Deep learning-based image enhancement for robust remote photoplethysmography in various illumination scenarios

Shutao Chen[2]*, Sui Kei Ho[1], Jing Wei Chin[2], Kin Ho Luo[2], Tsz Tai Chan[1,2], Richard H.Y. So[1,2], Kwan Long Wong[1,2]

[1]The Hong Kong University of Science and Technology, [2]PanopticAI Ltd.

Abstract

Remote photoplethysmography (rPPG) is a non-invasive and convenient approach for measuring human vital signs using a camera. However, accurate measurement can be challenging due to the different illumination of the surrounding environment. In this study, we present a deep learning-based image enhancement model (IEM) inspired by the Retinex theory to improve the robustness of rPPG signal extraction and heart rate (HR) estimation in different lighting conditions. We fine-tuned the IEM with a time-shifted negative Pearson correlation between the PPG signal ground truth and the predicted rPPG signal from a pre-trained 3D CNN (PhysNet). We evaluated our method using a publicly available dataset (BH-rPPG) of various lighting scenarios and our own private dataset. Our results demonstrate that our proposed model is generalizable and significantly improves rPPG extraction and HR estimation accuracies across a range of illumination intensities.

1. Introduction

Vital sign measurement from a remote photoplethysmography (rPPG) signal has gained significant attention in recent years due to its non-invasive and convenient nature [5, 15]. RPPG enables contactless vital sign monitoring, which is particularly advantageous in situations where physical contact with the patient is not desirable, such as in intensive care units, when highly contagious disease is present, or in remote telemedicine consultations outbreaks [6, 7, 21]. However, robust rPPG signal extraction is challenging due to factors such as motion artifacts, different skin types, and the lighting conditions of the surrounding environment [22, 23, 25, 28].

Traditional methods, such as POS [22] and CHROM [8]), have been proposed to eliminate the influence of illumination intensity on rPPG signal extraction. However, the actual performance of these methods is unsatisfactory when tested on a publicly available dataset, as shown in Fig.1. When the focus shifts to deep learning-based methods, the issues become more apparent. Most existing deep learning studies for HR estimation from rPPG are based on supervised training of 2D [4] or 3D [27] convolutional neural networks (CNNs). While these models have shown remarkable accuracy in HR estimation compared to traditional methods, they tend to be influenced by different illumination intensities and lack generalization to different environmental illuminations. The performance of these models tends to degrade when the test environment becomes very dark or bright [26].

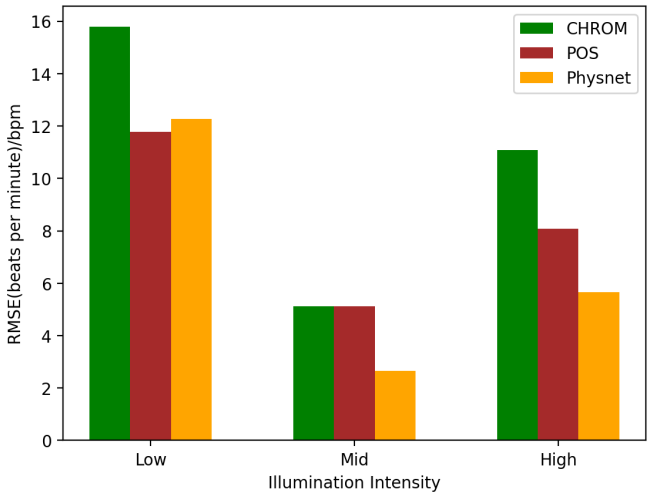

Figure 1. HR estimation error of traditional (CHROM and POS) and deep learning (PhysNet) rPPG extraction methods for subjects from the BH-rPPG dataset under low (8 lux), mid (42.4 lux), and high (104 lux) illuminations. All methods performed well under normal lighting conditions but degraded when the illumination becomes low or high.

Shafer's dichromatic model hypothesizes that the reflection components of diffuse and specular reflections are

*First author

modulated by the illumination of the surrounding environment [22]. Since the rPPG signal is measured by capturing reflectance of subtle changes in skin color caused by variations in blood volume, we hypothesize that image enhancement methods can isolate the illumination factor and improve the performance of rPPG extraction under low-light conditions [11]. In this study, we propose to combine a pretrained image enhancement model (IEM) with an rPPG extraction network to overcome the challenges of robust rPPG signal extraction and heart rate estimation in environment with different levels of illumination. We use a state-of-the-art rPPG extraction model to guide task-driven transfer learning of the IEM to recover the diffuse reflection component containing physiological rPPG information from the raw video by decomposition of the illumination mask. We focus on image-wise instead of video-wise enhancement based on the considerations that the illumination of consecutive frames will not vary greatly, and that employment of video-wise enhancement might be costly.

The paper is organized as follows. In Section 2, we give some background about rPPG extraction-related work. In Section 3, we introduce our proposed Image Enhancement Model + PhysNet framework and ShiftLoss function for fine-tuning. Three experiments are designed to validate the performance of our proposed methods under different illumination and the results are shown in Section 4 and Section 5.

