Supplementary Material of DD-rPPGNet : De-interfering and Descriptive Feature Learning for Unsupervised rPPG Estimation

In this supplementary, in Section I, we first present the detailed derivations elucidating the nature of the cross-correlation between rPPG and interference signals. Next, in Section II, we show that Temporal Difference Convolution (TDC) [1] is a special case of the proposed 3D Learnable Descriptive Convolution (3DLDC). Finally, in Section III, we provide the detailed description of the proposed protocols.

I. Derivation: the nature of the cross-correlation between rPPG and interference signals

We perform a running correlation between the interference-carrying rPPG signal $\hat{r}$ and the estimated interference signal n^{bg} to calculate their correlation $c_{\hat{r},n}$ by,

$$\begin{aligned} c_{\hat{r},n}[\tau] &= NC(\hat{r}[v-\tau], n^{bg}[v]), \\ &= NC(\hat{r}[\bar{v}], n^{bg}[v]), \end{aligned} \tag{1}$$

where

$$NC(a[v], b[v]) = \frac{\sum_l^L a[l] \cdot b[l]}{\sqrt{\sum_l^L (a[l])^2}\sqrt{\sum_l^L (b[l])^2}} \tag{2}$$

is the operation of normalized correlation between the signals $a[v]$ and $b[v]$, L is the length of signal, $a[l]$ is the l-th sample of the signal $a[v]$, $c_{\hat{r},n}[\tau]$ is the τ-th sample of the correlation signal $c_{\hat{r},n}$, $\hat{r}[\bar{v}] = \hat{r}[v-\tau]$ is the signal $\hat{r}[v]$ shifted by τ samples, $\tau = -(L-1), \cdots, 0, \cdots, L-1$ is the lag. Note that, when $\tau > 0$, $\hat{r}[v-\tau]$ is equal to the signal $\hat{r}[v]$ shifted u samples to the right, while when $\tau < 0$, $\hat{r}[v-\tau]$ is equal to the signal $\hat{r}[v]$ shifted τ samples to the left. In addition, we use zero-padding to keep the same length between $\hat{r}[\bar{v}]$ and $n^{bg}[v]$.

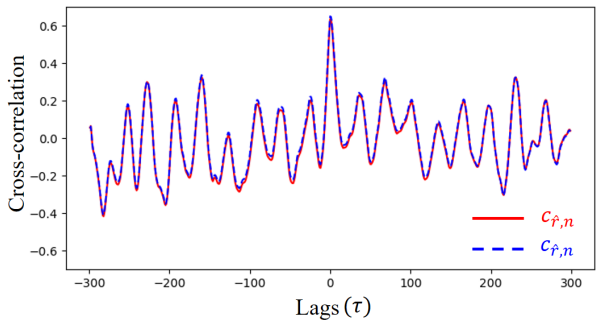

Fig. 1. The correlation signals produced by using Equation (1) (red solid line) and Equation (3) (blue dotted line) are highly similar.

Considering the interference-carrying rPPG signals $\hat{r}$, which consist of the genuine rPPG signal r and the interference signals n^{fg}, we submit $\hat{r} = r + n^{fg}$ into (1) to rewrite Equation (1) as,

$$\begin{aligned} c_{\hat{r},n}[u] &= NC(r[v-\tau] + n^{fg}[v-\tau], n^{bg}[v]) \\ &= NC(r[\bar{v}] + n^{fg}[\bar{v}], n^{bg}[v]) \\ &= \frac{\sum_l^L (r[l] + n^{fg}[l]) \cdot (n^{bg}[l])}{\sqrt{\sum_l^L (r[l] + n^{fg}[l])^2}\sqrt{\sum_l^L (n^{bg}[l])^2}} \\ &= \frac{\sum_l^L r[l] \cdot n^{bg}[l]}{\sqrt{\sum_l^L (r[l] + n^{fg}[l])^2}\sqrt{\sum_l^L (n^{bg}[l])^2}} \\ &\quad + \frac{\sum_l^L n^{fg}[l] \cdot n^{bg}[l]}{\sqrt{\sum_l^L (r[l] + n^{fg}[l])^2}\sqrt{\sum_l^L (n^{bg}[l])^2}} \\ &= \frac{\sqrt{\sum_l^L (r[l])^2}}{\sqrt{\sum_l^L (r[l] + n^{fg}[l])^2}} \cdot \frac{\sum_l^L r[l] \cdot n^{bg}[l]}{\sqrt{\sum_l^L (r[l])^2}\sqrt{\sum_l^L (n^{bg}[l])^2}} \\ &\quad + \frac{\sqrt{\sum_l^L (n^{fg}[l])^2}}{\sqrt{\sum_l^L (r[l] + n^{fg}[l])^2}} \cdot \frac{\sum_l^L n^{fg}[l] \cdot n^{bg}[l]}{\sqrt{\sum_l^L (r[l])^2}\sqrt{\sum_l^L (n^{bg}[l])^2}} \\ &= \frac{\sqrt{\sum_l^L (r[l])^2}}{\sqrt{\sum_l^L (r[l] + n^{fg}[l])^2}} \cdot NC(r[\bar{v}], n^{bg}[v]) \\ &\quad + \frac{\sqrt{\sum_l^L (n^{fg}[l])^2}}{\sqrt{\sum_l^L (r[l] + n^{fg}[l])^2}} \cdot NC(n^{fg}[\bar{v}], n^{bg}[v]) \\ &= \alpha NC(r[\bar{v}], n^{bg}[v]) + \beta NC(n^{fg}[\bar{v}], n^{bg}[v]), \end{aligned} \tag{3}$$

