Improving Video-based Heart Rate and Respiratory Rate Estimation via Pulse-Respiration Quotient

Yuzhuo Ren [1] **Braeden Syrnyk** [1 2] **Niranjan Avadhanam** [1]

Abstract

Remote physiological measurement, *e.g.*, heart rate and respiratory rate measurement, becomes more and more important when contact instrument-based measurement is inaccessible and non-preferable under the COVID-19 pandemic. Non-contact camera based physiological measurement enables Telehealth, remote health monitoring and smart hospital applications. Remote physiological signal measurement has challenges such as environment illumination variations, head motion, facial expression, etc. We propose a convolutional neural network to jointly estimate heart rate and respiratory rate with camera video as input in a multitask fashion, which leverages the correlation between heart rate and respiratory rate. Specifically, we propose a novel loss function which integrates the frequency correlation between heart rate and respiratory rate to improve robustness of both heart rate and respiratory rate estimation. Furthermore, we propose a post processing filter based on correlation between heart rate and respiratory rate which further improve prediction accuracy. Extensive experiments demonstrate that our proposed system significantly improves heart rate and respiratory rate measurement accuracy.

1. Introduction

Remote physiological measurement research draws significant attention especially remote health monitoring becomes preferable during the COVID-19 pandemic. Heart rate and respiratory rate are the most important physiological signals. Measuring heart rate and respiratory rate is the essential step to identifying many diseases. Cameras are one of the popular sensors for remote heart rate and respiratory rate estimation. The underlying principle for camera based heart rate and respiratory rate measurement is capturing subtle skin color changes (Wu et al., 2012) or subtle motions (Balakrishnan et al., 2013) caused by blood circulation and respiration. Pulse oximeter uses Photoplethysmography (PPG) technology to detect blood volume changes by measuring light absorption from the skin. Imaging Photoplethysmography (iPPG), also called remote Photoplethysmography (rPPG), technology is based on the measurement of subtle changes in light reflected from the skin from camera images in non-contact way. Skin color change and motion caused by blood circulation and respiration are so subtle which makes camera based heart rate and respiratory rate estimation challenging, especially under uncontrolled environments such as lightening variations, facial expression, head motion, etc.

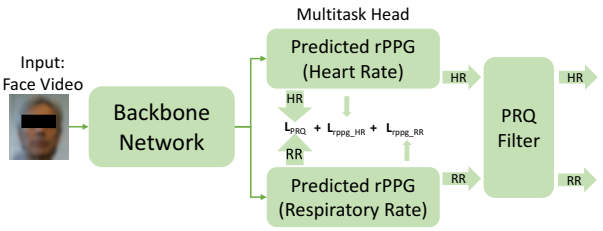

Figure 1. Overview of our proposed neural network to jointly estimate heart rate and respiratory rate leveraging correlation between heart rate and respiratory rate, *i.e.*, Pulse-Respiration Quotient (PRQ) loss. Our PRQ loss can be applied to any backbone multitask network architecture to further improve heart rate and respiratory rate estimation accuracy. We also propose on using the PRQ information as a post-processing filter to further improve heart rate and respiratory rate accuracy.

Traditional heart rate and respiratory rate estimation uses face tracking (Wang et al., 2016) and/or skin segmentation (Tasli et al., 2014) to detect a region of interest (ROI) which contains strong physiological signals, such as cheek, forehead, nose regions, and then extract color changes from these ROI across frames, with frequency bandpass filtering (Wang et al., 2017) and principle component analysis (PCA) (Lewandowska et al., 2011) to filter noise and extract heart rate or respiratory rate. To improve the algorithm robustness under a noise environment and leverage big data

[1]NVIDIA, Santa Clara, USA [2]University of Waterloo, Ontario, Canada. Correspondence to: Yuzhuo Ren <yren@nvidia.com>.

Proceedings of the 1st Workshop on Healthcare AI and COVID-19, ICML 2022.

and supervised learning, a deep learning based method has been proposed (Chen & McDuff, 2018) to enable end-to-end learning of the heart rate and respiratory rate. Deep learning based methods outperform traditional handcrafted feature based methods especially for challenging scenarios, such as head motion, under compression artifacts (Nowara et al., 2021), etc. Especially, multitask learning (Liu et al., 2020) enables learning heart rate and respiratory rate jointly using the same backbone network but with two heads, one for heart rate, another for respiratory rate. Multitask learning reduces memory by using one network to estimate both of the signals. However, the loss is simply the summation of the heart rate rPPG waveform Mean Squared Error (MSE) loss and respiratory rate rPPG waveform MSE loss which results in an accuracy drop because the multitask network is trained to optimize on two tasks rather than a dedicated network to optimize on a single task. Leveraging the correlation between heart rate and respiratory rate to improve multitask learning accuracy has never been explored.

In our work, we propose a Pulse-Respiration Quotient (PRQ) loss to integrate into the multitask loss to improve multitask learning accuracy. PRQ loss takes the correlation between the heart rate and respiratory rate into account to make the network optimize PRQ as well as rPPG. Furthermore, we propose on using the PRQ information to further remove inaccurate heart rate or respiratory rate estimation as a post-processing step. We demonstrate in experiments that adding heart rate and respiratory waveform PRQ loss into multitask loss improves network accuracy and using PRQ information as a post processing filter can further improve the result accuracy.

2. Related Work

2.1. Traditional Hand-crafted Feature Method

Lewandowska (Lewandowska et al., 2011) proposed using channel selection and the PCA algorithm to separate heart rate signal and noise. CHROM (De Haan & Jeanne, 2013) method leverages light absorption differences among R, G and B channels to conduct noise reduction among RGB channels to improve physiological signal robustness. Wang (Wang et al., 2015) improved robustness of the CHROM method by using spatial redundancy of image sensor to improve motion robustness. POS (Wang et al., 2016) method extracts the pulse using a projection plane orthogonal to the skin tone. Wang (Wang et al., 2017) proposed sub-band pulse extraction to suppress periodic motions to particularly improve heart rate estimation robustness in fitness scenarios. RGBIR sensor has also been proposed for physiological signal estimation in order to leverage an additional IR channel to improve signal robustness and reduce noise (Wang & den Brinker, 2020).

