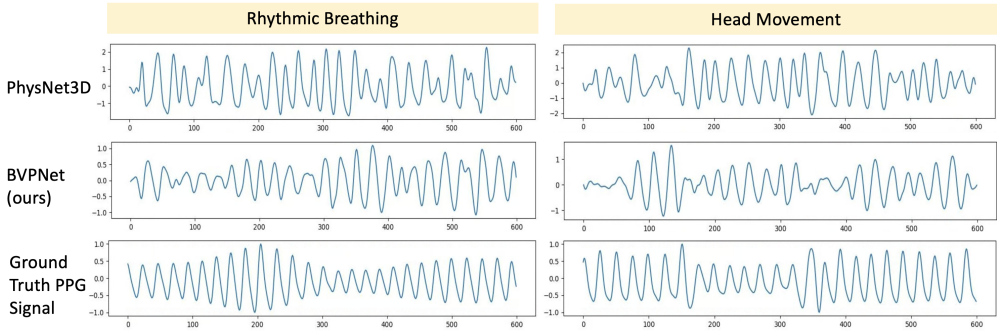

Figure 6. Qualitative comparison of the estimated BVP signals from thermal video frames.

4. Discussion and Conclusions

The experiments with the SOTA methods and the proposed iBVPNet model highlight the usefulness of the introduced iBVP dataset in training and validating the rPPG methods. While most of the rPPG methods estimate HR, it is advantageous to estimate the PPG signal, which can then reliably be used to extract various HR- and HRV-related metrics. The presence of noise artifacts in the ground-truth PPG signals can obscure the training stage for end-to-end supervised models. Here, the SQA-PhysMD implemented in this work for inferring a dense-signal-quality measure for PPG signals (an extended version of *SQA-Phys* [20]) has played a key role in eliminating noisy segments not only from the ground-truth PPG signals but also from the corresponding video frames.

As the SQA-PhysMD can assess signal quality for any type of PPG signals, it can be independently applied to the existing datasets for producing high-resolution-signal-quality labels as in the iBVP dataset. This can help with automatically removing noisy segments in the existing rPPG datasets, reducing tedious manual work and efforts, which are otherwise required to be made by researchers [38,75]. Furthermore, the ground-truth PPG signals acquired from the ear lobe closely match the phase and the morphology of the rPPG signals extracted from the facial video frames. Therefore, the iBVP dataset can significantly contribute toward improving robustness of rPPG methods. It is, however, to be noted that the dataset presents limited diversity with respect to skin color and varying lighting conditions. Further, the dataset does not present well-balanced gender distribution, which may slightly bias the training of the supervised rPPG methods. Future work to extend this dataset shall take these limitations into consideration.

Some of the existing RGB imaging-based rPPG datasets are available after applying the video compression techniques (e.g., motion JPEG, JPEG). It is noteworthy that the performance of SOTA models can be severely affected owing to the loss of BVP information from the compressed videos [35,37]. To circumvent this, one recent work implemented a generative method to reconstruct the original video frames from the compressed video frames as an initial step, followed by an architecture to estimate BVP signals [76]. However, this approach adds significant overhead in processing the video frames, and therefore alternative ways are required to address the BVP extraction from the compressed videos. Thus, the iBVP dataset offers raw RGB-thermal image frames, without the compression methods.

In agreement with a previous finding on rPPG with infrared (IR) video frames [10], this work reports poor performance of the existing SOTA rPPG methods as well as our proposed iBVPNet model on thermal video frames. It is worth noting that thermal video frames require tailored pre-processing since various factors including ambient temperature and quantization methods [44] can significantly impact the rPPG extraction. Alternatively, raw thermal frames augmented with a thermal augmentation module such as that proposed in a recent work on the segmentation of thermal facial frames [77] can be incorporated while training the models for rPPG estimation. In addition, approaches combining the RGB and thermal modalities can be investigated to compare the model performance with that trained individually on RGB and thermal modalities. Further investigations on assessing the potential of thermal infrared imaging in extracting BVP signals are therefore required, to which our dataset can contribute in future work.

Lastly, we highlight the use of MACC as an effective metric, which can be used along with the widely used metrics to report rPPG methods. The unique benefit MACC offers is that it is a more direct metric to compare the waveforms, and does not become impacted by the method used to compute the heart rate (e.g., based on FFT or counting of peaks in the waveform).

