

Privacy-Phys: Facial Video-Based Physiological Modification for Privacy Protection

Zhaodong Sun  and Xiaobai Li , *Member, IEEE*

Abstract—The invisible remote photoplethysmography (rPPG) signals in facial videos can reveal the cardiac rhythm and physiological status. Recent studies show that rPPG is a non-contact way for emotion recognition, disease detection, and biometric identification, which means there is a potential privacy problem about physiological information leakage from facial videos. Therefore, it is essential to process facial videos to prevent rPPG extraction in privacy-sensitive situations such as online video meetings. In this letter, we propose Privacy-Phys, a novel method based on a pre-trained 3D convolutional neural network, to modify rPPG in facial videos for privacy protection. Our experimental results show that our approach can modify rPPG signals in facial videos more effectively and efficiently than the previous baseline. Our method can be applied to process facial videos in online video meetings or video-sharing platforms to prevent rPPG from being captured maliciously.

Index Terms—Privacy protection, remote photoplethysmography (rPPG), facial videos.

I. INTRODUCTION

Facial videos are widely captured and shared by ubiquitous cameras in our daily life. Some works [1]–[3] have realized that visible facial appearance in facial videos has privacy issues. However, invisible physiological signals in facial videos also present a potential risk of private information leakage.

Several studies [4]–[13] have shown that it is feasible to extract remote photoplethysmography (rPPG) signals from facial videos to derive physiological parameters such as heart rates. These physiological parameters from rPPG can be used for emotion detection [14]–[16], health monitoring [17], [18], and biometric identification [19]. However, the studies above also raise privacy issues of physiological signals in facial videos. Attackers might measure rPPG from facial videos to obtain emotion, health status, and biometric information without others' permission. Recently, rPPG was used to estimate heart rates of athletes in Olympic shooting games, which incurred criticism of widespread biometrics capture [20]. Therefore, it is essential to develop a method to conceal rPPG signals in facial videos.

Two papers [21], [22] explored to conceal rPPG in facial videos for privacy protection. Chen *et al.* [21] used Laplacian

Manuscript received 1 May 2022; revised 16 June 2022; accepted 20 June 2022. Date of publication 24 June 2022; date of current version 13 July 2022. This work was supported in part by the Academy of Finland under Grants 323287 and 345948, and in part by the Finnish Work Environment Fund under Grant 200414. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Yue Deng. (*Corresponding author: Xiaobai Li.*)

The authors are with the Center for Machine Vision and Signal Analysis, University of Oulu, 90014 Oulu, Finland (e-mail: zhaodong.sun@oulu.fi; xiaobai.li@oulu.fi).

Digital Object Identifier 10.1109/LSP.2022.3185964

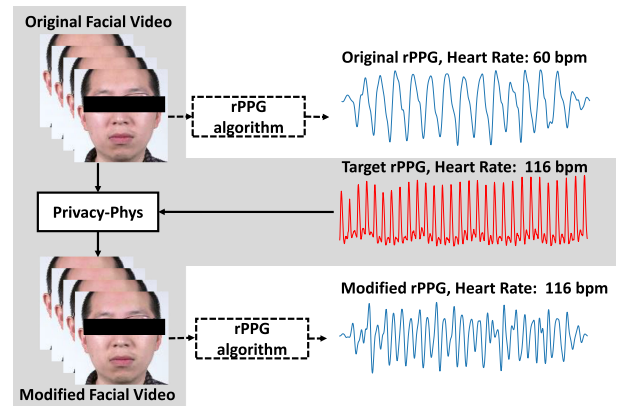


Fig. 1. The diagram of rPPG modification. The gray area is the rPPG modification steps.

pyramid representation to remove the color changes induced by rPPG signals in facial videos. However, their method was designed to eliminate the rPPG, which might produce unrealistic videos as mentioned by [23]. Recently, PulseEdit [22] was proposed to modify rPPG in facial videos. They generated a target rPPG signal and solved an optimization problem with the target and original rPPG signals to get the perturbation. The perturbation was uniformly added to the facial region to embed the target rPPG. However, there are three limitations of PulseEdit. 1) PulseEdit was only validated on simple datasets which are not close to realistic application situations with sudden head motions and facial expressions. 2) The rPPG modification is less effective when deep learning-based rPPG methods try to extract rPPG from the modified videos. 3) The running speed is not high enough to support real-time video processing, especially during online video meetings.