2.2. Convolutional Neural Network Method

Recent CNN based solutions provide an end-to-end solution for physiological signal estimation and enables large scale data training to improve physiological signal estimation. Chen (Chen & McDuff, 2018) proposed a Convolutional Attention Network (CAN). CAN takes two consecutive frames' face crop difference as motion map and original frame's face crop as appearance map as inputs to estimate blood volume pulse. Ren (Ren et al., 2021) integrated efficient channel attention (Wang et al., 2020) to Chen's spatial attention network and built a dual attention network to recalibrate features in both spatial domain and feature domain. 3D convolution was proposed in (Liu et al., 2020) to replace 2D convolution in Chen (Chen & McDuff, 2018)'s architecture, although it gives better accuracy, 3D modules complexity is much higher than 2D module. Liu (Liu et al., 2020) further improved Chen (Chen & McDuff, 2018)'s network by adding temporal shift module (TSM) (Lin et al., 2019) and multitask learning for heart rate and respiratory rate estimation, and called it Multitask Temporal Shift Convolutional Attention Network (MTTS-CAN). TSM shifts part of the channels along the temporal dimension to exchange information among neighborhood frames. TSM achieves the accuracy of 3D CNN and maintains 2D CNN's complexity as well. Niu (Niu et al., 2019) proposed to use spatial-temporal map with a deeper backbone network for end-to-end heart rate estimation. Niu (Niu et al., 2020) proposed cross-verified feature disentangling from pairwise face video training to improve heart rate estimation accuracy. Several CNN architectures are proposed to estimate heart rate from a highly compressed video (Rapczynski et al., 2019; Yu et al., 2019; Nowara et al., 2021).

2.3. Loss Function

The loss function of Chen's model (Chen et al., 2018) is the MSE between the estimated and ground truth physiological signal derivative. Liu (Liu et al., 2020) summed heart rate and respiratory rate rPPG MSE loss as the multitask learning loss for joint learning. Niu (Niu et al., 2019) proposed a smooth Mean Absolute Error (MAE) loss function to constrain the smoothness of adjacent heart rate measurements based on the fact that the variance of the subjects' heart rate is small during a very small period of time (Niu et al., 2017). Niu (Niu et al., 2020) combined physiological signal loss, heart rate estimation loss and cross-verified disentangling loss into the loss function. Gideon (Gideon & Stent, 2021) introduced maximum cross-correlation (MCC) as a new loss function for rPPG supervised training which is more robust to synchronization error between video and physiological ground truth. Revanur (Revanur et al., 2022) performed the MCC in the frequency domain instead of time domain.

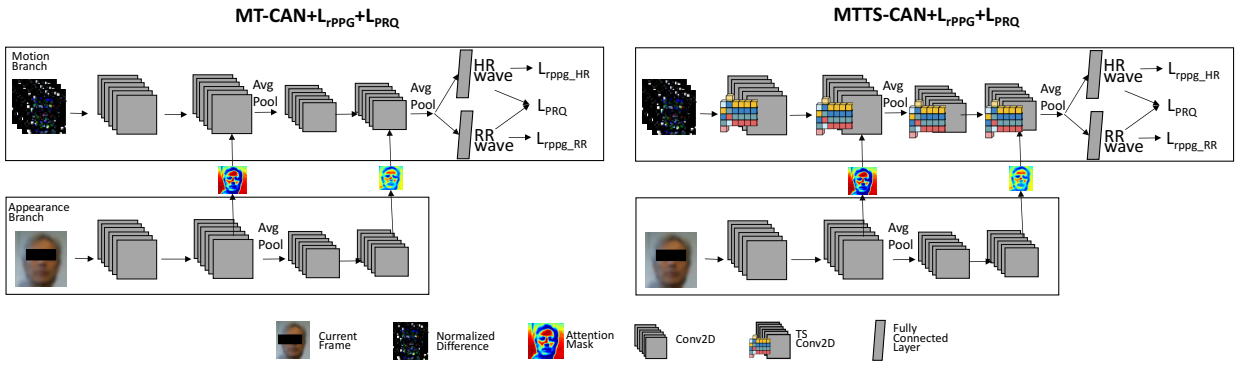

Figure 2. Multitask Convolutional Attention Network (MT-CAN (Liu et al., 2020)) and Multitask Temporal Shift Convolutional Attention Network (MTTS-CAN (Liu et al., 2020)) architecture by adding PRQ loss in multitask learning.

2.4. Heart Rate and Respiratory Rate Correlation

A lof of research (Bahmed et al., 2016; Scholkmann & Wolf, 2019; von Bonin et al., 2014) has been done to investigate the frequency correlation between cardiovascular and respiratory systems, and defined its ratio as the Pulse-Respiration Quotient (PRQ). PRQ has correlation with general health, physical activities, etc. Scholkmann (Scholkmann & Wolf, 2019) found that PRQ may change during different physical activities and different times of the day (non-sleep, sleep), however PRQ is in a certain range regardless of higher or lower heart rate or respiratory rate. In our multitask learning model, we leverage the correlation between heart rate and respiratory rate and integrate the PRQ loss into multitask learning to improve both heart rate and respiratory rate.