Author Contributions: Conceptualization, Y.C. and J.J.; methodology, J.J. and Y.C.; programming, J.J.; study validation, J.J.; artifact validation, J.J. and Y.C.; investigation, Y.C.; data collection and analysis, J.J. and Y.C.; manuscript preparation, J.J. and Y.C.; visualization, J.J.; overall supervision, Y.C.; project administration, Y.C.; funding acquisition, Y.C. All authors have read and agreed to the published version of the manuscript.

Funding: UCL CS PhD Studentship; GDI—Physiological Computing and Artificial Intelligence.

Institutional Review Board Statement: The study protocol is approved by the University College London Interaction Centre ethics committee (ID Number: UCLIC/1920/006/Staff/Cho, Approval date: 20 May 2020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: We have released the iBVP dataset and the source codes can be obtained from our GitHub page.

Acknowledgments: Authors thank Katherine Wang for assisting in recruitment of the participants and supporting data collection. The authors also thank our participants who participated in the study.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

1D-CNN	1-dimensional convolutional neural network
BPM	Beats per minute
BVP	Blood volume pulse
ECG	Electrocardiogram
HR	Heart rate
PPG	Photoplethysmography
RGB	Color images with red, green, and blue frames

Appendix A. Detailed Results of Multifold Evaluation

Below, we present the fold-wise comparison between different rPPG methods, separately for the models trained on RGB and thermal video frames.

Appendix A.1. RGB Video Frames

Table A1. Detailed performance evaluation for rPPG estimation with RGB video frames of iBVP dataset.

Folds	MACC (Avg)			SNR (Avg)			RMSE (HR)			Corr (HR)		
	PhysNet3D	RTrPPG	iBVPNet (Ours)	PhysNet3D	RTrPPG	iBVPNet (Ours)	PhysNet3D	RTrPPG	iBVPNet (Ours)	PhysNet3D	RTrPPG	iBVPNet (Ours)
0	0.767	0.669	**0.790**	0.532	0.250	**0.762**	2.829	6.058	**1.476**	0.846	0.568	**0.860**
1	**0.734**	0.654	0.710	0.373	0.190	**0.423**	8.412	12.480	**5.325**	**0.538**	0.258	0.376
2	0.830	0.773	**0.860**	0.709	0.475	**0.972**	2.937	6.213	**1.412**	0.888	0.587	**0.934**
3	**0.718**	0.637	0.660	**0.305**	0.113	0.291	5.848	7.591	**4.542**	**0.800**	0.674	0.679
4	**0.851**	0.763	0.836	0.637	0.402	**0.740**	2.330	3.993	**1.681**	**0.955**	0.879	0.945
5	**0.867**	0.801	0.853	0.601	0.373	**0.808**	2.092	3.508	**1.113**	0.966	0.905	**0.973**
6	0.780	0.689	**0.824**	0.573	0.297	**0.825**	5.114	7.682	**2.342**	0.898	0.826	**0.945**
7	**0.821**	0.751	**0.821**	0.603	0.342	**0.806**	2.943	5.051	**2.652**	**0.903**	0.781	0.830
8	0.702	0.603	**0.744**	0.329	0.113	**0.604**	4.103	11.395	**2.692**	**0.772**	0.655	0.724
9	0.743	0.680	**0.746**	0.445	0.271	**0.535**	6.222	5.044	**3.932**	0.909	**0.911**	0.870

Appendix A.2. Thermal Video Frames

Table A2. Detailed performance evaluation for rPPG estimation with thermal video frames of iBVP dataset.