where $\alpha = \frac{\sqrt{\sum_l^L (r[l])^2}}{\sqrt{\sum_l^L (r[l]+n^{fg}[l])^2}}$ and $\beta = \frac{\sqrt{\sum_l^L (n^{fg}[l])^2}}{\sqrt{\sum_l^L (r[l]+n^{fg}[l])^2}}$, $r[l]$ is the l-th sample of the signal $r[\bar{v}]$, $n^{fg}[l]$ is the l-th sample of the signal $n^{fg}[\bar{v}]$.

To evaluate the correctness of Equation (3), we compare the correlation signals produced by Equation (1) and Equation (3). In particular, we consider using the ground truth rPPG signal as r and adopting the off-the-shelf rPPG estimator [2] to extract interference signals from the two different non-facial regions, denoted as n^{fg} and n^{bg}. Next, we propose to combine r and n^{fg} to simulate the interference rPPG signal $\hat{r} = r + n^{fg}$. Finally, we Equation (1) and Equation (3) to calculate their correlations, shown in Figure 1. We see that the correlations produced by using Equation (1) and Equation (3) yield nearly identical results.

II. Derivation: TDC as a Special Case of 3DLDC

We first [1] to formulate Temporal Difference Convolution (TDC) within local temporal regions $\mathcal{R}^{t-1}$, $\mathcal{R}^t$, and $\mathcal{R}^{t+1}$ on the feature level by,

$$\begin{aligned} g(p^t) = &\sum_{p_k^{t-1} \in \mathcal{R}^{t-1}} w(p_k^{t-1}) \cdot (f(p^{t-1} + p_k^{t-1}) - \theta \cdot f(p^t)) \\ &+ \sum_{p_k^{t+1} \in \mathcal{R}^{t+1}} w(p_k^{t+1}) \cdot (f(p^{t+1} + p_k^{t+1}) - \theta \cdot f(p^t)) \\ &+ \sum_{p_k^{t} \in \mathcal{R}^{t}} w(p_k^{t}) \cdot f(p^{t} + p_k^{t}), \end{aligned} \tag{4}$$

where w is the convolution kernel, f is the input feature map, p^t is the pixel of current location in the t-th frame, p_k^t is the location of neighboring pixels in $\mathcal{R}^t = \{(-1,-1),(-1,0),...,(0,1),(1,1)\}$, g is the output feature map, and θ is a hyperparemter for controling the contribution of temporal difference.

Next, we propose the 3D Learnable Descriptive Convolution (3DLDC) to adaptively focus on blood volume changes in optical information for learning distinctive rPPG features by,

$$\begin{aligned} g(p^t) = (1-\epsilon) &\underbrace{\sum_{i} \sum_{p_k^{t+i} \in \mathcal{R}^{t+i}} w(p_k^{t+i}) \cdot f(p^{t+i} + p_k^{t+i})}_{\text{vanilla convolution}} \\ + \epsilon &\underbrace{\sum_{i} \sum_{p_k^{t+i} \in \mathcal{R}^{t+i}} w(p_k^{t+i}) \cdot (f(p^{t+i} + p_k^{t+i}) \cdot m(p_k^{t+i}))}_{\text{3D learnable descriptive convolution}} \end{aligned} \tag{5}$$

where m is the learnable descriptor, w is the convolution kernel, and ϵ is a hyperparameter for controling the contribution of 3D LDC. Note that, m and w are both of the same size $3 \times 3 \times 3$.