3. Methodology

3.1. Skin Reflection Model

For the theoretical optical principle of the deep neural network model to estimate rPPG, Shafer's dichromatic reflection model (DRM) (Wang et al., 2016) is applied to model the lighting and physiological signals (Chen & McDuff, 2018; Liu et al., 2020). RGB value of the k-th skin pixel in an image can be defined by a time-varying function:

$$\begin{aligned}\mathbf{C}_k(t) = &\underbrace{I_0 \cdot (1 + \Psi(m(t), \Theta(p(t), r(t))))}_{I(t)} \\ &\cdot (\underbrace{\mathbf{u}_s \cdot (s_0 + \Phi(m(t), \Theta(p(t), r(t))))}_{\mathbf{v}_s(t)} \\ &+ \underbrace{\mathbf{u}_d \cdot d_0 + \mathbf{u}_p \cdot \Theta(p(t), r(t))}_{\mathbf{v}_d(t)}) + \mathbf{v}_n(t)\end{aligned} \tag{1}$$

where $\mathbf{C}_k(t)$ denotes a vector of the RGB values; $I(t)$ is the illuminace intensity; $\mathbf{v}_s(t)$ and $\mathbf{v}_d(t)$ are specular and diffusion reflection respectively; $\mathbf{v}_n(t)$ denotes camera sensor's quantization noise. $I(t)$, $\mathbf{v}_s(t)$ and $\mathbf{v}_d(t)$ can all be decomposed into stationary part (*i.e.* I_0, $\mathbf{u}_s \cdot s_0$, $\mathbf{u}_d \cdot d_0$) and time-varying part (*i.e.* $I_0 \cdot \Psi(\cdot)$, $\mathbf{u}_s \cdot \Phi(\cdot)$, $\mathbf{u}_p \cdot \Theta(\cdot)$) (Wang et al., 2016). $\Theta(p(t), r(t))$ denotes a combination of both pulse $p(t)$ and respiration $r(t)$ signal. The relation between $\mathbf{C}_k(t)$ and $\Theta(p(t), r(t))$ is non-linear and the non-linear is caused by illuminance variation, head motion, facial expression, camera compression, etc. $m(t)$ denotes all non-physiological variations; $\Psi(\cdot)$ denotes the intensity variation observed by camera; $\Phi(\cdot)$ denotes the specular reflections varying parts; $\mathbf{u}_s$ and $\mathbf{u}_d$ denotes the unit color vector of the light source and skin-tissue respectively; $\mathbf{u}_p$ denotes the relative pulse strengths. I_0 denotes stationary part of illuminance intensity; s_0 and d_0 denotes the stationary specular and diffusion reflection respectively. Deep learning based heart rate and respiration rate estimation methods (Chen & McDuff, 2018; Liu et al., 2020) model the relation between $\mathbf{C}_k(t)$ and $\Theta(p(t), r(t))$ by supervision from training data. For example, adding head motion data into training can make the neural network predict heart rate and respiratory rate more accurately under the head motion case.

3.2. Architecture

We use MT-CAN and MTTS-CAN (Liu et al., 2020) as the backbone network for multitask training. Their network architectures are shown in Figure 2. MT-CAN is a two branch network, *i.e.* motion branch and appearance branch, with spatial attention applied from appearance branch. Motion branch takes N consecutive frames' face ROI difference as input. Appearance branch takes current frame's face ROI as input. There are two spatial attention layers that get multiplied to motion branch to select informative spatial features. Spatial attention mask can help localize the rPPG signal since the strength of the rPPG signal on face differs with spatial location. Forehead and cheek regions have stronger heart rate rPPG signal while nose and neck have stronger

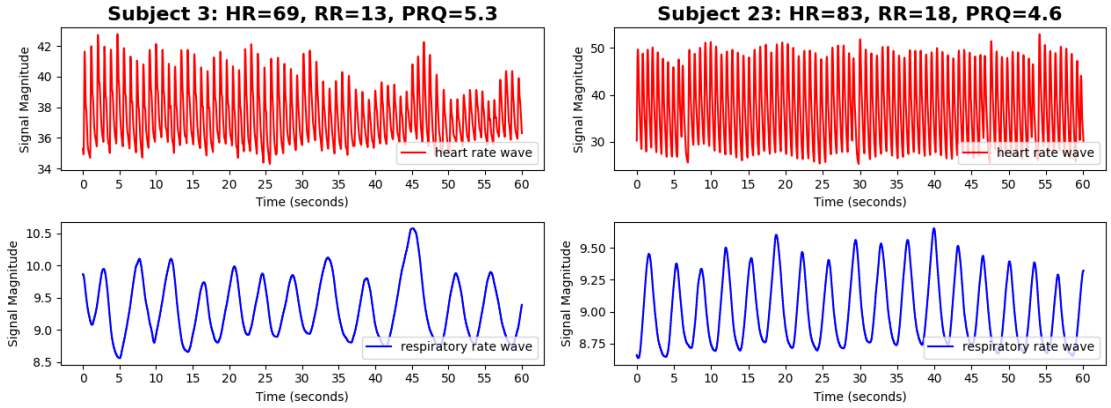

Figure 3. Ground truth heart rate and respiratory rate waveform from two subjects in COHFACE dataset. Horizontal axis is recording time in second, vertical axis is the signal magnitude from ground truth sensor recording.

respiratory rate rPPG signal. Compared to MT-CAN, temporal shift module (Lin et al., 2019) is applied before 2D convolution layers in the motion branch in MTTS-CAN. Temporal shit module helps incorporate temporal information when predicting rPPG which makes rPPG signal more robust. Different from (Liu et al., 2020), the novel part of our network architecture is that we add heart rate and respiratory rate ratio error as a loss function, *i.e.*, L_{PRQ} to multitask loss besides rPPG loss from heart rate and respiratory rate. The L_{PRQ} reinforces the correlation between heart rate branch and respiratory branch to maintain the heart rate and respiratory rate ratio in a valid range and to avoid the network over optimize towards one branch.