Folds	MACC (Avg)		SNR (Avg)		RMSE (HR)		Corr (HR)	
	PhysNet3D	iBVPNet (Ours)	PhysNet3D	iBVPNet (Ours)	PhysNet3D	iBVPNet (Ours)	PhysNet3D	iBVPNet (Ours)
0	0.377	**0.469**	−0.099	**0.363**	6.496	**3.144**	0.092	**0.136**
1	0.352	**0.403**	−0.110	**0.109**	6.932	**5.557**	**0.286**	0.065
2	0.389	**0.437**	−0.071	**0.266**	5.599	**4.731**	**−0.139**	−0.218
3	0.378	**0.409**	−0.151	**0.171**	5.856	**5.037**	0.093	**0.589**
4	0.367	**0.401**	−0.120	**0.138**	5.475	**5.401**	**0.065**	−0.060
5	0.368	**0.442**	−0.149	**0.232**	5.628	**4.856**	−0.141	**−0.046**
6	0.350	**0.430**	−0.114	**0.213**	6.815	**5.865**	0.014	**0.365**
7	0.338	**0.386**	−0.150	**0.113**	**5.453**	6.015	**−0.238**	−0.247
8	0.358	**0.431**	−0.144	**0.264**	6.409	**4.245**	−0.063	**0.238**
9	**0.326**	0.322	**−0.238**	−0.279	**8.732**	9.152	−0.162	**0.129**

Appendix B. Performance Comparison of SOTA rPPG Methods on Existing Bench-Marking Datasets and iBVP Dataset

Table A3. Performance comparison of SOTA rPPG methods on existing bench-marking datasets and iBVP Dataset.

Datasets	rPPG Method	RMSE (HR)	Corr (HR)
PURE [8]	PhysNet3D [35]	2.60	0.99
	rPPGNet [76]	1.21	1.00
	SAM-rPPGNet [37]	1.21	1.00
MANHOB-HCI [7]	PhysNet3D [35]	8.76	0.69
	rPPGNet [76]	5.93	0.88
VIPL-HR [10]	PhysNet3D [35]	14.80	0.20
	AutoHR [78]	8.68	0.72
iBVP Dataset (ours)	PhysNet3D [35]	4.28	0.85
	RTrPPG [38]	6.90	0.70
	iBVPNet (ours)	2.72	0.81

Note: Only 3D CNN methods that estimate BVP signals are chosen. For iBVP dataset, only RGB imaging modality is considered for this comparison, as it is the only common modality among the datasets).

Appendix C. Evaluation of PPG-Signal-Quality Assessment Methods

In the below table, we compare the performance of the novel PPG-signal-quality assessment method, SQA-PhysMD, with the existing state-of-the-art methods that include Segade [23] and TinyPPG [22].

Table A4. Performance evaluation of PPG-signal-quality assessment methods on bench-marking and iBVP dataset.

SQ Method	PPG DaLiA Testing Set		iBVP Dataset	
	DICE Score (%)	Accuracy (%)	DICE Score (%)	Accuracy (%)
Segade [23]	86.47	85.14	48.61	96.93
TinyPPG [22]	87.03	85.99	52.56	97.59
SQA-PhysMD (ours)	**87.61**	**86.62**	**67.47**	**98.72**

Appendix D. Demographic Information of the Study Participants

In the below table, we provide demographic details of the study participants. Please note following: (i) participant 'p09' is excluded from the released dataset, and (ii) the data of 'p16' is not to be used for any presentation material such as figures in the research papers or published videos.

Table A5. Participant Demographics.

PID	Gender	Age	Ethnicity †
p01	M	33	A
p02	F	19	A
p03	M	21	A
p04	F	20	A
p05	F	32	D
p06	F	19	C
p07	M	18	C
p08	F	30	C
p09	F	25	C
p10	F	23	A
p11	M	32	A
p12	F	30	A
p13	F	25	C
p14	F	23	A
p15	F	24	A
p16	F	20	A
p17	M	28	E
p18	F	24	C
p19	F	21	A
p20	M	27	C
p21	F	27	C
p22	M	28	A
p23	F	34	A
p24	F	25	C
p25	F	27	C
p26	F	22	B
p27	M	45	A
p28	F	31	C
p29	F	28	A
p30	F	35	B
p31	M	33	A
p32	F	24	A
p33	M	29	D

† Ethnic groups are coded as follows: **A:** Asian/Asian British (Indian, Pakistani, Bangladeshi, Chinese, any other Asian background); **B:** Black/Black British (African, Caribbean, any other Black/African/Caribbean background); **C:** Caucasian; **D:** Mixed / Multiple ethnic groups; **E:** Other.