The contributions of our paper are as follows:

- We proposed a novel deep learning-based IEM + rPPG framework for robust rPPG signal extraction and heart rate estimation in different illuminations. We validate our method on publicly available and private datasets and our model significantly outperforms existing state-of-the-art methods.
- To address the misalignment between the ground truth PPG signal and predicted rPPG signal during transfer learning of the IEM, we propose a novel loss function that minimizes the time-shifted negative Pearson correlation.

2. Related work

2.1. Traditional methods for rPPG extraction

Several traditional methods based on signal processing and linear operations have been proposed for rPPG extraction, such as ICA, POS, CHROM, GREEN, LGI, PBV [8, 9, 17, 18, 20, 22]. Each method employs a skin model or mathematical assumptions to extract the rPPG signal. For example, ICA [18] decomposes signals into a linear mixture of sources based on independence and non-Gaussianity assumptions. The CHROM [8] algorithm proposes a dichromatic reflection model to remove irrelevant specular reflections. Similarly, POS [22] removes specular reflections by utilizing a plane orthogonal to the skin tone in a normalized RGB space. Methods like POS and CHROM account for the influence of illumination intensities by taking the temporal average of consecutive frames within a short time window. Consequently, traditional methods are relatively more robust to changes in environmental illumination than deep learning-based methods, particularly when the measurement noise is negligible. The authors of [3] propose an rPPG toolbox that includes a summary, evaluation, and implementation of popular rPPG signal processing methods.

2.2. Deep learning-based rPPG extraction methods

Most current deep learning studies on HR estimation from rPPG involve training a 2D or 3D Convolutional Neural Network (CNN) in a supervised manner using pairs of videos and their corresponding PPG or BVP signals. DeepPhys [4] utilizes motion and appearance 2D CNN models to process input videos and combines their outputs to predict the final rPPG signal. The authors propose a normalized frame difference as their input to the motion model to capture the spatial-temporal information of the input videos. PhysNet [27] employs a 3D CNN to learn rPPG features in both the spatial and temporal domains and uses a temporal encoder-decoder network to exploit temporal information more effectively. The authors conducted ablation studies to demonstrate the performance gain by the temporal encoder-decoder and the superior performance of 3DCNN over 2DCNN+RNN. While deep learning methods show superior performance under normal illumination, their performance declines significantly when models trained under normal illumination are tested on low-illumination videos.

2.3. rPPG extraction in low illumination

Many rPPG datasets were collected in laboratory settings with sufficient illumination, which may not reflect real-life situations. Even in indoor environments with stationary participants, existing analysis [24] indicates that low-illumination videos can degrade the performance of all previously mentioned methods. Yang et al. [26] conducted an assessment of deep learning-based heart rate estimation using rPPG of subjects under different illuminations. They investigated the effects of unsupervised image enhancement methods or data augmentation during model training to overcome the poor performance of deep learning models, but the results were unsatisfactory.

3. Methods

3.1. Dichromatic model

Shafer's dichromatic model [22] has been widely acknowledged as accurately modeling the optical flow for rPPG signal extraction. It assumes that the reflection of light from the skin is composed of specular reflection and

diffuse reflection, which is modulated by the illumination intensities as shown in the following equation,

$$C(t) = I(t) \cdot (v_s(t) + v_d(t)) + v_n(t) \tag{1}$$

where $I(t)$ is the time-varying illumination intensity, $v_s(t)$ is the specular reflection, $v_d(t)$ is the diffuse reflection and $v_n(t) \sim \mathcal{N}(0, \sigma^2)$ stands for the noise during measurement. The pulsatile rPPG signal is included within the diffuse reflection.

For deep learning-based methods, the variation of pixel values in different scenarios makes deep learning-based models harder to generalize [26]. To reduce the effect, we employ the normalized frame difference in [4], which is defined as,

$$\tilde{C}(t) = \frac{C(t) - C(t-1)}{C(t) + C(t-1)}, \qquad t \geq 1 \tag{2}$$

The frame difference reduces the influence of motion artifacts and background pixels. It also provides temporal context to the underlying network. Based on empirical data shown in Table 1, we found that the normalized frame difference (to zero mean and unit variance) could improve HR estimation accuracy to some extent. We utilized normalized frame difference as an input to the framework.

Method	MAE/bpm	RMSE/bpm	PEARSON
PhysNet (Original Standardized frame)	1.904	4.758	0.964
PhysNet (Normalized frame difference)	**0.963**	**2.411**	**0.990**

Table 1. The HR estimation accuracy in terms of Mean Absolute Error (MAE), Root-Mean-Square Error (RMSE) and Pearson correlation for PhysNet model with standardized original input and normalized frame difference on the UBFC-rPPG dataset in 2-fold cross validation. The PhysNet model with normalized frame difference achieved better performance on HR estimation accuracy.

3.2. Traditional Image Enhancement Methods

Traditional image enhancement has been applied to preprocess raw images with image enhancement methods to standardize them for rPPG extraction [26]. We selected Gamma Correction (GC) [19] and Histogram Equalization (HE) [1] as representations for traditional methods. The gamma value was set to 2.5 and 0.8 for low and high-light conditions, respectively, as suggested in [26].