3.3. Loss Function

3.3.1. Heart Rate and Respiratory Rate rPPG Loss

L_{rPPG}, defined in Equation (2), is the summation of heart rate rPPG loss $L_{\text{rPPG\_HR}}$ and respiratory rPPG loss $L_{\text{rPPG\_RR}}$. L_{rPPG} measures the MSE error between predicted and ground truth rPPG physiological signal.

$$L_{\text{rPPG}} = \alpha \underbrace{\frac{1}{T}\sum_{t=1}^{T}(p(t) - p(t)^{'})^2}_{L_{\text{rPPG\_HR}}} + \beta \underbrace{\frac{1}{T}\sum_{t=1}^{T}(r(t) - r(t)^{'})^2}_{L_{\text{rPPG\_RR}}}, \tag{2}$$

where T is the time window, $p(t)$ and $r(t)$ are time variant pulse rPPG sequence and respiratory rPPG sequence respectively.

3.3.2. Pulse-Respiration Quotient (PRQ) Loss

Figure 3 visualizes two subjects heart rate and respiratory rate ground truth physiological signal recording. We can see the trend that subject with lower heart rate has lower respiratory rate. Figure 4 plots heart rate and respiratory rate values for each subject in COHFACE dataset. Heart rate and respiratory rate are weakly linear correlated, *i.e.*, when heart rate increases respiratory rate increases. Heart rate and respiratory rate Pearson correlation is 0.3, which is computed using the following equation,

$$\rho_{\text{Pulse,Respiration}} = \frac{Cov(\text{Pulse}, \text{Respiration})}{\sigma_{\text{Pulse}}\sigma_{\text{Respiration}}}, \tag{3}$$

where Cov is the co-variance, σ is the standard deviation. Figure 5 plots the PRQ histogram for all the subjects in COHFACE dataset. $\text{PRQ}_{\text{max}} = 14.83$, $\text{PRQ}_{\text{min}} = 2.54$, $\text{PRQ}_{\text{mean}} = 5.57$, most of the subjects have $\text{PRQ} = 4$ which aligns with the research (Bahmed et al., 2016; Scholkmann & Wolf, 2019; von Bonin et al., 2014) that PRQ maintains within a certain range regardless of heart rate and respiratory rate values.

We define PRQ loss in Equation (4), which is the mean absolute error (MAE) between predicted PRQ and ground truth PRQ. Our intuition is that if the predicted PRQ and ground truth PRQ has large difference, then it's likely that either of the signal is not predicted accurately.

$$L_{\text{PRQ}} = |\text{PRQ}_{\text{pred}} - \text{PRQ}_{\text{GT}}|, \tag{4}$$

where

$$\text{PRQ}_{\text{pred}} = \frac{\text{HR}_{\text{pred}}}{\text{RR}_{\text{pred}}}$$

$$\text{PRQ}_{\text{GT}} = \frac{\text{HR}_{\text{GT}}}{\text{RR}_{\text{GT}}}$$

HR and RR are computed by first applying bandpass filter ([0.67, 4]Hz for HR, [0.08, 0.5]Hz for RR) and then extract dominant frequency component using frequency analysis method.

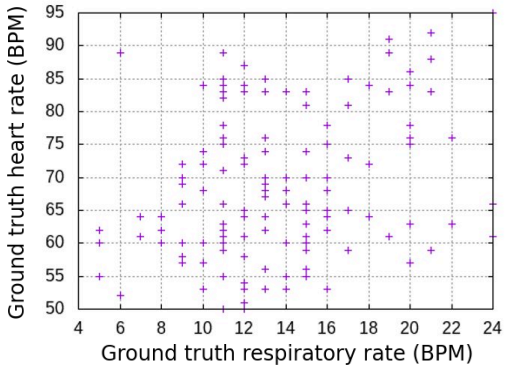

Figure 4. Heart rate and respiratory rate dot plot. Each dot represents ground truth heart rate and respiratory rate extracted from the one minute recording from COHFACE dataset.

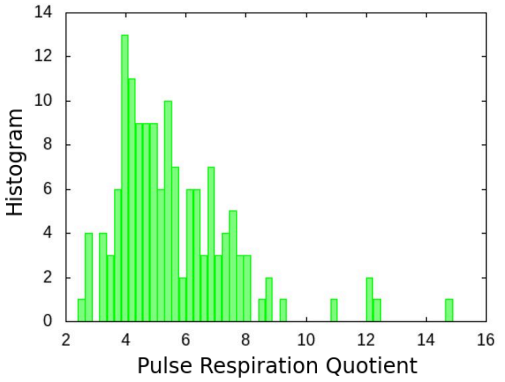

Figure 5. Pulse-Respiration Quotient (PRQ) Histogram of data in COHFACE dataset.

3.3.3. Multitask Loss

The multitask loss proposed by Liu *et al.* (Liu et al., 2020) is simply the summation of heart rate and respiratory rate rPPG loss, which does not take the correlation between heart rate and respiratory rate into account. There can be case when the neural network has lower loss for heart rate rPPG and respiratory rate rPPG estimation in multitask learning, however, PRQ is not in the valid range. Please note that PRQ loss alone is not enough for supervising heart rate and respiratory rate joint training, since $\text{PRQ} = \frac{\text{HR}}{\text{RR}}$ is a ratio, there exists a constant k that would result in the same PRQ as $\text{PRQ} = \frac{k*\text{HR}}{k*\text{RR}}$. Therefore, for example, if $(\text{HR}_{\text{pred}}, \text{RR}_{\text{pred}}) = (60, 10)$ and $(\text{HR}_{\text{GT}}, \text{RR}_{\text{GT}}) = (120, 20)$, the PRQ loss would give zero error, in which case network could not learn meaningful information with PRQ loss only. Thus, PRQ loss has to be combined with rPPG loss to improve the accuracy. Incorporating both rPPG loss and PRQ loss into multitask learning loss can make neural network back propagate the rPPG loss and PRQ loss and make predicted rPPG aligned with ground truth rPPG meanwhile predicted PRQ does not deviate from ground truth PRQ. We integrate rPPG loss (Equation (2)) and PRQ loss (Equation (4)) into the multitask loss. Our multitask loss is defined in Equation (5),

$$L = L_{\text{rPPG}} + \gamma L_{\text{PRQ}}, \tag{5}$$

where multitask loss L is the summation of rPPG loss L_{rPPG} defined in Equation (2) and PRQ loss L_{PRQ} defined in Equation (4). γ is empirical parameter to balance L_{rPPG} and L_{PRQ}. We set $\alpha = \beta = \gamma = 1$ in our experiment.