References

1. Hertzman, A.B. The Blood Supply of Various Skin Areas as Estimated by the Photoelectric Plethysmograph. *Am. J. Physiol.-Leg. Content* **1938**, *124*, 328–340. [CrossRef]
2. Verkruysse, W.; Svaasand, L.O.; Nelson, J.S. Remote Plethysmographic Imaging Using Ambient Light. *Opt. Express* **2008**, *16*, 21434–21445. [CrossRef] [PubMed]
3. Cho, Y.; Bianchi-Berthouze, N.; Julier, S.J. DeepBreath: Deep Learning of Breathing Patterns for Automatic Stress Recognition Using Low-Cost Thermal Imaging in Unconstrained Settings. In Proceedings of the 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), San Antonio, TX, USA, 23–26 October 2017; pp. 456–463. [CrossRef]
4. Cho, Y. Rethinking Eye-blink: Assessing Task Difficulty through Physiological Representation of Spontaneous Blinking. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems, CHI '21, New York, NY, USA, 8–13 May 2021; pp. 1–12. [CrossRef]

5. Reşit Kavsaoğlu, A.; Polat, K.; Recep Bozkurt, M. A Novel Feature Ranking Algorithm for Biometric Recognition with PPG Signals. *Comput. Biol. Med.* **2014**, *49*, 1–14. [CrossRef] [PubMed]
6. Xiao, H.; Liu, T.; Sun, Y.; Li, Y.; Zhao, S.; Avolio, A. Remote Photoplethysmography for Heart Rate Measurement: A Review. *Biomed. Signal Process. Control* **2024**, *88*, 105608. [CrossRef]
7. Soleymani, M.; Lichtenauer, J.; Pun, T.; Pantic, M. A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Trans. Affect. Comput.* **2012**, *3*, 42–55. [CrossRef]
8. Stricker, R.; Müller, S.; Gross, H.M. 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, Edinburgh, UK, 25–29 August 2014; pp. 1056–1062. [CrossRef]
9. Zhang, Z.; Girard, J.M.; Wu, Y.; Zhang, X.; Liu, P.; Ciftci, U.; Canavan, S.; Reale, M.; Horowitz, A.; Yang, H.; et al. Multimodal Spontaneous Emotion Corpus for Human Behavior Analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3438–3446.
10. Niu, X.; Han, H.; Shan, S.; Chen, X. VIPL-HR: A Multi-modal Database for Pulse Estimation from Less-Constrained Face Video. In Proceedings of the Computer Vision—ACCV 2018, Perth, Australia, 2–6 December 2018; Jawahar, C., Li, H., Mori, G., Schindler, K., Eds.; Lecture Notes in Computer Science; Springer: Cham, Switzerland, 2019; pp. 562–576. [CrossRef]
11. Bobbia, S.; Macwan, R.; Benezeth, Y.; Mansouri, A.; Dubois, J. Unsupervised Skin Tissue Segmentation for Remote Photoplethysmography. *Pattern Recognit. Lett.* **2019**, *124*, 82–90. [CrossRef]
12. Sabour, R.M.; Benezeth, Y.; De Oliveira, P.; Chappé, J.; Yang, F. UBFC-Phys: A Multimodal Database For Psychophysiological Studies of Social Stress. *IEEE Trans. Affect. Comput.* **2023**, *14*, 622–636. [CrossRef]
13. Revanur, A.; Li, Z.; Ciftci, U.A.; Yin, L.; Jeni, L.A. The First Vision for Vitals (V4V) Challenge for Non-Contact Video-Based Physiological Estimation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 2760–2767.
14. McDuff, D.; Wander, M.; Liu, X.; Hill, B.L.; Hernandez, J.; Lester, J.; Baltrusaitis, T. SCAMPS: Synthetics for Camera Measurement of Physiological Signals. *arXiv* **2022**, arXiv:2206.04197. [CrossRef]
15. Špetlík, R. Visual Heart Rate Estimation with Convolutional Neural Network. In Proceedings of the British Machine Vision Conference, Newcastle, UK, 9–13 September 2018.