3.3. Deep learning-based Image Enhancement Model

There is extensive literature on general low-illumination image enhancement using deep learning methods. A literature review with benchmarking can be found at [11]. Many state-of-the-art studies were based on Retinex theory [10], according to which the color image received by the camera $C(t)$ can be decomposed into two factors: reflectance $R(t)$, which refers to the appearance of the object; and illumination map $I(t)$, which are highly dependent on the surrounding light of the environment. The relation of the two factors is as followed,

$$C(t) = I(t) \cdot R(t) \tag{3}$$

where $C(t)$ is the raw image received by the photographic source, and $I(t)$ and $R(t)$ stand for the illumination map and reflection, respectively.

Compared with the dichromatic model, we found that the desired rPPG signal is contained in the reflectance. Therefore, we propose that fine-tuning the existing deep learning image enhancement modules using data from the low-illumination domain would boost the performance of existing rPPG neural networks. We applied the architecture in Fig.2, using pretrained weights in the initial image enhancement module and keeping the rPPG extractor frozen. However, the preliminary result showed that the pretrained model cannot be directly used to improve rPPG extraction performance as the model was trained for general image enhancement purposes. Therefore, we fine-tuned the original model to enable the module to reduce gaps between the low-illumination domain and training domain of the rPPG extractor while keeping the gain in illumination.

We made use of a lightweight image enhancement network recently proposed in [14]. In the original work, the generalization of the model was considered, and the authors suggested to fine-tune the model to fit into different domains. The model is a good entry point to assess the capability of general image enhancement modules for the specific task of enhancing low-illumination facial videos for rPPG extraction. In later sections, this network is referred to as the Image Enhancement Model (IEM) and we fine-tune it for rPPG extraction.

3.4. Transfer learning for rPPG extraction

In our work, we selected PhysNet [27] for its superior performance in rPPG extraction for videos captured under normal lighting conditions. PhysNet is cascaded after the IEM and all the parameters of PhysNet are frozen during the retraining process. Given the input video frames, the reflectance within each frame is extracted by the IEM and then fed into the PhysNet to predict the rPPG signal. Heart rate is estimated by taking the peak absolute values of the Fourier transform of the rPPG signal. The negative Pear-

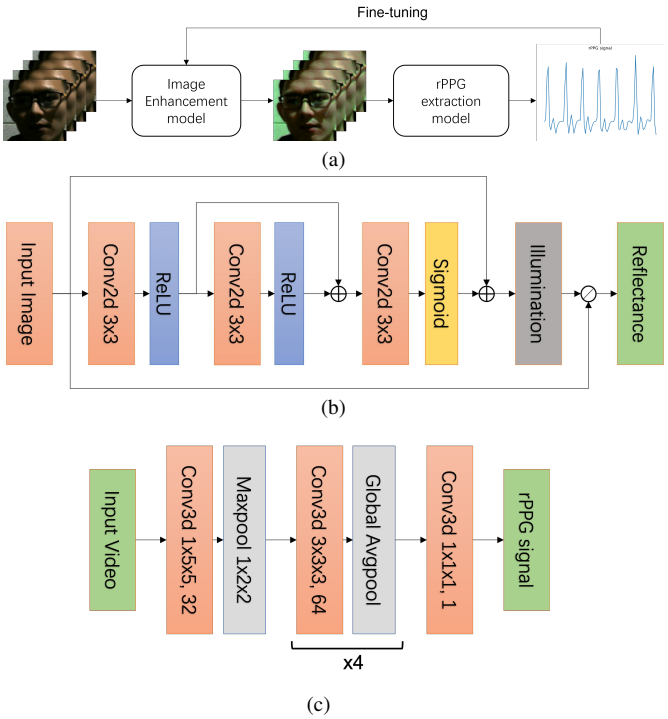

Figure 2. (a) Flow chart of the proposed framework: The input face images are processed by the image enhancement model, in which the outputs are fed into the rPPG extraction network to predict the rPPG. Prediction loss is back-propagated to fine-tune the image enhancement model. (b) Structure of Image enhancement model. (c) Structure of PhysNet rPPG extraction model.

son correlation (NPC) between the rPPG signal and ground truth PPG signal is used as a loss function to back-propagate to the IEM for fine-tuning its parameters. This loss function is defined as,

$$Loss_{NPC} = 1 - \frac{\sum_{t=0}^{T}(s_{rPPG}(t) - \bar{s}_{rPPG})(s_{PPG}(t) - \bar{s}_{PPG})}{\sqrt{\sigma(s_{rPPG}) \cdot \sigma(s_{PPG})}} \quad (4)$$

where $s_{rPPG}(t)$ is the predicted rPPG signal, $s_{PPG}(t)$ is the ground truth PPG signal and $\sigma(\cdot)$ is the standard deviation operator. We used NPC as a loss function because our target is to estimate the HR by Fourier transform and NPC penalizes more heavily on frequency prediction errors rather than errors on instantaneous values between the waves.