We show that Temporal Difference Convolution (TDC) [1] is a special case of 3DLDC when the matrix m in Equation (5) is

$$m = \mathbf{1}_{3\times3\times3} + \left[m_0, \begin{bmatrix} 0 & 0 & 0 \\ 0 & w_s & 0 \\ 0 & 0 & 0 \end{bmatrix}, m_0 \right], \tag{6}$$

where

$$w_s = -\frac{1}{w(p^t)} \Big(\sum_{p_k^{t-1} \in \mathcal{R}^{t-1}} w(p_k^{t-1}) + \sum_{p_k^{t+1} \in \mathcal{R}^{t+1}} w(p_k^{t+1}) \Big) \tag{7}$$

is the weight for special case of 3DLDC, m_0 is all-zero matrix $\mathbf{0}_{3\times3}$, and the base matrix $\mathbf{1}_{3\times3\times3}$ is an all-ones matrix.

By merging vanilla convolution and 3DLDC, we rewrite Equation (5) as,

$$g(p^t) = \sum_{i=-1}^{1} \sum_{p_k^{t+i} \in \mathcal{R}^{t+i}} \hat{w}(p_k^{t+i}) \cdot f(p^{t+i} + p_k^{t+i}), \tag{8}$$

where

$$\begin{aligned} \hat{w} &= \left[\begin{bmatrix} (1-\epsilon)w_{\text{-}1,\text{-}1}^{t\text{-}1} + \epsilon\, w_{\text{-}1,\text{-}1}^{t\text{-}1} m_{\text{-}1,\text{-}1}^{t\text{-}1} & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & (1-\epsilon)w_{1,1}^{t\text{-}1} + \epsilon\, w_{1,1}^{t\text{-}1} m_{1,1}^{t\text{-}1} \end{bmatrix}, \right. \\ &\quad \begin{bmatrix} (1-\epsilon)w_{\text{-}1,\text{-}1}^{t} + \epsilon\, w_{\text{-}1,\text{-}1}^{t} m_{\text{-}1,\text{-}1}^{t} & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & (1-\epsilon)w_{1,1}^{t} + \epsilon\, w_{1,1}^{t} m_{1,1}^{t} \end{bmatrix}, \\ &\quad \left. \begin{bmatrix} (1-\epsilon)w_{\text{-}1,\text{-}1}^{t+1} + \epsilon\, w_{\text{-}1,\text{-}1}^{t+1} m_{\text{-}1,\text{-}1}^{t+1} & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & (1-\epsilon)w_{1,1}^{t+1} + \epsilon\, w_{1,1}^{t+1} m_{1,1}^{t+1} \end{bmatrix} \right] \\ &= \left[\begin{bmatrix} w_{\text{-}1,\text{-}1}^{t\text{-}1}(1-\epsilon+\epsilon\, m_{\text{-}1,\text{-}1}^{t\text{-}1}) & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & w_{1,1}^{t\text{-}1}(1-\epsilon+\epsilon\, m_{1,1}^{t\text{-}1}) \end{bmatrix}, \right. \\ &\quad \begin{bmatrix} w_{\text{-}1,\text{-}1}^{t}(1-\epsilon+\epsilon\, m_{\text{-}1,\text{-}1}^{t}) & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & w_{1,1}^{t}(1-\epsilon+\epsilon\, m_{1,1}^{t}) \end{bmatrix}, \\ &\quad \left. \begin{bmatrix} w_{\text{-}1,\text{-}1}^{t+1}(1-\epsilon+\epsilon\, m_{\text{-}1,\text{-}1}^{t+1}) & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & w_{1,1}^{t+1}(1-\epsilon+\epsilon\, m_{1,1}^{t+1}) \end{bmatrix} \right]. \end{aligned} \tag{9}$$

where $w_{p_k}^{t+i} = w(p_k^{t+i}), p_k \in \{(-1,-1),...,(1,1)\}$. By substituting m in Equation (6) into Equation (9), we rewrite Equation (9) as,

$$\begin{aligned} \hat{w} = &\left[\begin{bmatrix} w_{\text{-}1,\text{-}1}^{t\text{-}1}(1-\epsilon+\epsilon) & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & w_{1,1}^{t\text{-}1}(1-\epsilon+\epsilon) \end{bmatrix}, \right. \\ &\quad \begin{bmatrix} w_{\text{-}1,\text{-}1}^{t}(1-\epsilon+\epsilon) & \cdots & \vdots \\ \vdots & w_{0,0}^{t}(1-\epsilon+\epsilon\,(1+w_s)) & \vdots \\ \vdots & \cdots & w_{1,1}^{t}(1-\epsilon+\epsilon) \end{bmatrix}, \\ &\quad \left. \begin{bmatrix} w_{\text{-}1,\text{-}1}^{t+1}(1-\epsilon+\epsilon) & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & w_{1,1}^{t+1}(1-\epsilon+\epsilon) \end{bmatrix} \right] \end{aligned}$$