16. Castaneda, D.; Esparza, A.; Ghamari, M.; Soltanpur, C.; Nazeran, H. A Review on Wearable Photoplethysmography Sensors and Their Potential Future Applications in Health Care. *Int. J. Biosens. Bioelectron.* **2018**, *4*, 195–202. [CrossRef] [PubMed]
17. Zhang, Z.; Pi, Z.; Liu, B. TROIKA: A General Framework for Heart Rate Monitoring Using Wrist-Type Photoplethysmographic Signals during Intensive Physical Exercise. *IEEE Trans. Biomed. Eng.* **2015**, *62*, 522–531. [CrossRef] [PubMed]
18. Chuang, C.H.; Chang, K.Y.; Huang, C.S.; Jung, T.P. IC-U-Net: A U-Net-based Denoising Autoencoder Using Mixtures of Independent Components for Automatic EEG Artifact Removal. *NeuroImage* **2022**, *263*, 119586. [CrossRef]
19. Jain, P.; Ding, C.; Rudin, C.; Hu, X. A Self-Supervised Algorithm for Denoising Photoplethysmography Signals for Heart Rate Estimation from Wearables. *arXiv* **2023**, arXiv:2307.05339. [CrossRef]
20. Joshi, J.; Wang, K.; Cho, Y. PhysioKit: An Open-Source, Low-Cost Physiological Computing Toolkit for Single- and Multi-User Studies. *Sensors* **2023**, *23*, 8244. [CrossRef] [PubMed]
21. Geng, Z.; Guo, M.H.; Chen, H.; Li, X.; Wei, K.; Lin, Z. Is Attention Better Than Matrix Decomposition? *arXiv* **2021**, arXiv:2109.04553. [CrossRef]
22. Zheng, Y.; Wu, C.; Cai, P.; Zhong, Z.; Huang, H.; Jiang, Y. Tiny-PPG: A Lightweight Deep Neural Network for Real-Time Detection of Motion Artifacts in Photoplethysmogram Signals on Edge Devices. *Internet Things* **2024**, *25*, 101007. [CrossRef]
23. Guo, Z.; Ding, C.; Hu, X.; Rudin, C. A Supervised Machine Learning Semantic Segmentation Approach for Detecting Artifacts in Plethysmography Signals from Wearables. *Physiol. Meas.* **2021**, *42*, 125003. [CrossRef] [PubMed]
24. Poh, M.Z.; McDuff, D.J.; Picard, R.W. Non-Contact, Automated Cardiac Pulse Measurements Using Video Imaging and Blind Source Separation. *Opt. Express* **2010**, *18*, 10762–10774. [CrossRef] [PubMed]
25. de Haan, G.; Jeanne, V. Robust Pulse Rate From Chrominance-Based rPPG. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 2878–2886. [CrossRef] [PubMed]
26. de Haan, G.; van Leest, A. Improved Motion Robustness of Remote-PPG by Using the Blood Volume Pulse Signature. *Physiol. Meas.* **2014**, *35*, 1913. [CrossRef] [PubMed]
27. Wang, W.; den Brinker, A.C.; Stuijk, S.; de Haan, G. Algorithmic Principles of Remote PPG. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 1479–1491. [CrossRef] [PubMed]
28. Pilz, C.S.; Zaunseder, S.; Krajewski, J.; Blazek, V. Local Group Invariance for Heart Rate Estimation From Face Videos in the Wild. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Salt Lake City, UT, USA, 18–22 June 2018; pp. 1254–1262.
29. Chen, W.; McDuff, D. DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks. In Proceedings of the Computer Vision—ECCV 2018, Munich, Germany, 8–14 September 2018; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer: Cham, Switzerland, 2018; Volume 11206, pp. 356–373. [CrossRef]
30. Liu, X.; Fromm, J.; Patel, S.; McDuff, D. Multi-Task Temporal Shift Attention Networks for On-Device Contactless Vitals Measurement. In Proceedings of the Advances in Neural Information Processing Systems, Online, 6 June 2020; Curran Associates, Inc.: Nice, France, 2020; Volume 33, pp. 19400–19411.