Using this process, we found that the PPG signal is not always aligned with the video in time, which leads to bad performance of the current NPC loss as shown in Figure 3. To solve this problem, we used a time-shifting NPC loss function, which is defined as,

$$Loss_{ShiftNPC} = 1 - \arg max_{\tau \in (-\delta_t, \delta_t)} \frac{\sum_{t=0}^{T}(s_{rPPG}(t+\tau) - \bar{s}_{rPPG})(s_{PPG}(t) - \bar{s}_{PPG})}{\sqrt{\sigma(s_{rPPG}) \cdot \sigma(s_{PPG})}} \quad (5)$$

The time-shifted NPC, which we refer to as ShiftLoss in later sections, makes it possible to handle misalignment between the PPG and rPPG signals with a time shift of no more than δ_t seconds. However, extra noise may be introduced even when the PPG signal and videos are perfectly aligned, and the max shift parameter δ_t should be carefully designed during the training. In our experiments, δ_t is fixed at 1/3 seconds. ShiftLoss should be used after obtaining a satisfactory rPPG extraction model, as this loss is not capable of finding the best shift of the rPPG signal due to the back-propagation mechanism. For an initialized rPPG extraction model, this loss may degrade performance.

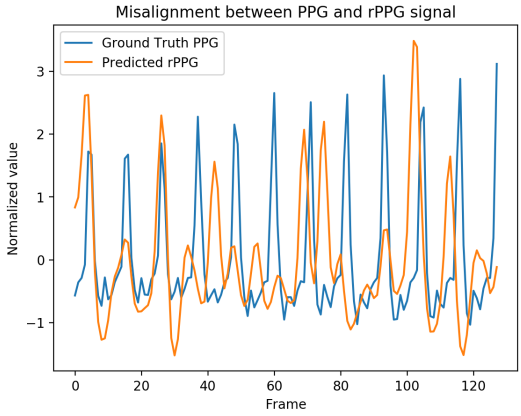

Figure 3. Ground truth PPG signal and rPPG signal predicted by PhysNet from selected input videos. Note the obvious phase shift between the PPG and rPPG signals.

It should be noted that the misalignment between predicted rPPG and ground truth PPG still exists objectively, but time-shifting NPC loss makes it more robust to predict heart rate.

3.5. rPPG-Toolbox

Different data pre/post-processing methods and training procedures can give significantly different results even for the same deep learning model. To resolve these complications, we use rPPG-ToolBox, which is a public GitHub repository at [16], in all of our training. The toolbox contains data loaders of different datasets, trainers for different deep learning models, rPPG extraction and evaluation scripts, and so forth. The introduction in [12] describes a reliable codebase for the replication of results and benchmarks for different models, databases, and training procedures.

Our adaptation unifies many aspects within the training including the Adam optimizer and learning rate scheduler used in training, and the same 2nd-order Butterworth filter (cut-off frequencies of 0.75 and 2.5 Hz) to filter the predicted PPG waveform. This allows a fair comparison of the proposed method and easy reproduction of our results.

4. Experimental Setup

4.1. Dataset

UBFC-rPPG dataset: UBFC-rPPG dataset [2] contains 42 videos of different subjects playing games in a bright environment. Recordings were made with a webcam (Logitech C920 HD pro) at 30fps with a resolution of 640x480 pixels in uncompressed 8-bit RGB format. A CMS50E transmissive pulse oximeter was used to obtain the ground truth PPG data.

BH-rPPG dataset: BH-rPPG dataset [26] contains 105 videos of 35 subjects in 3 different illuminations: low, medium, and high-light corresponding to the luminosities of 8, 42.4, and 104 lux, respectively. The videos were recorded using a webcam (Logitech HD pro-C310 color) with a resolution of 640×480 pixels and fps of around 15Hz. The PPG signal was collected using an oximeter (CONTEC CMS50E).

Private dataset: We collected a private dataset of 12 subjects under two illumination intensities: 25 lux and 100 lux. The videos were recorded using the Logitech C270 webcam with a resolution of 640×480 pixels and fps of 15Hz, in which each condition was recorded for 30 seconds. The PPG signal was collected using Philip's reusable oximeter (M1196A).

4.2. Training settings

The UBFC-rPPG dataset was used to generate the Baseline PhysNet. The other two datasets were used for transfer learning of the image enhancement model and testing. We trained the Baseline PhysNet with UBFC-rPPG dataset [2] as suggested in the original paper, and the same subjects did not appear in both the training and testing phases. We also attempted to train on the BH-medium illumination (42.4 lux) dataset, but the results were unsatisfactory as the PPG and rPPG signals did not align well, resulting in a low Pearson correlation between them. According to Figure 1, the Baseline PhysNet showed good performance on the medium illumination videos in the BH-rPPG dataset.

