Supplementary Material:
Non-Contrastive Unsupervised Learning of Physiological Signals from Video

Jeremy Speth, Nathan Vance, Patrick Flynn, Adam Czajka
University of Notre Dame
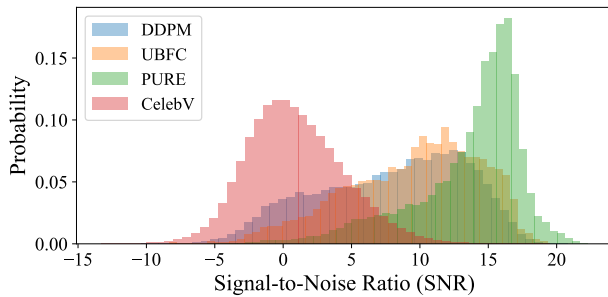{jspeth,nvance1,flynn,aczajka}@nd.edu

Figure 1. Distribution of SNR values from the POS [7] algorithm over the three main rPPG datasets and CelebV [8]. Although the downloaded videos are of the highest quality available, there is a clear downwards shift in signal quality due to compression and diverse settings.

1. rPPG Signal Quality on CelebV-HQ

Given the enormous quantity of video data present in CelebV-HQ, we were surprised to find that SiNC was unable to correctly learn the pulse signal. To better understand why training on CelebV-HQ is challenging, we predicted rPPG signals with the POS [7] algorithm, which is a reputable baseline approach that tends to transfer well across different sources of video data. We calculated the signal-to-noise ratio (SNR) from all of the predictions and compared them with predictions on traditional rPPG datasets (DDPM [4, 6], UBFC [1], and PURE [5]). Figure 2 shows the histograms of SNRs for each dataset. The SNR for CelebV-HQ is much lower than the other datasets, indicating a lower signal quality. The drop in quality is likely due to video compression, which may have even occurred multiple times before download.

2. Justification for Frequency Bounds

To justify our selection of the lower and upper bounds of 40 bpm and 180 bpm, we plotted the distribution of ground truth pulse rates over DDPM, UBFC, and PURE. In general we see that very few pulse rates approach 40, and the
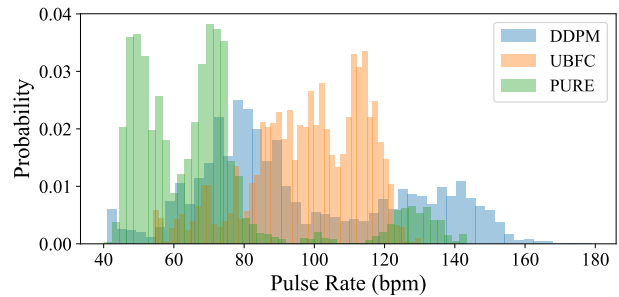

Figure 2. Distribution of ground truth pulse rates over the three main rPPG datasets explored in this paper.

highest pulse rates are just beyond 160 bpm.

3. Allowing Second Harmonic in Sparsity Loss

Many rPPG papers allow signal power in the second harmonic when evaluating their approaches [2, 3]. We chose not to incorporate higher harmonics to keep the sparsity loss simple and avoid the risk of amplifying the dicrotic notch of the waveform. We verified this empirically by training and testing models on PURE while allowing energy in the second harmonic. The performance dropped due to the peak frequency occasionally occurring in the second harmonic (MAE of 3.29 ± 1.69). For the purpose of pulse rate estimation, including the dicrotic notch can actually introduce inaccuracies.

4. Impact of Batch Size

The variance component of the loss depends on the batch size, since a normalized sum over the batch is calculated. To verify that the batch size can safely be reduced for memory-constrained environments we performed an ablation study with batch sizes in $\{5, 10, 15, 20\}$ when training on the PURE dataset. For within-dataset testing, models gave MAEs of $\{0.73 \pm 0.08, 0.65 \pm 0.09, 1.75 \pm 1.39, 0.61 \pm 0.06\}$, respectively. Overall, the batch size does not seem to have a large influence on performance.

5. Augmentations are Critical

Several augmentations are used while training SiNC, some of which can even influence the underlying pulse rate distributions (see frequency augmentations in Sec. 3.2). To verify that the augmentations play a key role, we trained models without them on PURE. Models trained without augmentations gave a MAE of 33.86 ± 0.83. Therefore, training with augmentations is critical within the SiNC approach for models to converge to the blood volume pulse.

References

[1] 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.

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

[3] Ewa M. Nowara, Daniel McDuff, and Ashok Veeraraghavan. Systematic analysis of video-based pulse measurement from compressed videos. *Biomedical Optics Express*, 12(1):494, 2021.

[4] Jeremy Speth, Nathan Vance, Adam Czajka, Kevin Bowyer, Diane Wright, and Patrick Flynn. Deception detection and remote physiological monitoring: A dataset and baseline experimental results. In *International Joint Conference on Biometrics (IJCB)*, pages 4264–4271, 2021.

[5] Ronny Stricker, Steffen Muller, and Horst Michael Gross. Non-contact video-based pulse rate measurement on a mobile service robot. *IEEE International Symposium on Robot and Human Interactive Communication*, pages 1056–1062, 2014.

[6] Nathan Vance, Jeremy Speth, Siamul Khan, Adam Czajka, Kevin W. Bowyer, Diane Wright, and Patrick Flynn. Deception detection and remote physiological monitoring: A dataset and baseline experimental results. *IEEE Transactions on Biometrics, Behavior, and Identity Science (TBIOM)*, pages 1–1, 2022.

[7] W. Wang, A. C. den Brinker, S. Stuijk, and G. de Haan. Algorithmic principles of remote PPG. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2017.

[8] Hao Zhu, Wayne Wu, Wentao Zhu, Liming Jiang, Siwei Tang, Li Zhang, Ziwei Liu, and Chen Change Loy. CelebV-HQ: A large-scale video facial attributes dataset. In *ECCV*, 2022.