In this letter, we propose Privacy-Phys, an rPPG modification method based on a pre-trained 3D convolutional neural network (3DCNN) [24], and solve the limitations above. The diagram of the rPPG modification is shown in Fig. 1. The original facial video is processed by the rPPG modification algorithm so that the modified facial video will be visibly identical to the original video while the rPPG signal is modified.

Our contributions are listed below. 1) We propose a novel method to modify rPPG in facial videos by leveraging a pre-trained 3DCNN model. 2) We demonstrate that our proposed method can effectively modify rPPG especially against deep learning-based rPPG measurement methods. 3) Our method runs significantly faster than PulseEdit and can be used for real-time video processing. The remainder of this letter is structured as follows. In Section II, we introduce how Privacy-Phys modifies

III. EXPERIMENTS

A. Experiment Setup

Data: PURE [30] and MMSE-HR [31] datasets are used for experiments. PURE has 60 facial videos recorded from ten subjects with 640×480 resolution and 30 frames per second (fps) in six different setups, including steady and simple motion tasks. MMSE-HR has 102 facial videos with 1040×1392 resolution and 25 fps from 40 subjects. The videos contain spontaneous facial expressions and sudden head motions, which are close to realistic situations. MMSE-HR is more challenging than PURE dataset due to sudden facial motions. The original videos are cropped by OpenFace [32] to get full face bounding boxes, and bounding boxes are resized to 128×128 . In order to explore conditions when only partial faces are available (e.g., due to occlusion), we also spatially crop the full face bounding boxes into four equal quarters of upper left, upper right, bottom left, and bottom right.

Experimental Protocol: The pre-trained 3DCNN model $G_{\theta^*}(x)$ was obtained from [8], which is the precondition to perform our rPPG modification method. We then perform rPPG modification on PURE and MMSE-HR datasets, respectively. These videos are divided into 15-sec video clips, and we use the proposed method to embed a 15-sec target PPG signal into these video clips. Similar to PulseEdit [22], we generate a sinusoid of frequency 120 beats per minute (bpm) as the target rPPG signal. For the parameters in the proposed method, we try different values of λ and report the results in the following. For each video clip, we perform the projected gradient descent with Adam optimizer [29] for 20 iterations which are enough to converge. The learning rate of the Adam optimizer is 0.1, and other parameters in the optimizer are set to the default values as in [29]. Skin masks S are obtained from the original videos by pixel thresholding used in [33].

rPPG Measurement Methods: The effect of rPPG modification needs to be evaluated against rPPG measurement algorithms. After getting the modified videos, we use several rPPG measurement algorithms to get the heart rates from the modified videos. We choose three traditional rPPG algorithms including **POS** [7], **CHROM** [5], **ICA** [4], and four latest deep learning-based rPPG algorithms including **PhysNet** [8] trained on OBF dataset [34], **TS-CAN** [9] trained on AFRL dataset [35], **Gideon2021** [10] trained on UBFC-rPPG dataset [36], and **PhysFormer** [11] trained on VIPL dataset [26]. Deep learning-based rPPG methods are trained according to the steps in their papers.

Evaluation Metrics: We use two metrics, including peak signal to noise ratio (PSNR) [37] and mean absolute errors (MAE) of heart rates. PSNR is used to evaluate the video similarity between modified videos and original videos. A higher PSNR means better video fidelity. To evaluate if the target PPG signal is successfully embedded into the modified video clips, we calculate the mean absolute error (MAE) between the heart rates in the modified face videos and the heart rates in the target PPG signal. The MAE values should be small if the target PPG is successfully embedded into the new video clips.

B. Experimental Results

The influence of λ : λ is a parameter to balance the fidelity term L_f and rPPG term L_{rppg} . Changing λ will cause the change of video quality and rPPG modification success. Fig. 3 shows

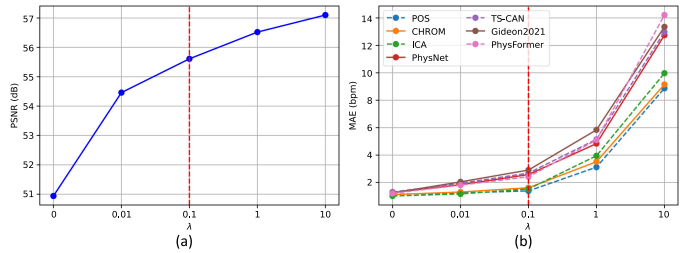


Fig. 3. (a) PSNR with respect to λ . (b) MAE with respect to λ .

