Augmentation of rPPG Benchmark Datasets: Learning to Remove and Embed rPPG Signals via Double Cycle Consistent Learning from Unpaired Facial Videos (Supplementary Material)

Cheng-Ju Hsieh, Wei-Hao Chung, and Chiou-Ting Hsu

National Tsing Hua University, Hsinchu, Taiwan
peter55180831@gmail.com, godofyax@gmail.com, and cthsu@cs.nthu.edu.tw

1 Overview

In this supplementary, we first demonstrate the necessity of the proposed Removal-Net G_R in Sec. 2.1 and then give more visualized examples of :

- Single-cycle framework (Section 3.1),
- Training without the Removal-Net (Section 3.1),
- Embedding different PPG signals into the same subject (Section 3.2),
- Embedding the same PPG signal into different subjects (Section 3.2), and
- Augmented videos on PURE dataset (Section 3.3).

2 Ablation Study

2.1 Training Without the Removal-Net

To demonstrate the necessity of Removal-Net, we show that, if without G_R, whether the Embedding-Net G_E alone can learn to successfully embed the specified PPG signal into the input videos. In Table 1, the setting "w/o Removal-Net" indicates that we remove G_R and its relevant losses from the total loss. The increased MAE and RMSE show that, although G_E learns to directly embed the specified PPG signal into the input video, the inherent rPPG signal in the input video still remains in the resulting video and thus decreases the rPPG estimation performance.

Table 1. Comparison of w/o and w/ Removal-Net on UBFC-rPPG.

Method	MAE↓	RMSE↓	R↑
Ours (w/o Removal-Net)	1.75	3.47	0.77
Ours (w/ Removal-Net)	0.71	1.48	0.96

3 Visualized Examples

3.1 Comparison with Single-Cycle Framework and Training Without the Removal-Net

In Fig. 1, we compare the perceptual quality of the generated videos when training the proposed RErPPG-NET using single-cycle framework, without the Removel-Net, and using the proposed total loss. Here we embed the original PPG signal of the input video into the rPPG-removed videos and show the resultant videos and the residuals between the input and the output videos.

Comparing Fig. 1 (b) with (d) , we show that the single-cycle framework tends to focus on the consistency of facial contents but is oblivious to the rPPG-related information. As to Fig. 1 (c), if without G_R, although the Embedding-Net G_E alone learns to embed the PPG signal into the input video, both the visual quality and the rPPG estimation performance in Table 1 are poorer than the proposed method.

3.2 Stability of the RErPPG-Net

In Fig. 2, we embed the PPG signals of subjects 33, 22, and 31 into the video of subject 1 on the UBFC-rPPG dataset. The results show that the three rPPG-embedded videos are perceptually indistinguishable from the input one and that their estimated rPPG signals are perfectly correlated with the specified PPG signals.

In Fig. 3, we embed the PPG signal of subject 1 into the videos of subjects 4, 38, and 48 on the UBFC-rPPG dataset. The three rPPG-embedded videos well preserve the perceptual quality of their original ones with only minor difference on the facial area; and the estimated rPPG signals all well aligned with the embedded one with 104.50 bpm.

3.3 Augmented Videos on PURE Dataset

In Fig. 4, we show an example of embedding the PPG signal of subject 7 into the video of subject 1 on PURE dataset. As shown in Fig. 4 (b) and (c), both the rPPG-removed and rPPG-embedded videos are visually indistinguishable from the input one. In Fig. 4 (e) and (f), the estimated rPPG signals from the rPPG-removed and rPPG-embedded videos are highly correlated with the background signal and the ground truth signal, respectively. These results demonstrate that the proposed RErPPG-Net successfully erases the rPPG signal from the input video and embeds the specified PPG signal into the rPPG-removed video.