3.4. PRQ Post Processing Filter

The multitask neural network output is the pulse waveform sequence and respiratory waveform sequence. To extract the heart rate and respiratory rate in beats per minute, we first applied bandpass filter (cut-off frequencies of 0.67 and 4 Hz for heart rate, and 0.08 and 0.50 Hz for respiratory rate). 10 second window is used as evaluation window to apply the Fourier transform to get the dominant frequencies as the heart rate and respiratory rate. We further remove the predicted heart rate and respiratory result with PRQ outside range [2.5, 16.0] based on the dataset statistics as shown in Figure 5. Using PRQ as a post processing helps reduce false positive prediction, *i.e.*, instead of showing wrong heart rate and respiratory rate estimation for current evaluation window, we use previous evaluation window's result until we get reliable prediction result which within valid PRQ range.

4. Experiments

We compare our methods with two networks for heart rate measurement and respiratory rate measurement: multitask convolutional attention network (MT-CAN) and multitask temporal shift network (MTTS-TAN) (Liu et al., 2020). Other than MT-CAN and MTTS-CAN, we are not aware of other networks that estimates heart rate and respiratory rate together in a multitask fashion.

4.1. Datasets

We run our experiments using COHFACE dataset(Heusch et al., 2017) since the dataset contains RGB videos of faces, synchronized with heart rate and respiratory rate wave of the recorded subjects. The face video is recorded with a Logitech webcam with resolution 640x480 pixels. The camera frame rate is 20Hz. Heart rate and respiratory rate waveform ground truth are recorded at 256 Hz. The dataset contains 40 subjects, in total 160 one-minute long video sequences. There are 4 videos from every client: 2 videos

Table 1. Heart rate and respiratory rate testing data accuracy benchmark with MT-CAN (Liu et al., 2020) on COHFACE dataset.

Settings	Heart Rate (HR)			Respiratory Rate (RR)			HR and RR Average		
	MAE	SNR	Availability	MAE	SNR	Availability	MAE	SNR	Availability
L_{rPPG}	11.016	2.177	0.684	4.217	15.534	0.980	7.617	8.855	0.832
$L_{\text{rPPG}} + F_{\text{PRQ}}$	8.470	2.233	0.688	4.071	15.557	0.988	6.271	8.895	0.838
$L_{\text{rPPG}} + L_{\text{PRQ}}$	6.727	2.772	0.735	4.115	15.159	0.972	5.421	8.965	0.854
$L_{\text{rPPG}} + L_{\text{PRQ}} + F_{\text{PRQ}}$	4.601	2.880	0.755	4.083	15.311	0.984	4.342	9.096	0.870

Table 2. Heart rate and respiratory rate testing data accuracy benchmark with MTTS-CAN(Liu et al., 2020) on COHFACE dataset.

Settings	Heart Rate (HR)			Respiratory Rate (RR)			HR and RR Average		
	MAE	SNR	Availability	MAE	SNR	Availability	MAE	SNR	Availability
L_{rPPG}	1.621	6.756	0.945	2.514	13.554	1.000	2.067	10.155	0.972
$L_{\text{rPPG}} + F_{\text{PRQ}}$	1.625	6.747	0.945	2.482	13.600	1.000	2.053	10.174	0.972
$L_{\text{rPPG}} + L_{\text{PRQ}}$	1.621	6.756	0.945	2.514	13.555	1.000	2.067	10.155	0.972
$L_{\text{rPPG}} + L_{\text{PRQ}} + F_{\text{PRQ}}$	1.625	6.747	0.945	2.482	13.601	1.000	2.053	10.174	0.972

with controlled conditions, another 2 videos with more natural conditions. Natural condition videos include lightening change and subjects natural head movement. We follow the training and testing data split protocol provided by COHFACE dataset in our experiment.

4.2. Experiment Details

We use the COHFACE(Heusch et al., 2017) dataset to evaluate our proposed heart rate and respiratory rate PRQ loss function and post processing filter. We use OpenCV face detector to get face crop and resize it to 72x72. And we sample the ground truth heart rate wave and breath rate wave at 20 Hz so that it is aligned with camera frame rate. Motion map is current frame and previous frame's face crop subtraction. Appearance map is the current frame's face crop. Both motion map and appearance map are normalized in the video.

To have a fair comparison with previous work, we use the same backbone architecture to train MT-CAN and MTTS-CAN. We use the same motion map and appearance map and the same post processing steps (same cut off frequencies for bandpass filter and frequency analysis method). The training parameters (learning rate, epoch, etc) are the same for fair comparison. The only difference is whether to apply PRQ loss function and PRQ post processing filter.

We did ablation on our proposed PRQ loss function and post processing filter. We compare accuracy by applying PRQ loss function and post processing PRQ filter in different network backbones to show how adding PRQ loss to multitask loss and applying post processing filter could improve network accuracy.