TABLE I
MAE AND PSNR FOR OUR METHOD WITH $\lambda = 0.1$ AND THE BASELINE METHOD PULSEEDIT [22] WITH THEIR DEFAULT SETTINGS ON PURE DATASET AND MMSE-HR DATASET WITH FULL AND PARTIAL FACES

		PURE Dataset		MMSE-HR Dataset		
		Ours	PulseEdit [22]	Ours	PulseEdit [22]	
Full Faces	MAE (bpm)	POS [7]	0.03	0.03	1.39	3.71
		CHROM [5]	0.02	0.03	1.62	3.85
		ICA [4]	0.03	0.03	1.53	3.81
		PhysNet [8]	0.35	1.17	2.58	5.80
		TS-CAN [9]	0.20	1.79	2.69	5.57
		Gideon2021 [10]	0.51	1.25	2.91	5.37
		PhysFormer [11]	0.31	1.13	2.42	5.30
	PSNR (dB)	58.38	58.14	55.61	54.76	
Partial Faces	MAE (bpm)	POS [7]	0.15	1.28	2.17	4.52
		CHROM [5]	0.20	1.30	2.35	4.79
		ICA [4]	0.18	1.33	2.71	5.44
		PhysNet [8]	1.35	2.97	3.24	7.46
		TS-CAN [9]	1.86	3.54	3.81	6.19
		Gideon2021 [10]	1.52	3.37	3.67	7.30
		PhysFormer [11]	2.35	3.89	3.72	7.51
	PSNR (dB)	57.41	57.17	54.85	54.12	

PSNR/MAE with respect to λ on MMSE-HR dataset with full faces. A larger λ will result in a higher PSNR and a larger MAE. From the security aspect, a smaller λ will have higher rPPG modification success and lower PSNR, which can make stronger privacy protection under the premise of lower video quality. We want to find a λ where PSNR is high while MAE is low. The results show that, when λ is larger than 0.1, MAE will increase sharply, and PSNR will not increase too much. This means $\lambda = 0.1$ is a suitable level where video quality is good, and most rPPG signals are successfully modified. Therefore, we choose $\lambda = 0.1$ to balance video quality and modification success.

Comparison with PulseEdit: Table I shows the MAE and PSNR for the proposed method with $\lambda = 0.1$ and PulseEdit with their default parameter on both PURE and MMSE-HR datasets with full faces or partial faces. The results show that the proposed method has a lower MAE, which indicates that the modification of the proposed method is more effective than that of PulseEdit [22]. The MAE values of PulseEdit on deep learning-based rPPG methods are much higher than traditional rPPG methods, while the proposed method can keep low MAE on both deep learning-based and traditional rPPG methods. In addition, Table I shows the proposed method has low MAE in challenging MMSE-HR dataset with sudden head motions and facial expressions. The results on the MMSE-HR dataset mean that the proposed method is more suitable for realistic situations than PulseEdit. Besides, our method also achieves higher PSNR than PulseEdit. The average results on partial faces with four

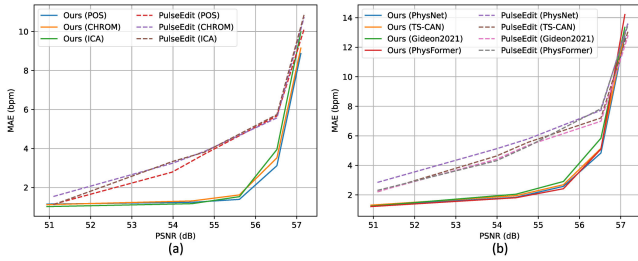


Fig. 4. MAE with respect to PSNR. (a) comparison between our method and PulseEdit on traditional rPPG methods. (b) comparison between our method and PulseEdit on deep learning-based rPPG methods.

quarters also demonstrate that our method is still effective when only partial faces are available.