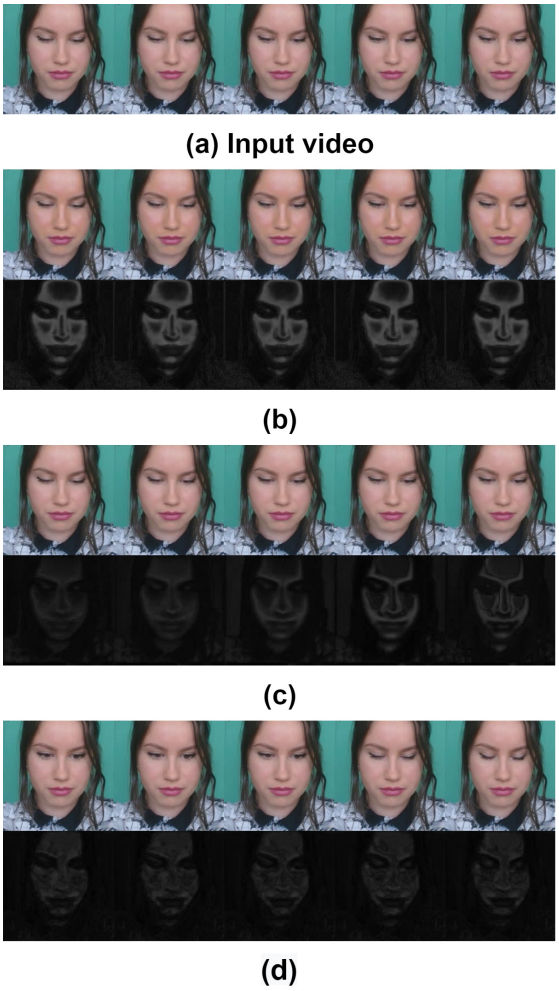

Fig. 1. Visualized examples on UBFC-rPPG dataset, when training the RErPPG-Net: (b) using single-cycle framework; (c) without the Removal-Net G_R; and (d) using the proposed total loss.

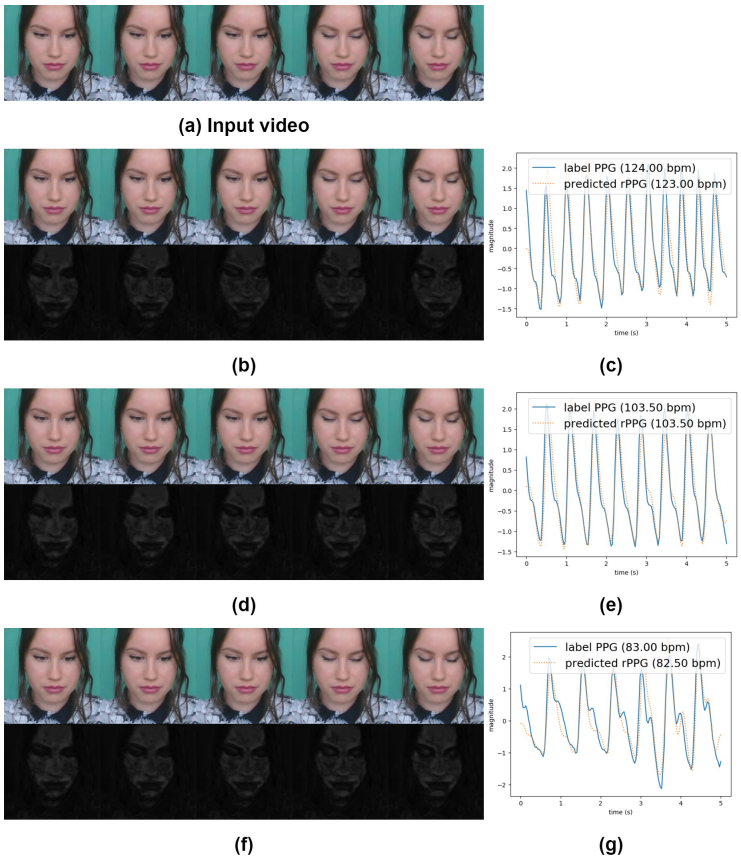

Fig. 2. Results of embedding different PPG signals into the video of subject 1 on UBFC-rPPG dataset. The embedded PPG signals are from: (b) subject 33; (d) subject 22; and (f) subject 31. The ground truth PPG signals (blue) and the predicted rPPG signals (orange) of (b), (d), and (f) are shown in (c), (e), and (g), respectively.