For each subject, the face was cropped using the Mediapipe algorithm [13], which generated 72×72 pixels images. The difference between successive frames was then taken and normalized to zero-mean with unit variance. We fine-tuned the image enhancement model from the pretrained weights provided in the original work for faster convergence and better performance [14]. A 2-fold split of the BH-low dataset was applied to test the performance gain after exposing the enhancement network to the desired domain. The enhancement module was trained with a frozen Baseline PhysNet using the Adam optimizer and OneCycleScheduler of maximum learning rate 3×10^{-4} with 0 weight decay, and the momentum was set to 0.9. The training process consisted of 20 epochs with a batch size of 4, and the model of the best epoch by training loss is used in testing. We used rPPG-toolbox [3] to facilitate implementation of the model.

We designed three experiments to validate and test the proposed framework. In Experiment 1, the IEM was trained and tested for low-light conditions using the BH-rPPG dataset. In Experiment 2, we validated the framework under high-light conditions with the same dataset. In Experiment 3, we tested for generalization by evaluating the model performance on our private dataset containing facial videos recorded by off-the-shelf webcams under both low-light conditions (25 lux) and high-light conditions (100 lux).

4.3. Exp. 1: Model performance on low-illumination intensity videos

In the first experiment, we constructed the IEM + PhysNet framework and conducted the training. We kept the Baseline PhysNet frozen and retrained the IEM from a checkpoint in its original work. The subjects in the BH-rPPG low-light dataset were split into two groups: one group for training and the other group for testing, and vice versa. This was to avoid having the same subject in both the training and testing phases.

The author acknowledges that the improvement partially comes from retraining on the BH-rPPG low-light dataset. Therefore, we also retrained the Baseline PhysNet with the same strategy on the BH-rPPG low-light dataset for a fair comparison. As a baseline, we also tested the effect of traditional image enhancement methods (Gamma Correction and histogram equalization with the same setting as in [26]).

We evaluated the effect of ShiftLoss in this part by comparing the results before retraining, after training with NP loss and training with ShiftLoss.

4.4. Exp. 2: Model performance on high-illumination intensity videos

It was shown from Fig.1 that the performance of PhysNet degraded when the illumination intensity was high. As the IEM is based on Retinex theory that decomposes the input image into illumination and reflectance, we hy- pothesized that the IEM should be robust to high illumination intensity. To test this hypothesis, we evaluated the model on the BH-rPPG high-light dataset. The BH-rPPG high-light dataset contains videos of 35 subjects under 100 lux illumination.

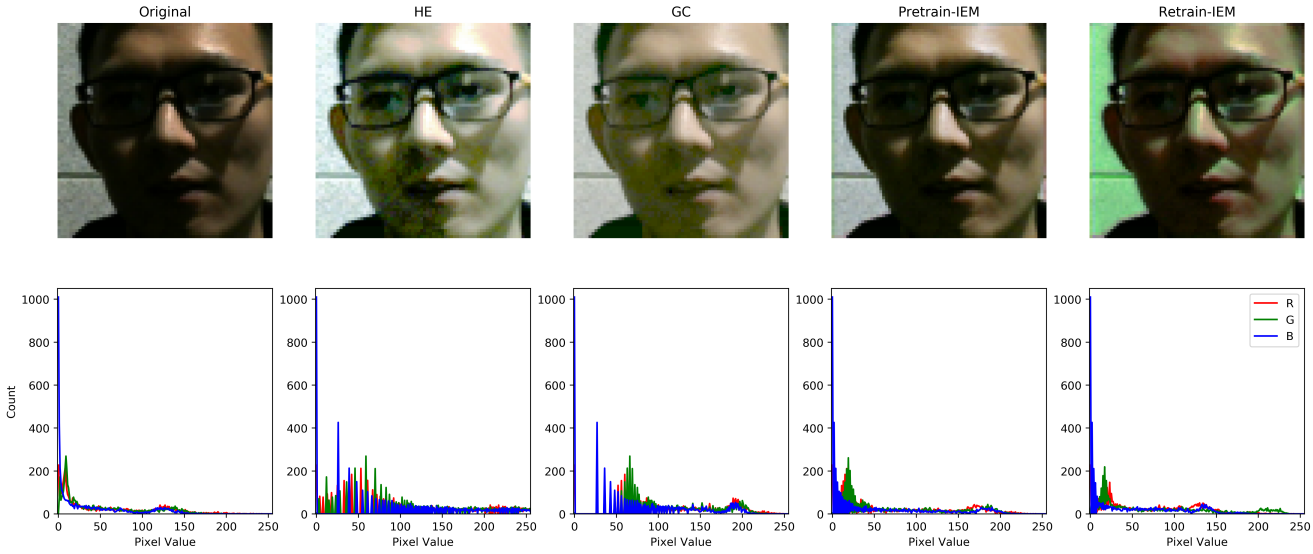

Figure 4. Comparison of output for different image enhancement methods (IEM). The images and their corresponding pixel value histograms of the RGB channel are shown on the top and bottom rows, respectively. The first column demonstrates the original face image of a selected subject. The second column shows the output of the histogram equalization (HE) algorithm. The third column illustrates the image processed with gamma correction (GC) by setting the gamma value to 2.5. The fourth and fifth columns show the outputs of the framework with pretrained IEM and retrained IEM, respectively, for rPPG extraction.

