1. Face ROI Extraction

The landmark points corresponding to different face regions are depicted in Figure 1 (a) in blue and yellow colors. The landmark points in yellow are used for defining the facial region with the significant rPPG information. The extracted face region is shown in (a) with blue color and the landmarks used for obtaining the chrominance signals, the filtered RGB signals, RGB temporal signals into two orthogonal vectors [2]. For obtaining the chrominance signals, the filtered RGB signals, RGB temporal signals are projected into orthogonal vectors

\[ \mathbf{\eta}_j = 3 \cdot \bar{r}_j - 2 \cdot \bar{g}_j \]

\[ \mathbf{\mu}_j = 1.5 \cdot \bar{r}_j + \bar{g}_j - 1.5 \cdot \bar{b}_j \]

where \( \bar{r}_j, \bar{g}_j, \) and \( \bar{b}_j \) are the chrominance signal, \( \mathbf{\eta}_j \) and \( \mathbf{\mu}_j \) obtained by:

\[ c_j = \tilde{\mathbf{\eta}}_j - \alpha \cdot \tilde{\mathbf{\mu}}_j, \text{where } \alpha = \frac{\sigma(\tilde{\mathbf{\eta}}_j)}{\sigma(\tilde{\mathbf{\mu}}_j)} \]

where, \( \sigma(\cdot) \) represents the standard deviation operator.

3. Synthetic Temporal Signals

The synthetic temporal signals used for pre-training our architecture are generated using sine waves and noise functions as described in [3]. To mimic the systolic and diastolic peaks in the synthetic temporal signals, we have used two waves with the same time period, corresponding to the pulse, with one of the waves having twice the amplitude of the other wave. Further, another wave, having the time period corresponding to the respiratory signal, is used. For adding the effect of noise, we have used a step function and the Gaussian noise function. The synthetic temporal signal \( s_{syn} \) is given by:

\[ s_{syn} = \kappa_1 \cdot \sin(\omega_1 t + \phi) + \kappa_2 \cdot \sin(\omega_2 t + \theta) + 0.5 \cdot \kappa_1 \cdot \sin(2\omega_1 t + \phi) + N(t) \]

\[ p_1 \cdot \text{step}(t - t_1) + p_2 \cdot \text{step}(t - t_2) \]

where, \( \kappa_1 \) and \( \kappa_2 \) are amplitudes of the sine waves sampled randomly from \([0,1]\); \( \omega_1 \) and \( \omega_2 \) are the HR and respiratory frequencies, respectively. Random phase \( \phi \) and \( \theta \) are sampled randomly from \([0, \pi]\). Also, \( \text{step}(t) \) denotes the step function used to add noise and the values \( t_1 \) and \( t_2 \) are chosen between \([0, T]\) randomly, where \( T \) is the video clip length. Furthermore, \( N \) represents the Gaussian noise. The values \( p_1 \) and \( p_2 \) are derived from Bernoulli distribution.

4. Effect of size of the ROIs used for rPPG estimation

The effect of different block sizes considered for ROI division on the performance of our method is provided in Table 1. The experiments are performed by changing the...
number of non-overlapping blocks in the horizontal direction. Hence, when we increase the number of blocks, the size of the patch decreases. Likewise, on decreasing the number of blocks, the patch size increases. Initially, when we decrease the size of the patch, the performance of our method improves. However, the performance saturates for an optimum size of the patches, after which a decrease in the performance is observed. As we go on decreasing the patch size, the total number of temporal signals obtained increases, hence, providing a better combination of rPPG information. However, the smaller patch sizes are susceptible to noise, which affects the quality of rPPG information, leading to performance degradation. RADIANT obtains optimal performance when 10 horizontal blocks are considered.

5. Qualitative Results

The figures 2 (a) and 2 (b) depict examples for HR estimation by our proposed method. The face videos and temporal signals are presented in the first row; the second row shows the estimated pulse signal and its Fourier Power Spectrum, and the third row presents the ground truth signal and its Power Spectrum. In Figure 2 (a), an example of successful HR estimation is provided with the estimated and ground truth pulse rate to be 98 BPM. The quality score for the temporal signals is 5.19 for this sample. Further, a higher correlation is observed in the estimated and ground truth pulse signal. Similarly, the Fourier Power Spectrum of the estimated pulse and the ground truth pulse signal show single peaks denoting less amount of noise in the estimated pulse signal. An example of unsuccessful HR estimation can be observed in Figure 2 (b) where the estimated HR is 63 BPM, and the ground truth is 92 BPM. The quality score of the temporal signals is 3.80. Further, there is little correlation between the estimated pulse signal and the ground truth pulse signal. Likewise, the Fourier Power Spectrum of the estimated pulse signal shows multiple peaks indicating a higher noise content resulting in incorrect HR estimation.

References

[1] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois. Unsupervised skin tissue seg-