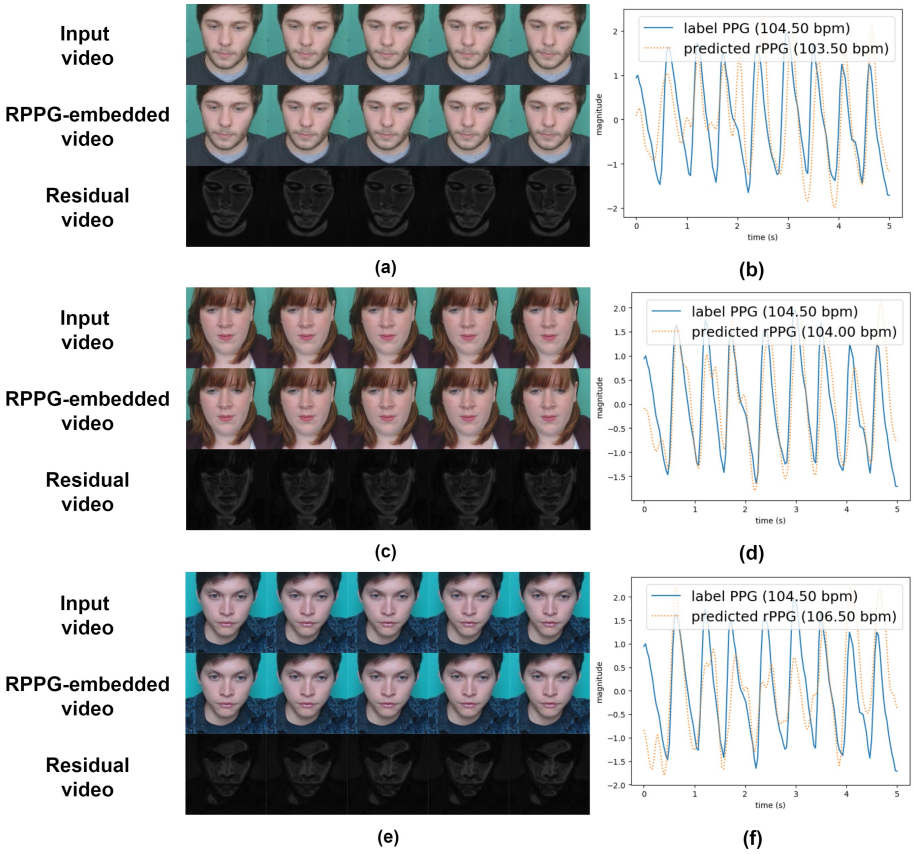

Fig. 3. Results of embedding the PPG signal from subject 1 into (a) subject 4; (c) subject 38; and (e) subject 49. The ground truth PPG signals and the predicted rPPG signals (orange) of (a), (c), and (e) are shown in (b), (d), and (f), respectively.

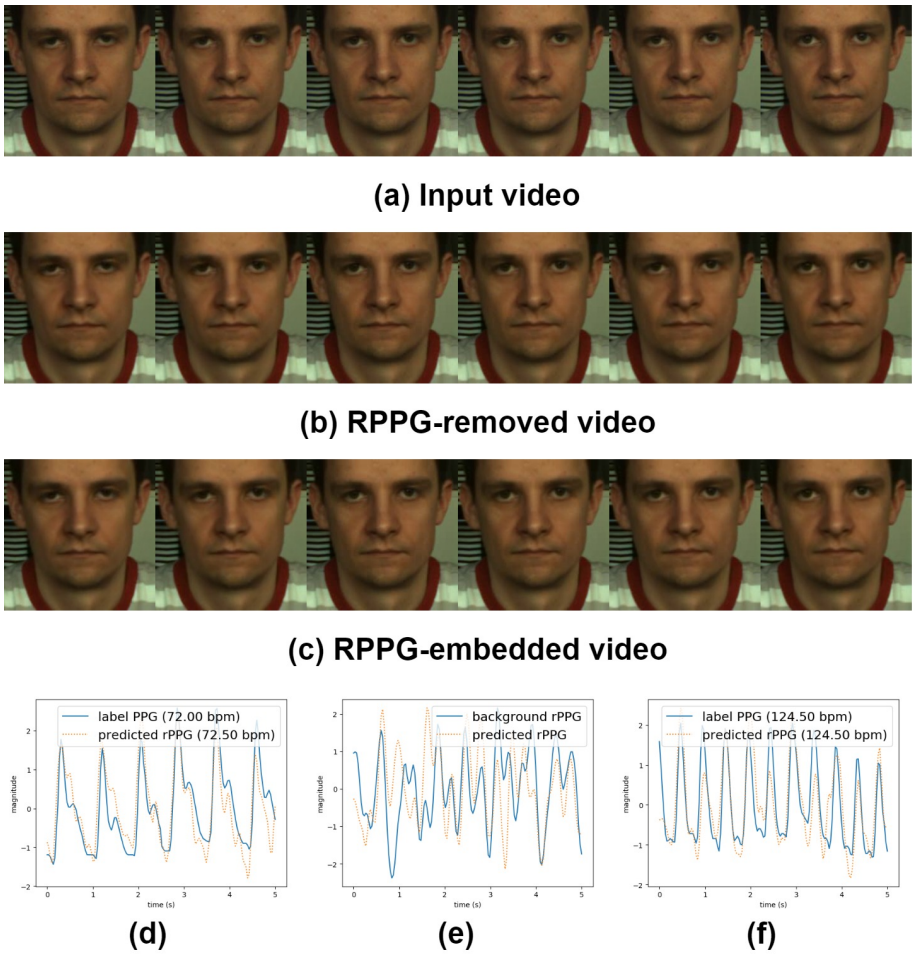

Fig. 4. (a) The input video x_i of subject 1 from PURE dataset; (b) The rPPG-removed video x_r; (c) The rPPG-embedded video x_t; (d) The ground truth PPG signal s_i (blue) and the predicted rPPG signal of x_i (orange); (e) The background signal s_{bg} (blue) and the predicted signal of x_r (orange); and (f) The specified PPG signal s_t (blue) and the predicted rPPG signal of x_t (orange).