The models tested here were consistent with Experiment 1 and not retrained for high-light conditions.

4.5. Exp. 3: Model performance on private dataset

To assess the generalization of our framework, we tested the performance of the model on our privately collected dataset. The framework came from the previous subsection and is no longer retrained to avoid the possibility that the improvement was because of the retraining process. We used 25 lux and 100 lux as representations for low and high illumination intensity, respectively.

4.6. Metrics

In this experiment, we computed the metrics with respect to the HR estimated from ground truth PPG signal, which included the MAE, RMSE, and Pearson coefficient as defined below:

$$MAE_{HR} = \frac{\sum_k |HR_k^{pred} - HR_k^{gt}|}{K} \tag{6}$$

$$RMSE_{HR} = \sqrt{\frac{\sum_k |HR_k^{pred} - HR_k^{gt}|^2}{K}} \tag{7}$$

$$PC_{HR} = \frac{\sum_k (HR_k^{pred} - \overline{HR}^{pred})(HR_k^{gt} - \overline{HR}^{gt})}{\sqrt{\sigma(HR_k^{pred}) \cdot \sigma(HR_k^{gt})}} \tag{8}$$

5. Results and Analysis

5.1. Exp. 1: Model performance on low-illumination intensity videos

The results are shown in Table 2. The Baseline PhysNet trained on UBFC-rPPG dataset performs badly for the BH-rPPG low-light dataset. This result comes from two reasons: the difference between the measurements in UBFC-rPPG and BH-rPPG dataset as analyzed in [26], and the low-illumination intensity that deviated from UBFC-rPPG dataset. To exclude the former reason, we retrained the PhysNet with the BH-rPPG low-light dataset. The result was significantly worse when NP loss is used. We then checked the extracted rPPG and PPG signal in the temporal axis and noted there was a phase shift between the two signals. We then retrained the PhysNet with ShiftLoss as a loss function and the HR estimation error was reduced by almost 50%. The correlation between the rPPG and PPG also improved significantly.

To test the traditional image enhancement methods, we first preprocessed the video frames with traditional methods by applying Gamma correction ($\gamma = 2.5$) and Histogram equalization before feeding them into PhysNet. The results in Table 2 illustrate these traditional image processing methods usually downgrade the performance. One reason is that these methods can change the distribution of the pixel values as shown in Fig. 4. We used a variety of gamma values, but the performance did not improve.

Method	MAE/bpm	RMSE/bpm	PEARSON
Traditional methods			
CHROM	9.815	15.780	0.347
POS	6.314	11.783	0.585
Baseline PhysNet	5.675	12.261	0.499
GC + Baseline PhysNet	9.392	14.248	0.272
HE + Baseline PhysNet	14.967	18.133	0.216
Transfer learning: PhysNet			
Retrained PhysNet (NP loss)	9.286	12.286	0.110
Retrained PhysNet (ShiftLoss)	**2.974**	**5.101**	**0.888**
Transfer learning: IEM + frozen PhysNet			
Pretrained IEM + Baseline PhysNet	13.232	18.878	-0.058
Retrained IEM (NP loss) + Baseline PhysNet	8.990	14.371	0.236
Retrained IEM (ShiftLoss) + Baseline PhysNet	**2.042**	**3.142**	**0.959**

Table 2. Performance of various rPPG extraction methods on the BH-rPPG low-light (8 lux) rPPG dataset in two transfer learning settings (IEM and non-IEM). For each setting, the performances of the original, and retrained networks via NP loss and ShiftLoss are shown. In both settings, the networks fine-tuned by the Shiftloss algorithm performed the best. Subsequently, they were chosen for the transfer learning stage.

For the pre-trained image enhancement model, we initially tested it without any further training on the dataset. The performance was unimpressive. We ascribed this lacklustre performance to the fact that the original training of the image enhancement model was based on metrics that were manually selected to be more friendly to human vision than to rPPG-related feature extraction. We demonstrated the output of different image enhancement algorithms including histogram equalization (HE), gamma correction (GC), pretrained image enhancement model (IEM) employed in our work before and after fine-tuning for rPPG extraction in Fig.4. The corresponding pixel value histograms are shown under each image respectively. It can be seen from Fig.4 that the distribution of pixel values in RGB color channels varied distinctly among different algorithms. This might explain the vary- ing performances of different models.

We then tested our proposed retraining of the IEM + PhysNet framework. Results illustrate that our model significantly improved the HR estimation accuracy, with the Pearson correlation coefficients exceeding 0.95. The author admits the improvement partially comes from retraining on the BH-rPPG low-light dataset, but when compared with retrained PhysNet with ShiftLoss, IEM + PhysNet still outperforms with the same training and testing data. This proves that applying task-driven fine-tuning of image enhancement with illumination and reflectance decomposition can benefit rPPG extraction and HR estimation accuracies.

Subsections II and III of Table 2 also clearly indicate the effectiveness of ShiftLoss on HR estimation accuracy in both Retrained PhysNet and Retrained IEM + Baseline PhysNet. There would be misalignment problems between the ground truth PPG signal and the rPPG signal predicted by the Baseline PhysNet model with NP loss. In the following study, the models with NP loss are not listed in the table due to the bad performance.