PulseEdit also has a parameter similar to λ to trade off video quality and modification success. We can change this parameter to get the curves of MAE with respect to PSNR. Fig. 4 plots the MAE with respect to PSNR to compare our method and PulseEdit on MMSE-HR dataset with full faces. At the same PSNR, our method can achieve a lower MAE than PulseEdit, which means our method can more effectively modify rPPG than PulseEdit. Our method also surpasses PulseEdit much more on deep learning-based rPPG methods than on traditional rPPG methods.

Running Speed: Our method and PulseEdit are tested on a Linux workstation with an Intel Xeon 6230 CPU, 16 GB memory, and a single Nvidia V100 GPU. The speed of PulseEdit on GPU is 13 fps. Our method runs at 42 fps on GPU, which is much faster than PulseEdit since our method uses a deep neural network that GPU can accelerate. However, PulseEdit does not employ parallel computation models, so GPU does not give PulseEdit a significant boost. The results show that our proposed method can be potentially used for real-time video processing during online video meetings.

C. Visual Validation

Fig. 5 shows a case where our method can successfully modify rPPG while PulseEdit fails. This case is challenging since the person has apparent facial expressions. We use PhysNet to measure rPPG signals from the original and modified videos. The rPPG signal from our modified video looks cleaner than that from PulseEdit's one. Power spectrum densities (PSDs) also indicate our method can successfully suppress the original rPPG with 88 bpm and add a new rPPG with 120 bpm. However, PulseEdit can only add the 120 bpm and fail to suppress the original rPPG signals with 88 bpm in this challenging case. The results indicate that it is still possible for attackers to obtain the original heart rates from the videos processed by PulseEdit when facial expressions occur.

Fig. 6 shows the RGB absolute difference between the original and modified frame for our method and PulseEdit. Our method has large perturbation in the cheek and forehead area, which follows the rPPG spatial distribution [38]. Our method also has the largest perturbation in the green channel, followed by blue and red channels, which follows the rPPG color channel distribution as illustrated in [7]. These observations demonstrate the correctness of our method. However, PulseEdit applies the perturbation uniformly to the face in spatial and color dimensions. Zhan *et al.* [39] have demonstrated that deep learning-based rPPG

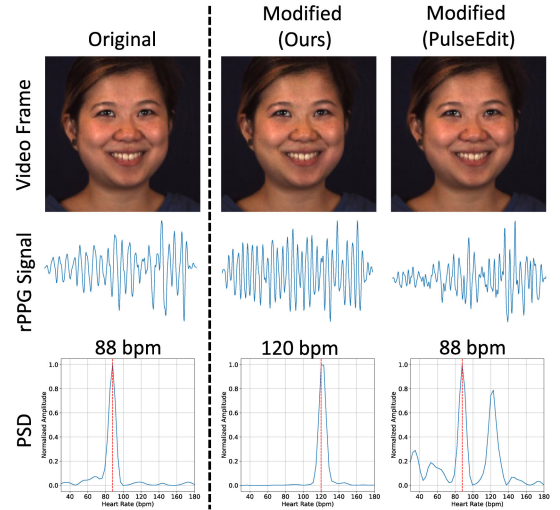


Fig. 5. Visualize the rPPG and PSD curves of one example video before and after modification. Our method successfully altered the rPPG signal and changed the HR from 88 to 120, while PulseEdit failed to, i.e., the original component of 88 is still detectable.

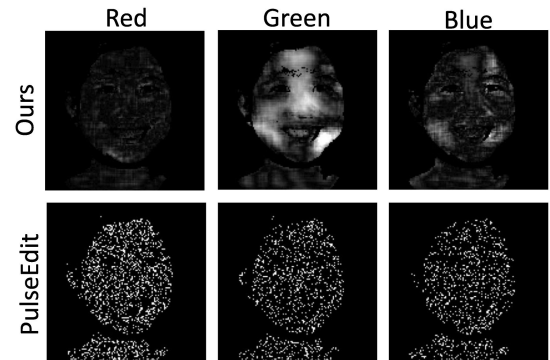


Fig. 6. The RGB absolute difference between the original video frame and the modified video frame (in Fig. 5) for our method and PulseEdit.

methods highly rely on the rPPG spatial and color distribution to extract rPPG signals, which means our perturbation is more compatible with deep learning-based rPPG methods.