4.3. Evaluation Metrics

The evaluation metrics were computed over all windows of all the test videos in a dataset, we used 10 seconds as evaluation window. We use the following metrics: (1) **Mean Absolute Error (MAE)**: The average absolute error between ground truth heart/respiratory rate and predicted heart/respiratory rate. (2) **Signal-to-Noise Ratio (SNR)**: We calculate blood volume pulse and respiration signal-to-noise ratios (SNR) according to the method proposed by De Haan (De Haan & Jeanne, 2013). The SNR is calculated in the frequency domain as the ratio between the energy around the first two harmonics and remaining frequencies within heart rate and respiratory rate frequency range. SNR captures the quality of predicted heart rate and respiratory rate. (3) **Availability**: We compute the percentage of SNR ≥ 0 in a video as availability. SNR < 0 indicates predicted heart/respiratory rate is not reliable since signal energy is less than noise energy. This metric captures percentage of the time the system is able to predict high quality heart/respiratory rate.

5. Results and Discussion

We did ablation study on our PRQ loss function and post processing PRQ filter using various network backbone and show that the PRQ loss and the PRQ filter's capability of improving heart rate and respiratory rate estimation accuracy. Specifically, we compare network testing accuracy with and without the PRQ loss function and post processing PRQ filter. We evaluated heart rate and respiratory rate estimation accuracy using the following rPPG loss (denoted as L_{rPPG}), PRQ loss function (denoted as L_{PRQ}) and PRQ filter (denoted as F_{PRQ}) combinations, which are in Table 1 and Table 2's row 1, row 2, row 3 and row 4 respectively:

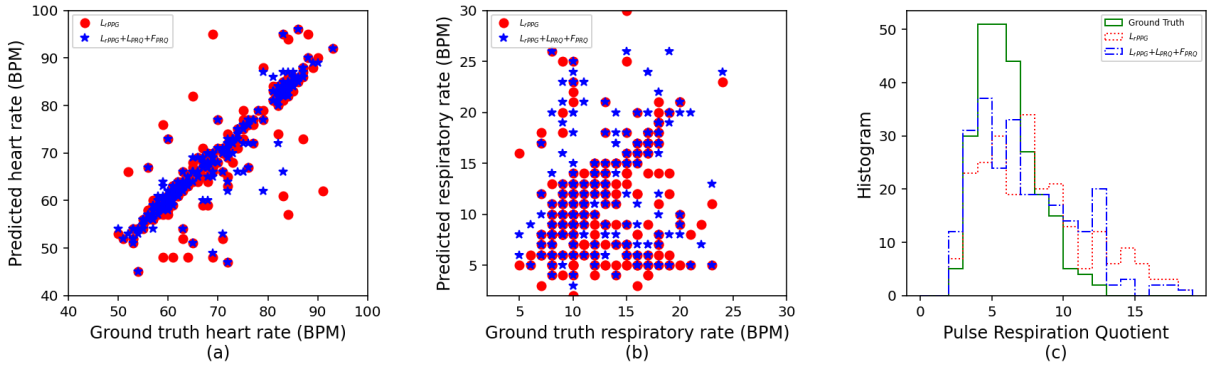

Figure 6. Comparison between L_{rPPG} and $L_{\text{rPPG}} + L_{\text{PRQ}} + F_{\text{PRQ}}$ setting using MT-CAN. (a): The scatter plot of the ground truth HR and the predicted HR on COHFACE dataset. (b): The scatter plot of the ground truth RR and the predicted RR on COHFACE dataset. (c): PRQ histogram.

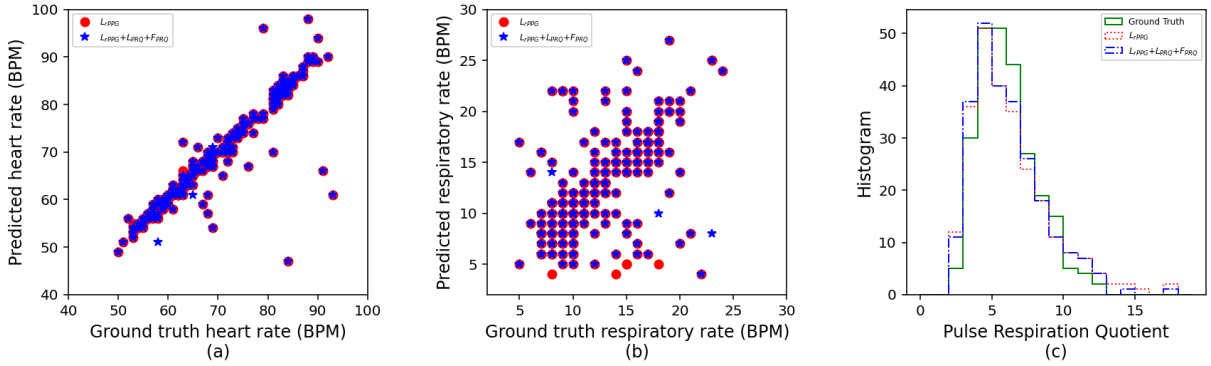

Figure 7. Comparison between L_{rPPG} and $L_{\text{rPPG}} + L_{\text{PRQ}} + F_{\text{PRQ}}$ setting using MTTS-CAN. (a): The scatter plot of the ground truth HR and the predicted HR on COHFACE dataset. (b): The scatter plot of the ground truth RR and the predicted RR on COHFACE dataset. (c): PRQ histogram.

1. $\boldsymbol{L_{\text{rPPG}}}$: Apply rPPG loss as multitask loss, which is the sum of MSE loss of heart rate rPPG loss and respiratory rate rPPG loss, defined in Equation (2).

2. $\boldsymbol{L_{\text{rPPG}} + F_{\text{PRQ}}}$: Apply rPPG loss as multitask loss defined in Equation (2) and also apply post processing PRQ filter. PRQ filter implementation detail is described in Section 3.4.

3. $\boldsymbol{L_{\text{rPPG}} + L_{\text{PRQ}}}$: Apply the multitask loss defined in Equation (5), *i.e.*, use the sum of rPPG loss and PRQ loss as multitask loss.