5.2. Exp. 2: Model performance on high-illumination intensity videos

We further compared the performance of different models on the BH-rPPG high-light dataset. As shown in Table 3, the Baseline PhysNet performed satisfactorily under well-lit conditions. However, the PhysNet retrained on BH-rPPG low-light dataset performed poorly. One possible explanation is an over-fit of retraining in low-light conditions. This indicates that solely retraining PhysNet will not make it generalizable to different lighting conditions.

For traditional image enhancement methods, the Gamma Correction (GC) with ($\gamma = 0.8$) seems to have little effect on performance. This is presumably because the saturation of illumination intensities is not affected by GC. In fact, the Histogram Equalization (HE) still significantly downgrades the performance because it changes the distribution of the pixel values as explained previously.

For IEM + PhysNet, it can be seen from the result that, even though this model is trained on a BH-rPPG low-light dataset, it still performs well under high-light conditions. This is supported by the underlying principle that IEM learns to separate illumination maps from the images so it can extract better rPPG-related features under different illumination intensities. It also indicates the feasibility of our proposed IEM + PhysNet to generalize to different illumination conditions.

Method	MAE/bpm	RMSE/bpm	PEARSON
Traditional methods			
CHROM	5.549	11.086	0.645
POS	3.243	8.073	0.802
Baseline PhysNet	2.511	5.649	0.876
GC + Baseline PhysNet	2.511	5.649	0.876
HE + Baseline PhysNet	10.472	16.005	0.251
Transfer learning: PhysNet			
Retrained PhysNet (Shift-Loss)	3.390	6.392	0.835
Transfer learning: IEM + frozen PhysNet			
Pretrained IEM + Baseline PhysNet	10.823	14.989	0.310
Retrained IEM (Shift-Loss) + Baseline PhysNet	**1.783**	**2.287**	**0.986**

Table 3. Performance of various rPPG extraction methods on the BH-high-light (104 lux) rPPG dataset. The Retrained IEM + Baseline PhysNet model achieves the best MAE, RMSE, and Pearson correlation for HR estimation.

5.3. Exp. 3: Model performance on the private dataset

In this experiment, we further tested the performance of retrained Physnet and IEM + PhysNet with baseline original PhysNet using our collected private dataset of low (25 lux) and high (100 lux) illumination intensity environment, and the results are shown in Table 4. It is evident that the Retrained PhysNet performs worst for low illumination intensity(25 lux), and even disastrously for high illumination intensity (100 lux). This resulted from the over-fitting during the retraining of PhysNet on the BH-rPPG dataset.

Our proposed IEM + PhysNet achieves the highest accuracy for both low-light and high-light conditions. For low-light conditions, even though it performs similarly to the original PhysNet, it still performs significantly better than the Retrained PhysNet. As the model is merely trained with BH-rPPG dataset, these results demonstrate the generalization capacity of our proposed IEM + PhysNet.

Method	Illu. /lux	MAE /bpm	RMSE /bpm	PEARSON
Baseline PhysNet	25	2.490	3.624	0.913
	100	2.832	6.508	0.858
Retrained PhysNet (ShiftLoss)	25	3.662	5.346	0.825
	100	14.795	22.607	0.102
Retrained IEM (ShiftLoss) + Baseline PhysNet (ours)	25	**2.344**	**3.516**	**0.924**
	100	**2.100**	**3.162**	**0.966**

Table 4. Performance of various rPPG extraction methods on the private dataset (25 lux and 100 lux): The Retrained IEM + Baseline PhysNet model achieves the best MAE, RMSE and Pearson correlation for HR estimation.

6. Conclusion

In this paper, we propose a deep learning-based IEM to improve rPPG signal extraction and HR estimation for subjects in different illuminations. Our approach utilizes transfer learning from the BH-rPPG low-light dataset, and we introduce a time-shifted NPC loss (ShiftLoss) function to address the issue of misalignment between PPG and videos. Our results demonstrate that our proposed model significantly improves the accuracy of rPPG extraction-based HR estimation under low-light conditions and high-light conditions that deviated from the training dataset. More importantly, our experiments indicate that the improvement in performance is not solely due to retraining on a new dataset, but also due to the advantages of the image enhancement model. Furthermore, our proposed model demonstrated generalization when tested under different illumination intensities using our own collected dataset.

Future work will focus on the domain generalization of the model for different datasets. Additionally, we aim to investigate the combination of the image enhancement model with other rPPG extraction networks to further improve the accuracy of HR estimation. Overall, our proposed method contributes to the advancement of rPPG technology and has the potential to improve the accuracy and convenience of vital sign monitoring in various healthcare and non-healthcare settings.