31. Niu, X.; Shan, S.; Han, H.; Chen, X. RhythmNet: End-to-End Heart Rate Estimation from Face via Spatial-Temporal Representation. *IEEE Trans. Image Process.* **2020**, *29*, 2409–2423. [CrossRef] [PubMed]
32. Lu, H.; Han, H. NAS-HR: Neural Architecture Search for Heart Rate Estimation from Face Videos. *Virtual Real. Intell. Hardw.* **2021**, *3*, 33–42. [CrossRef]
33. Song, R.; Chen, H.; Cheng, J.; Li, C.; Liu, Y.; Chen, X. PulseGAN: Learning to Generate Realistic Pulse Waveforms in Remote Photoplethysmography. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1373–1384. [CrossRef] [PubMed]
34. Lu, H.; Han, H.; Zhou, S.K. Dual-GAN: Joint BVP and Noise Modeling for Remote Physiological Measurement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 12404–12413.
35. Yu, Z.; Li, X.; Zhao, G. Remote Photoplethysmograph Signal Measurement from Facial Videos Using Spatio-Temporal Networks. *arXiv* **2019**, arXiv:1905.02419. [CrossRef]
36. Bousefsaf, F.; Pruski, A.; Maaoui, C. 3D Convolutional Neural Networks for Remote Pulse Rate Measurement and Mapping from Facial Video. *Appl. Sci.* **2019**, *9*, 4364. [CrossRef]
37. Hu, M.; Qian, F.; Wang, X.; He, L.; Guo, D.; Ren, F. Robust Heart Rate Estimation With Spatial–Temporal Attention Network From Facial Videos. *IEEE Trans. Cogn. Dev. Syst.* **2022**, *14*, 639–647. [CrossRef]
38. Botina-Monsalve, D.; Benezeth, Y.; Miteran, J. RTrPPG: An Ultra Light 3DCNN for Real-Time Remote Photoplethysmography. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 19–20 June 2022; pp. 2146–2154.
39. Yu, Z.; Shen, Y.; Shi, J.; Zhao, H.; Torr, P.H.S.; Zhao, G. PhysFormer: Facial Video-Based Physiological Measurement with Temporal Difference Transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 4186–4196.
40. Yu, Z.; Shen, Y.; Shi, J.; Zhao, H.; Cui, Y.; Zhang, J.; Torr, P.; Zhao, G. PhysFormer++: Facial Video-Based Physiological Measurement with SlowFast Temporal Difference Transformer. *Int. J. Comput. Vis.* **2023**, *131*, 1307–1330. [CrossRef]
41. Liu, X.; Hill, B.; Jiang, Z.; Patel, S.; McDuff, D. EfficientPhys: Enabling Simple, Fast and Accurate Camera-Based Cardiac Measurement. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 2–7 January 2023; pp. 5008–5017.
42. Liu, L.; Xia, Z.; Zhang, X.; Peng, J.; Feng, X.; Zhao, G. Information-Enhanced Network for Noncontact Heart Rate Estimation from Facial Videos. In *IEEE Transactions on Circuits and Systems for Video Technology*; IEEE: Toulouse, France, 2023; p. 1. [CrossRef]
43. Zhang, X.; Xia, Z.; Dai, J.; Liu, L.; Peng, J.; Feng, X. MSDN: A Multistage Deep Network for Heart-Rate Estimation from Facial Videos. *IEEE Trans. Instrum. Meas.* **2023**, *72*, 1–15. [CrossRef]
44. Cho, Y.; Julier, S.J.; Marquardt, N.; Bianchi-Berthouze, N. Robust Tracking of Respiratory Rate in High-Dynamic Range Scenes Using Mobile Thermal Imaging. *Biomed. Opt. Express* **2017**, *8*, 4480–4503. [CrossRef] [PubMed]