4. $\boldsymbol{L_{\text{rPPG}} + L_{\text{PRQ}} + F_{\text{PRQ}}}$: Apply the multitask loss defined in Equation (5) and also apply post processing PRQ filter. PRQ filter implementation detail is described in Section 3.4.

Under each of the above four experiment settings, we compare heart rate and respiratory rate estimation MAE, SNR and Availability.

5.1. Ablation on L_{PRQ} and F_{PRQ} using MT-CAN as Backbone Network

Table 1 first row and third row compares without and with L_{PRQ} using MT-CAN as backbone network on COHFACE test dataset. Adding L_{PRQ} on top of L_{rPPG} greatly reduces HR MAE from 11.016 to 6.727 and reduces RR MAE from 4.217 to 4.115. Since HR rPPG and RR rPPG are jointly optimized in multitask learning, thus we average evaluation metrics for HR and RR, which are in the third column. HR and RR averaged MAE is reduced from 7.617 to 5.421; averaged SNR increases from 8.855 to 8.965; averaged availability increases from 0.832 to 0.854. The results show that L_{PRQ} could improve network accuracy by imposing HR and RR's correlation during training.

Table 1 first row and second row compares without and with F_{PRQ}. Adding F_{PRQ} on top of L_{rPPG} reduces HR MAE from 11.016 to 8.470 and reduces RR MAE from 4.217 to 4.071. HR and RR averaged MAE is reduced from 7.617 to 6.271; averaged SNR increases from 8.855 to 8.895; averaged availability increases from 0.832 to 0.838. Therefore, F_{PRQ} also improves HR and RR accuracy.

Table 1 fourth row shows the result by adding both L_{PRQ} and F_{PRQ} on top of L_{rPPG}. For this experiment setting, we would like to evaluate whether applying both L_{PRQ} and F_{PRQ} could further improve result compared to apply L_{PRQ} or F_{PRQ} individually. HR and RR averaged MAE further reduces to 4.342; Averaged SNR further increases to 9.096; Averaged availability further increases to 0.870. This result shows that L_{PRQ} and F_{PRQ} can improve HR and RR estimation accuracy. Furthermore, applying both L_{PRQ} and F_{PRQ} can achieve better accuracy improvement than applying L_{PRQ} or F_{PRQ} individually, which indicates that L_{PRQ} and F_{PRQ} is complementary on reducing inaccurate HR and RR predictions.

Figure 6 (a) shows scatter plot of ground truth and predicted HR values from L_{rPPG} setting and $L_{\text{rPPG}} + L_{\text{PRQ}} + F_{\text{PRQ}}$ setting. Figure 6 (b) shows scatter plot of ground truth and predicted RR values from L_{rPPG} setting and $L_{\text{rPPG}}+L_{\text{PRQ}}+F_{\text{PRQ}}$ setting. $L_{\text{rPPG}} + L_{\text{PRQ}} + F_{\text{PRQ}}$ setting gives better HR and RR estimation result with higher correlation with ground truth. Figure 6 (c) shows PRQ histogram of L_{rPPG} setting and $L_{\text{rPPG}} + L_{\text{PRQ}} + F_{\text{PRQ}}$ setting compared to ground truth. $L_{\text{rPPG}} + L_{\text{PRQ}} + F_{\text{PRQ}}$ setting gives better correlation with ground truth, especaillly we can see that $L_{\text{rPPG}}+L_{\text{PRQ}}+F_{\text{PRQ}}$ setting reduces outlier PRQ outside ground truth PRQ range (*i.e.*, PRQ larger than 16.0 which is based on COHFACE dataset statistics, as shown in Figure 5). The predictions outside valid PRQ range even after applying F_{PRQ} are those at the beginning of the video when there's no previous reliable prediction to propagate to current evaluation window.

5.2. Ablation on L_{PRQ} and F_{PRQ} using MTTS-CAN as Backbone Network

Table 2 first row and third row compares without and with PRQ loss function using MTTS-CAN as backbone network on COHFACE test dataset. The accuracy does not improve because MTTS-CAN is a stronger backbone network and has much higher accuracy than MT-CAN. Therefore, there's not much outlier training samples outside valid PRQ range which L_{PRQ} could help.

Table 2 first row and second row compares without and with F_{PRQ}. F_{PRQ} reduces averaged HR and RR MAE from 2.067 to 2.053 and increases averaged SNR from 10.155 to 10.174. As post processing filter, F_{PRQ} is able to improve HR and RR estimation accuracy by removing outlier estimation outside valid PRQ range.

Figure 7 shows scatter plot of ground truth and predicted HR values and RR values, as well as PRQ histogram using MTTS-CAN backbone network. $L_{\text{rPPG}}+L_{\text{PRQ}}+F_{\text{PRQ}}$ setting gives better accuracy than L_{rPPG} setting since F_{PRQ} reduces predicted HR and RR outside valid PRQ range.

5.3. Discussion

Both L_{PRQ} and F_{PRQ} can reduce HR and RR prediction error by leveraging the correlation between HR and RR. L_{PRQ} leverages the correlation by back-propagating PRQ loss during neural network training while F_{PRQ} leverages the correlation by replacing wrong HR and RR outside valid PRQ range with neighborhood reliable prediction in a video. For the time window when neural network is not able to predict HR and RR accurately which can be due to under challenging environment (*e.g.*, low lightening, head motion, compression artifacts, etc), F_{PRQ} could reduce false predictions by maintaining PRQ within a reasonable range.

Both L_{PRQ} and F_{PRQ} provide higher accuracy improvement for the weaker backbone network (*i.e.*, MT-CAN) than the stronger backbone network (*i.e.*, MTTS-CAN). F_{PRQ} can improve accuracy for stronger backbone network MTTS-CAN. This demonstrates that both L_{PRQ} and F_{PRQ} can be integrated to improve prediction accuracy especially for the weaker backbone and for the harder training samples. For the weaker backbone network and the harder training samples, L_{PRQ} helps the network learn the hard training data which otherwise L_{rPPG} could not differentiate.