IV. CONCLUSION

We propose Privacy-Phys to modify rPPG signals in facial videos by leveraging a pre-trained 3DCNN model. Experimental results demonstrate that the proposed method can modify rPPG more effectively and faster than the baseline method (PulseEdit). Privacy-Phys can be applied to process facial videos in privacy-sensitive situations such as online video meetings to prevent rPPG from being captured maliciously. Future works can focus on concealing other physiological signals such as respiration-induced motions in facial videos for privacy protection.

ACKNOWLEDGMENT

The authors would like to thanks CSC-IT Center for Science, Finland, for providing computational resources.

REFERENCES

- [1] J. Chen, J. Konrad, and P. Ishwar, "VGAN-based image representation learning for privacy-preserving facial expression recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2018, pp. 1570–1579.
- [2] V. Mirjalili, S. Raschka, and A. Ross, "PrivacyNet: Semi-adversarial networks for multi-attribute face privacy," *IEEE Trans. Image Process.*, vol. 29, pp. 9400–9412, 2020.
- [3] B. Meden *et al.*, "Privacy-enhancing face biometrics: A comprehensive survey," *IEEE Trans. Inf. Forensics Secur.*, vol. 16, pp. 4147–4183, 2021.
- [4] M.-Z. Poh, D. J. McDuff, and R. W. Picard, "Advancements in noncontact, multiparameter physiological measurements using a webcam," *IEEE Trans. Biomed. Eng.*, vol. 58, no. 1, pp. 7–11, Jan. 2011.
- [5] G. De Haan and V. Jeanne, "Robust pulse rate from chrominance-based rPPG," *IEEE Trans. Biomed. Eng.*, vol. 60, no. 10, pp. 2878–2886, Oct. 2013.
- [6] X. Li, J. Chen, G. Zhao, and M. Pietikainen, "Remote heart rate measurement from face videos under realistic situations," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2014, pp. 4264–4271.
- [7] W. Wang, A. C. den Brinker, S. Stuijk, and G. De Haan, "Algorithmic principles of remote PPG," *IEEE Trans. Biomed. Eng.*, vol. 64, no. 7, pp. 1479–1491, Jul. 2017.
- [8] Z. Yu, X. Li, and G. Zhao, "Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks," in *Proc. 30th Brit. Mach. Vis. Conf.*, 2019, pp. 29.1–29.12.
- [9] X. Liu, J. Fromm, S. Patel, and D. McDuff, "Multi-task temporal shift attention networks for on-device contactless vitals measurement," *Adv. Neural Inf. Process. Syst.*, vol. 33, pp. 19400–19411, 2020.
- [10] J. Gideon and S. Stent, "The way to my heart is through contrastive learning: Remote photoplethysmography from unlabelled video," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, 2021, pp. 3995–4004.
- [11] Z. Yu, Y. Shen, J. Shi, H. Zhao, P. H. Torr, and G. Zhao, "Physformer: Facial video-based physiological measurement with temporal difference transformer," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 4186–4196.
- [12] 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," *IEEE Signal Process. Lett.*, vol. 27, pp. 1245–1249, 2020.
- [13] S. B. Park, G. Kim, H. J. Baek, J. H. Han, and J. H. Kim, "Remote pulse rate measurement from near-infrared videos," *IEEE Signal Process. Lett.*, vol. 25, no. 8, pp. 1271–1275, Aug. 2018.
- [14] D. McDuff, S. Gontarek, and R. Picard, "Remote measurement of cognitive stress via heart rate variability," in *Proc. IEEE 36th Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2014, pp. 2957–2960.
- [15] Z. Yu, X. Li, and G. Zhao, "Facial-video-based physiological signal measurement: Recent advances and affective applications," *IEEE Signal Process. Mag.*, vol. 38, no. 6, pp. 50–58, Nov. 2021.
- [16] R. Meziati Sabour, Y. Benezeth, P. De Oliveira, J. Chappe, and F. Yang, "UBFC-Phys: A multimodal database for psychophysiological studies of social stress," *IEEE Trans. Affect. Comput.*, early access, Feb. 13, 2021, doi: 10.1109/TAFFC.2021.3056960.
- [17] J. Shi, I. Alikhani, X. Li, Z. Yu, T. Seppänen, and G. Zhao, "Atrial fibrillation detection from face videos by fusing subtle variations," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2781–2795, Aug. 2020.