$$= \begin{bmatrix} \begin{bmatrix} w^{t-1}_{-1,-1} & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & w^{t-1}_{1,1} \end{bmatrix}, \\ \begin{bmatrix} w^{t}_{-1,-1} & \cdots & \vdots \\ \vdots & \boldsymbol{w^t_{0,0}(1+\epsilon\, w_s)} & \vdots \\ \vdots & \cdots & w^{t}_{1,1} \end{bmatrix}, \\ \begin{bmatrix} w^{t+1}_{-1,-1} & \cdots & \vdots \\ \vdots & \ddots & \vdots \\ \vdots & \cdots & w^{t+1}_{1,1} \end{bmatrix} \end{bmatrix} a \quad (10)$$

where

$$\begin{aligned} \boldsymbol{w^t_{0,0}(1+\epsilon\, w_s)} &= w^t_{0,0} + \epsilon\, w^t_{0,0} w_s \\ &= w^t_{0,0} - \epsilon\, (\sum_{p^{t-1}_k \in \mathcal{R}^{t-1}} w(p^{t-1}_k) + \sum_{p^{t+1}_k \in \mathcal{R}^{t+1}} w(p^{t+1}_k)) \end{aligned} \quad (11)$$

By substituting Equation (10) into Equation (8), we rewrite Equation (8) as,

$$\begin{aligned} g(p^t) &= \underbrace{\sum_{i=-1}^{1} \sum_{p^{t+i}_k \in \mathcal{R}^{t+i}} w(p^{t+i}_k) \cdot f(p^{t+i} + p^{t+i}_k)}_{\text{vanilla convolution}} \\ &\quad \underbrace{- \epsilon \cdot (\sum_{p^{t-1}_k \in \mathcal{R}^{t-1}} w(p^{t-1}_k) + \sum_{p^{t+1}_k \in \mathcal{R}^{t+1}} w(p^{k+1}_k)) \cdot f(p^t).}_{\text{TDC}} \\ &= \sum_{p^{t-1}_k \in \mathcal{R}^{t-1}} w(p^{t-1}_k) \cdot (f(p^{t-1} + p^{t-1}_k) - \epsilon \cdot f(p^t)) \\ &\quad + \sum_{p^{t+1}_k \in \mathcal{R}^{t+1}} w(p^{t+1}_k) \cdot (f(p^{t+1} + p^{t+1}_k) - \epsilon \cdot f(p^t)) \\ &\quad + \sum_{p^{t}_k \in \mathcal{R}^{t}} w(p^{t}_k) \cdot f(p^{t} + p^{t}_k), \end{aligned} \quad (12)$$

When $\epsilon = \theta$, 3DLDC apparently becomes TDC in Equation (4).

III. The proposed protocols

In this section, we provide the detailed description of the proposed protocols. In protocol 1, we use the training and testing data involved rapid head movements from the original setup "Fast Translation" in the dataset **PURE** [3]. Next, in protocol 2, we use the training and testing data involved facial expressions from the original setup "Talking" in **PURE**. Moreover, in protocol 3, we the training data from the original setup "Steady" in **PURE**. To simulate compression artifacts, we compress the testing videos from the same setup "Talking" in **PURE** 10 times by adopting a JPEG ratio of 10:1 as the testing data. Furthermore, in protocol 4, we use the training and testing data from the original setup "NATURE" in the dataset **COHFACE** dataset [4]. Finally, in protocol 5, we use the training data from the dataset **UBFC-rPPG** [5], and follow [2] to add the periodic interferences into both the facial region and non-facial regions of testing videos from **UBFC-rPPG** as the testing data.

References

[1] Z. Yu, X. Li, X. Niu, J. Shi, and G. Zhao, "Autohr: A strong end-to-end baseline for remote heart rate measurement with neural searching," *Proc. IEEE Signal Process. Lett.*, vol. 27, pp. 1245–1249, 2020.

[2] Z. Sun and X. Li, "Contrast-phys: Unsupervised video-based remote physiological measurement via spatiotemporal contrast," in *Proc. European Conf. Comput. Vis.*, 2022, pp. 492–510.

[3] R. Stricker, S. Müller, and H. Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," in *Proc. 23rd IEEE Int. Symp. Robot Human Interactive Commun.*, 2014, pp. 1056–1062.

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

[5] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," *Pattern Recog. Lett.*, vol. 124, pp. 82–90, 2019.