6. Conclusions

We proposed leveraging correlation between heart rate and respiratory rate to improve remote camera-based heart rate and respiratory rate estimation accuracy. We evaluated various ways of leveraging heart rate and respiratory rate correlation, including a novel PRQ loss function which can be inserted to any multitask learning backbone and a PRQ filter as a post processing to remove false estimation result outside a valid PRQ range. We integrated PRQ loss function into multitask learning network to back propagate PRQ loss to jointly optimize heart rate and respiratory rate. It was demonstrated by experimental results that the heart rate and respiratory rate correlation improve the estimation accuracy on benchmark dataset.

7. Acknowledgements

We thank Zhiding Yu for the insightful discussion and providing feedback on multitask learning and the discussion on multitask loss function.

References

Bahmed, F., Khatoon, F., Reddy, B. R., and Bahmed, F. Relation between respiratory rate and heart rate–a comparative study. *Indian Journal of Clinical Anatomy and Physiology*, 3(4):436–439, 2016.

Balakrishnan, G., Durand, F., and Guttag, J. Detecting pulse from head motions in video. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3430–3437, 2013.

Chen, W. and McDuff, D. Deepphys: Video-based physiological measurement using convolutional attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pp. 349–365, 2018.

Chen, X., Cheng, J., Song, R., Liu, Y., Ward, R., and Wang, Z. J. Video-based heart rate measurement: Recent advances and future prospects. *IEEE Transactions on Instrumentation and Measurement*, 68(10):3600–3615, 2018.

De Haan, G. and Jeanne, V. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.

Gideon, J. and Stent, S. The way to my heart is through contrastive learning: Remote photoplethysmography from unlabelled video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 3995–4004, 2021.

Heusch, G., Anjos, A., and Marcel, S. A reproducible study on remote heart rate measurement. *arXiv preprint arXiv:1709.00962*, 2017.

Lewandowska, M., Rumiński, J., Kocejko, T., and Nowak, J. 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)*, pp. 405–410. IEEE, 2011.

Lin, J., Gan, C., and Han, S. Tsm: Temporal shift module for efficient video understanding. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 7083–7093, 2019.

Liu, X., Fromm, J., Patel, S., and McDuff, D. Multi-task temporal shift attention networks for on-device contactless vitals measurement. *arXiv preprint arXiv:2006.03790*, 2020.

Niu, X., Han, H., Shan, S., and Chen, X. Continuous heart rate measurement from face: A robust rppg approach with distribution learning. In *2017 IEEE International Joint Conference on Biometrics (IJCB)*, pp. 642–650. IEEE, 2017.

Niu, X., Shan, S., Han, H., and Chen, X. Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation. *IEEE Transactions on Image Processing*, 29:2409–2423, 2019.

Niu, X., Yu, Z., Han, H., Li, X., Shan, S., and Zhao, G. Video-based remote physiological measurement via cross-verified feature disentangling. In *European Conference on Computer Vision*, pp. 295–310. Springer, 2020.

Nowara, E. M., McDuff, D., and Veeraraghavan, A. Systematic analysis of video-based pulse measurement from compressed videos. *Biomedical Optics Express*, 12(1): 494–508, 2021.

Rapczynski, M., Werner, P., and Al-Hamadi, A. Effects of video encoding on camera-based heart rate estimation. *IEEE Transactions on Biomedical Engineering*, 66(12): 3360–3370, 2019.

Ren, Y., Syrnyk, B., and Avadhanam, N. Dual attention network for heart rate and respiratory rate estimation. In *2021 IEEE 23rd International Workshop on Multimedia Signal Processing (MMSP)*, pp. 1–6. IEEE, 2021.

Revanur, A., Dasari, A., Tucker, C. S., and Jeni, L. A. Instantaneous physiological estimation using video transformers. *arXiv preprint arXiv:2202.12368*, 2022.

Scholkmann, F. and Wolf, U. The pulse-respiration quotient: A powerful but untapped parameter for modern studies about human physiology and pathophysiology. *Frontiers in physiology*, 10:371, 2019.

Tasli, H. E., Gudi, A., and den Uyl, M. Remote ppg based vital sign measurement using adaptive facial regions. In *2014 IEEE international conference on image processing (ICIP)*, pp. 1410–1414. IEEE, 2014.

von Bonin, D., Grote, V., Buri, C., Cysarz, D., Heusser, P., Moser, M., Wolf, U., and Laederach, K. Adaption of cardio-respiratory balance during day-rest compared to deep sleep—an indicator for quality of life? *Psychiatry research*, 219(3):638–644, 2014.

Wang, Q., Wu, B., Zhu, P., Li, P., Zuo, W., and Hu, Q. Eca-net: Efficient channel attention for deep convolutional neural networks, 2020 ieee. In *CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE*, 2020.

Wang, W. and den Brinker, A. C. Modified rgb cameras for infrared remote-ppg. *IEEE Transactions on Biomedical Engineering*, 67(10):2893–2904, 2020.

Wang, W., Stuijk, S., and De Haan, G. A novel algorithm for remote photoplethysmography: Spatial subspace rotation. *IEEE transactions on biomedical engineering*, 63 (9):1974–1984, 2015.

Wang, W., den Brinker, A. C., Stuijk, S., and De Haan, G. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2016.

Wang, W., den Brinker, A. C., Stuijk, S., and de Haan, G. Robust heart rate from fitness videos. *Physiological measurement*, 38(6):1023, 2017.

Wu, H.-Y., Rubinstein, M., Shih, E., Guttag, J., Durand, F., and Freeman, W. Eulerian video magnification for revealing subtle changes in the world. *ACM transactions on graphics (TOG)*, 31(4):1–8, 2012.

Yu, Z., Peng, W., Li, X., Hong, X., and 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*, pp. 151–160, 2019.