References

[1] Tinku Acharya and Ajoy K Ray. *Image processing: principles and applications*. John Wiley & Sons, 2005. 3

[2] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 124:82–90, 2019. 5

[3] Giuseppe Boccignone, Donatello Conte, Vittorio Cuculo, Alessandro D'Amelio, Giuliano Grossi, and Raffaella Lanzarotti. An open framework for remote-ppg methods and their assessment. *IEEE Access*, 8:216083–216103, 2020. 2, 5

[4] Weixuan Chen and Daniel McDuff. DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks, Aug. 2018. arXiv:1805.07888 [cs]. 1, 2, 3

[5] Chun-Hong Cheng, Kwan-Long Wong, Jing-Wei Chin, Tsz-Tai Chan, and Richard H. Y. So. Deep Learning Methods for Remote Heart Rate Measurement: A Review and Future Research Agenda. *Sensors*, 21(18):6296, Sept. 2021. 1

[6] Juan-Carlos Cobos-Torres, Mohamed Abderrahim, and José Martínez-Orgado. Non-contact, simple neonatal monitoring by photoplethysmography. *Sensors*, 18(12):4362, 2018. 1

[7] Juan-Carlos Cobos-Torres, Mohamed Abderrahim, and José Martínez-Orgado. Non-contact, simple neonatal monitoring by photoplethysmography. *Sensors*, 18(12):4362, 2018. 1

[8] Gerard De Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013. 1, 2

[9] G De Haan and A Leest. Improved motion robustness of remote-ppg by using the blood volume pulse signature. *Physiological Measurement*, 35, Aug. 2014. 2

[10] Edwin H Land. The retinex theory of color vision. *Scientific american*, 237(6):108–129, 1977. 3

[11] Jiaying Liu, Dejia Xu, Wenhan Yang, Minhao Fan, and Haofeng Huang. Benchmarking low-light image enhancement and beyond. *Int. J. Comput. Vision*, 129(4):1153–1184, apr 2021. 2, 3

[12] Zhang-Xiaoyu Narayanswamy Girish Zhang Yuzhe Wang Yuntao Patel Shwetak McDuff Daniel Liu, Xin. Deep physiological sensing toolbox. Oct. 2022. 4

[13] Camillo Lugaresi, Jiuqiang Tang, Hadon Nash, Chris McClanahan, Esha Uboweja, Michael Hays, Fan Zhang, Chuo-Ling Chang, Ming Guang Yong, Juhyun Lee, Wan-Teh Chang, Wei Hua, Manfred Georg, and Matthias Grundmann. Mediapipe: A framework for building perception pipelines. *CoRR*, abs/1906.08172, 2019. 5

[14] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. Toward Fast, Flexible, and Robust Low-Light Image Enhancement. In *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5627–5636, New Orleans, LA, USA, June 2022. IEEE. 3, 5

[15] Daniel J McDuff, Justin R. Estepp, Alyssa M. Piasecki, and Ethan B. Blackford. A survey of remote optical photoplethysmographic imaging methods. Nov. 2015. 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC) ; Conference date: 25-08-2015 Through 29-08-2015. 1

[16] Tang McJack. rppg-toolbox, 2022. [Online; accessed 2023.03.06]. 4

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

[18] Ming-Zher Poh, Daniel J. McDuff, and Rosalind W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 18(10):10762, May 2010. 2

[19] C. Poynton. *Digital Video and HD: Algorithms and Interfaces*. Electronics & Electrical. Elsevier Science, 2003. 3

[20] Wim Verkruysse and Nelson J Svaasand, Lars and. Remote plethysmographic imaging using ambient light. *Optics Express*, 16, Dec. 2008. 2

[21] Tom Vogels, Mark van Gastel, Wenjin Wang, and Gerard de Haan. Fully-automatic camera-based pulse-oximetry during sleep. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2018. 1

[22] Wenjin Wang, Albertus C. den Brinker, Sander Stuijk, and Gerard de Haan. Algorithmic Principles of Remote PPG. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, July 2017. 1, 2

[23] Lin Xi, Weihai Chen, Changchen Zhao, Xingming Wu, and Jianhua Wang. Image enhancement for remote photoplethysmography in a low-light environment. *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*, 2020. 1

[24] Lin Xi, Xingming Wu, Weihai Chen, Jianhua Wang, and Zhao Changchen. Weighted combination and singular spectrum analysis based remote photoplethysmography pulse extraction in low-light environments. *Medical Engineering Physics*, 105, May 2022. 2

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

[26] Ze Yang, Haofei Wang, and Feng Lu. Assessment of Deep Learning-based Heart Rate Estimation using Remote Photoplethysmography under Different Illuminations, May 2022. arXiv:2107.13193 [cs]. 1, 2, 3, 5, 6

[27] Zitong Yu, Xiaobai Li, and Guoying Zhao. Remote Photoplethysmograph Signal Measurement from Facial Videos Using Spatio-Temporal Networks, July 2019. arXiv:1905.02419 [cs]. 1, 2, 3

[28] Sebastian Zaunseder, Andreas Heinke, Alexander Trumpp, and Hagen Malberg. Heart beat detection and analysis from videos. Aug. 2014. 2014 IEEE 34th International Scientific Conference on Electronics and Nanotechnology (ELNANO) ; Conference date: 15-04-2014 Through 18-04-2014. 1