45. Tonacci, A.; Billeci, L.; Burrai, E.; Sansone, F.; Conte, R. Comparative Evaluation of the Autonomic Response to Cognitive and Sensory Stimulations through Wearable Sensors. *Sensors* **2019**, *19*, 4661. [CrossRef] [PubMed]
46. Birkett, M.A. The Trier Social Stress Test Protocol for Inducing Psychological Stress. *J. Vis. Exp. JoVE* **2011**, *56*, 3238. [CrossRef]
47. Cho, Y.; Julier, S.J.; Bianchi-Berthouze, N. Instant Stress: Detection of Perceived Mental Stress through Smartphone Photoplethysmography and Thermal Imaging. *JMIR Ment. Health* **2019**, *6*, e10140. [CrossRef] [PubMed]
48. Johnson, K.T.; Narain, J.; Ferguson, C.; Picard, R.; Maes, P. The ECHOS Platform to Enhance Communication for Nonverbal Children with Autism: A Case Study. In Proceedings of the Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, CHI EA'20, New York, NY, USA, 25–30 April 2020; pp. 1–8. [CrossRef]
49. Casado, C.Á.; López, M.B. Face2PPG: An Unsupervised Pipeline for Blood Volume Pulse Extraction from Faces. *IEEE J. Biomed. Health Inform.* **2023**, *27*, 5530–5541. [CrossRef] [PubMed]
50. Allen, J.; Murray, A. Effects of Filtering on Multisite Photoplethysmography Pulse Waveform Characteristics. In Proceedings of the Computers in Cardiology, Chicago, IL, USA, 19–22 September 2004; pp. 485–488. [CrossRef]
51. Patterson, J.A.; McIlwraith, D.C.; Yang, G.Z. A Flexible, Low Noise Reflective PPG Sensor Platform for Ear-Worn Heart Rate Monitoring. In Proceedings of the 2009 Sixth International Workshop on Wearable and Implantable Body Sensor Networks, Berkeley, CA, USA, 3–5 June 2009; pp. 286–291. [CrossRef]
52. Lindberg, L.G.; Öberg, P.Å. Photoplethysmography. *Med. Biol. Eng. Comput.* **1991**, *29*, 48–54. [CrossRef] [PubMed]
53. Huang, F.H.; Yuan, P.J.; Lin, K.P.; Chang, H.H.; Tsai, C.L. Analysis of Reflectance Photoplethysmograph Sensors. *Int. J. Biomed. Biol. Eng.* **2011**, *5*, 622–625.
54. Elgendi, M. Optimal Signal Quality Index for Photoplethysmogram Signals. *Bioengineering* **2016**, *3*, 21. [CrossRef] [PubMed]
55. Sukor, J.A.; Redmond, S.J.; Lovell, N.H. Signal Quality Measures for Pulse Oximetry through Waveform Morphology Analysis. *Physiol. Meas.* **2011**, *32*, 369. [CrossRef] [PubMed]
56. Song, J.; Li, D.; Ma, X.; Teng, G.; Wei, J. PQR Signal Quality Indexes: A Method for Real-Time Photoplethysmogram Signal Quality Estimation Based on Noise Interferences. *Biomed. Signal Process. Control* **2019**, *47*, 88–95. [CrossRef]
57. Goh, C.H.; Tan, L.K.; Lovell, N.H.; Ng, S.C.; Tan, M.P.; Lim, E. Robust PPG Motion Artifact Detection Using a 1-D Convolution Neural Network. *Comput. Methods Programs Biomed.* **2020**, *196*, 105596. [CrossRef] [PubMed]
58. Gao, H.; Wu, X.; Shi, C.; Gao, Q.; Geng, J. A LSTM-Based Realtime Signal Quality Assessment for Photoplethysmogram and Remote Photoplethysmogram. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 19–25 June 2021; pp. 3831–3840.