- [18] B. P. Yan *et al.*, "Contact-free screening of atrial fibrillation by a smartphone using facial pulsatile photoplethysmographic signals," *J. Amer. Heart Assoc.*, vol. 7, no. 8, 2018, Art. no. e008585.
- [19] O. R. Patil, W. Wang, Y. Gao, W. Xu, and Z. Jin, "A non-contact PPG biometric system based on deep neural network," in *Proc. IEEE 9th Int. Conf. Biometrics Theory Appl. Syst.*, 2018, pp. 1–7.
- [20] J.-M. Park and P. Dave, "Soaring heart rates laid bare on TV as archers okay new tech," *Reuters*. [Online]. Available: <https://www.reuters.com/lifestyle/sports/archery-soaring-heart-rates-laid-bare-tv-archers-okay-new-tech-2021-07-30/>
- [21] W. Chen and R. W. Picard, "Eliminating physiological information from facial videos," in *Proc. IEEE 12th Int. Conf. Autom. Face Gesture Recognit.*, 2017, pp. 48–55.
- [22] M. Chen, X. Liao, and M. Wu, "PulseEdit: Editing physiological signals in facial videos for privacy protection," *IEEE Trans. Inf. Forensics Secur.*, vol. 17, pp. 457–471, 2022.
- [23] D. McDuff and E. Nowara, "'Warm Bodies': A post-processing technique for animating dynamic blood flow on photos and avatars," in *Proc. CHI Conf. Human Factors Comput. Syst.*, 2021, pp. 1–9.
- [24] D. Tran, L. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "Learning spatiotemporal features with 3D convolutional networks," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2015, pp. 4489–4497.
- [25] W. Chen and D. McDuff, "Deepphys: Video-based physiological measurement using convolutional attention networks," in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 349–365.
- [26] X. Niu, S. Shan, H. Han, and X. Chen, "RhythmNet: End-to-end heart rate estimation from face via spatial-temporal representation," *IEEE Trans. Image Process.*, vol. 29, pp. 2409–2423, 2020.
- [27] W. Chen and D. McDuff, "DeepMag: Source-specific change magnification using gradient ascent," *ACM Trans. Graph.*, vol. 40, no. 1, pp. 1–14, 2020.
- [28] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light," *Opt. Exp.*, vol. 16, no. 26, pp. 21434–21445, 2008.
- [29] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. 3rd Int. Conf. Learn. Representations*, 2015, pp. 1–15.
- [30] R. Stricker, S. Müller, and H.-M. Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," in *Proc. IEEE 23rd Int. Symp. Robot Human Interactive Commun.*, 2014, pp. 1056–1062.
- [31] Z. Zhang *et al.*, "Multimodal spontaneous emotion corpus for human behavior analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 3438–3446.
- [32] T. Baltrusaitis, A. Zadeh, Y. C. Lim, and L.-P. Morency, "OpenFace 2.0: Facial behavior analysis toolkit," in *Proc. IEEE 13th Int. Conf. Autom. Face Gesture Recognit.*, 2018, pp. 59–66.
- [33] D. McDuff and E. Blackford, "iPhys: An open non-contact imaging-based physiological measurement toolbox," in *Proc. IEEE 41st Annu. Int. Conf. Eng. Med. Biol. Soc.*, 2019, pp. 6521–6524.
- [34] X. Li *et al.*, "The OBF database: A large face video database for remote physiological signal measurement and atrial fibrillation detection," in *Proc. IEEE 13th Int. Conf. Autom. Face Gesture Recognit.*, 2018, pp. 242–249.
- [35] J. R. Estepp, E. B. Blackford, and C. M. Meier, "Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2014, pp. 1462–1469.
- [36] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," *Pattern Recognit. Lett.*, vol. 124, pp. 82–90, 2019.
- [37] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.
- [38] M. Kumar, A. Veeraraghavan, and A. Sabharwal, "DistancePPG: Robust non-contact vital signs monitoring using a camera," *Biomed. Opt. Exp.*, vol. 6, no. 5, pp. 1565–1588, 2015.
- [39] Q. Zhan, W. Wang, and G. de Haan, "Analysis of CNN-based remote-PPG to understand limitations and sensitivities," *Biomed. Opt. Exp.*, vol. 11, no. 3, pp. 1268–1283, 2020.