59. Roh, D.; Shin, H. Recurrence Plot and Machine Learning for Signal Quality Assessment of Photoplethysmogram in Mobile Environment. *Sensors* **2021**, *21*, 2188. [CrossRef] [PubMed]
60. Desquins, T.; Bousefsaf, F.; Pruski, A.; Maaoui, C. A Survey of Photoplethysmography and Imaging Photoplethysmography Quality Assessment Methods. *Appl. Sci.* **2022**, *12*, 9582. [CrossRef]
61. Moscato, S.; Lo Giudice, S.; Massaro, G.; Chiari, L. Wrist Photoplethysmography Signal Quality Assessment for Reliable Heart Rate Estimate and Morphological Analysis. *Sensors* **2022**, *22*, 5831. [CrossRef] [PubMed]
62. Shin, H. Deep Convolutional Neural Network-Based Signal Quality Assessment for Photoplethysmogram. *Comput. Biol. Med.* **2022**, *145*, 105430. [CrossRef] [PubMed]
63. Feli, M.; Azimi, I.; Anzanpour, A.; Rahmani, A.M.; Liljeberg, P. An Energy-Efficient Semi-Supervised Approach for on-Device Photoplethysmogram Signal Quality Assessment. *Smart Health* **2023**, *28*, 100390. [CrossRef]
64. Pereira, T.; Ding, C.; Gadhoumi, K.; Tran, N.; Colorado, R.A.; Meisel, K.; Hu, X. Deep Learning Approaches for Plethysmography Signal Quality Assessment in the Presence of Atrial Fibrillation. *Physiol. Meas.* **2019**, *40*, 125002. [CrossRef] [PubMed]
65. Reiss, A.; Indlekofer, I.; Schmidt, P.; Van Laerhoven, K. Deep PPG: Large-Scale Heart Rate Estimation with Convolutional Neural Networks. *Sensors* **2019**, *19*, 3079. [CrossRef]
66. Schmidt, P.; Reiss, A.; Duerichen, R.; Marberger, C.; Van Laerhoven, K. Introducing WESAD, a Multimodal Dataset for Wearable Stress and Affect Detection. In Proceedings of the 20th ACM International Conference on Multimodal Interaction, ICMI '18, New York, NY, USA, 16–20 October 2018; pp. 400–408. [CrossRef]
67. Stark, Z. Chengstark, Segade, 2024. Available online: https://github.com/chengstark/Segade (accessed on 10 May 2023).
68. Py-Feat: Python Facial Expression Analysis Toolbox—Py-Feat. Available online: https://py-feat.org/pages/intro.html# (accessed on 12 September 2023).
69. Deng, J.; Guo, J.; Ververas, E.; Kotsia, I.; Zafeiriou, S. RetinaFace: Single-Shot Multi-Level Face Localisation in the Wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5203–5212.
70. Li, X.; Alikhani, I.; Shi, J.; Seppanen, T.; Junttila, J.; Majamaa-Voltti, K.; Tulppo, M.; Zhao, G. The OBF Database: A Large Face Video Database for Remote Physiological Signal Measurement and Atrial Fibrillation Detection. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 242–249. [CrossRef]
71. Hendrycks, D.; Mu, N.; Cubuk, E.D.; Zoph, B.; Gilmer, J.; Lakshminarayanan, B. AugMix: A Simple Data Processing Method to Improve Robustness and Uncertainty. *arXiv* **2020**, arXiv:1912.02781. [CrossRef]
72. Pytorchvideo.Transforms—PyTorchVideo Documentation. Available online: https://pytorchvideo.readthedocs.io/en/latest/api/transforms/transforms.html (accessed on 20 November 2020).
73. Orphanidou, C.; Bonnici, T.; Charlton, P.; Clifton, D.; Vallance, D.; Tarassenko, L. Signal-Quality Indices for the Electrocardiogram and Photoplethysmogram: Derivation and Applications to Wireless Monitoring. *IEEE J. Biomed. Health Inform.* **2015**, *19*, 832–838. [CrossRef] [PubMed]
74. Liu, X.; Narayanswamy, G.; Paruchuri, A.; Zhang, X.; Tang, J.; Zhang, Y.; Sengupta, R.; Patel, S.; Wang, Y.; McDuff, D. rPPG-Toolbox: Deep Remote PPG Toolbox. *Adv. Neural Inf. Process. Syst.* **2024**, *36*, 152.
75. Deividbotina-Alv/Rtrppg: Python Implementation of the 3DCNN-based Real-Time rPPG Network (RTrPPG). Available online: https://github.com/deividbotina-alv/rtrppg (accessed on 14 September 2023).
76. Yu, Z.; Peng, W.; Li, X.; Hong, X.; Zhao, G. Remote Heart Rate Measurement from Highly Compressed Facial Videos: An End-to-End Deep Learning Solution with Video Enhancement. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 151–160.
77. Joshi, J.; Bianchi-Berthouze, N.; Cho, Y. Self-Adversarial Multi-scale Contrastive Learning for Semantic Segmentation of Thermal Facial Images. In Proceedings of the 33rd British Machine Vision Conference 2022, London, UK, 21–24 November 2022; p. 864.
78. Yu, Z.; Li, X.; Niu, X.; Shi, J.; Zhao, G. AutoHR: A Strong End-to-End Baseline for Remote Heart Rate Measurement with Neural Searching. *IEEE Signal Process. Lett.* **2020**, *27*, 1245–1249. [CrossRef]