sensors

MDPI

Review

Deep Learning Methods for Remote Heart Rate Measurement: A Review and Future Research Agenda

Chun-Hong Cheng [1,2,†], Kwan-Long Wong [2,3,*,†], Jing-Wei Chin [2,4], Tsz-Tai Chan [2,4] and Richard H. Y. So [2,4]

[1] Department of Computer Science, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China; chchengak@connect.ust.hk
[2] PanopticAI, Hong Kong Science and Technology Parks, New Territories, Hong Kong, China; jwchin@connect.ust.hk (J.-W.C.); ttchanac@connect.ust.hk (T.-T.C.); rhyso@ust.hk (R.H.Y.S.)
[3] Department of Bioengineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
[4] Department of Industrial Engineering and Decision Analytics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
* Correspondence: klwongaz@connect.ust.hk
† These authors contributed equally to this work.

check for updates

Citation: Cheng, C.-H.; Wong, K.-L.; Chin, J.-W.; Chan, T.-T.; So, R.H.Y. Deep Learning Methods for Remote Heart Rate Measurement: A Review and Future Research Agenda. *Sensors* **2021**, *21*, 6296. https://doi.org/10.3390/s21186296

Academic Editors: Wenjin Wang, Steffen Leonhardt, Daqing Zhang and Bert den Brinker

Received: 30 July 2021
Accepted: 16 September 2021
Published: 20 September 2021

Abstract: Heart rate (HR) is one of the essential vital signs used to indicate the physiological health of the human body. While traditional HR monitors usually require contact with skin, remote photoplethysmography (rPPG) enables contactless HR monitoring by capturing subtle light changes of skin through a video camera. Given the vast potential of this technology in the future of digital healthcare, remote monitoring of physiological signals has gained significant traction in the research community. In recent years, the success of deep learning (DL) methods for image and video analysis has inspired researchers to apply such techniques to various parts of the remote physiological signal extraction pipeline. In this paper, we discuss several recent advances of DL-based methods specifically for remote HR measurement, categorizing them based on model architecture and application. We further detail relevant real-world applications of remote physiological monitoring and summarize various common resources used to accelerate related research progress. Lastly, we analyze the implications of research findings and discuss research gaps to guide future explorations.

Keywords: noncontact monitoring; heart rate measurement; remote photoplethysmography; rPPG; deep learning

1. Introduction

Human vital signs, such as heart rate (HR), body temperature (BT), respiratory rate (RR), blood oxygen saturation (SpO2), heart rate variability (HRV), and blood pressure (BP), are common indicators used for monitoring the physiological status of the human body [1–4]. They can be used to estimate and analyze a person's physical health, detect possible diseases, and monitor recovery. In particular, closely monitoring a person's HR can enable early detection and prevention of cardiovascular problems, such as atherosclerosis (heart block) and arrhythmia (irregular heart rate) [5].

Photoplethysmography (PPG) is a common method for measuring HR. It utilizes a light source and photodetector to measure the volumetric changes of blood vessels under the skin [6,7]. As the light source illuminates the tissue, small variations in reflected or transmitted light intensity from blood flow are captured by the photodetector, yielding the so-called PPG signal [7]. The absorption of light follows the Beer–Lambert law, which states that the light absorbed by blood is proportional to the penetration of light into the skin and the concentration of hemoglobin in the blood [8]. During the cardiac cycle, minute variations in hemoglobin concentration cause fluctuations in the amount of light absorbed by the blood vessels, resulting in changes of skin intensity values. Pulse oximeters are commonly used for non-invasive measurement of these slight variations in the skin through PPG.

However, as with other wearables and contact-based devices (e.g., smartwatches), they are unsuitable for monitoring newborns or patients with fragile skin [9,10]. Furthermore, long-term monitoring may lead to discomfort and even the risk of skin infections [11]. As a result, remote PPG (rPPG) methods have emerged as an attractive alternative.

During the last decade, rPPG methods have gained significant traction. In rPPG, a digital camera (e.g., webcam, standard RGB camera, near-infrared camera) functions as a photodetector that captures subtle color changes of the skin; ambient light typically serves as the light source [12]. Figure 1 illustrates the principle of rPPG with the dichromatic reflection model (DRM) [13]. According to the DRM, the signals captured by the digital camera are a combination of specular reflections (surface reflections) and diffuse reflections (body reflections). Specular reflections occur at the interface of the incident light and the skin, which do not contain meaningful physiological signals. Thus, rPPG methods utilize signal processing techniques to separate the specular reflections and extract the diffuse reflections associated with the underlying signals of interest. The ability for contactless measurement can significantly reduce monitoring costs and enable applications where traditional contact sensors would be suboptimal [14]. However, while rPPG technology will undoubtedly play a pivotal role in the future of digital healthcare, the extracted signals are inherently much weaker and require meticulous processing.

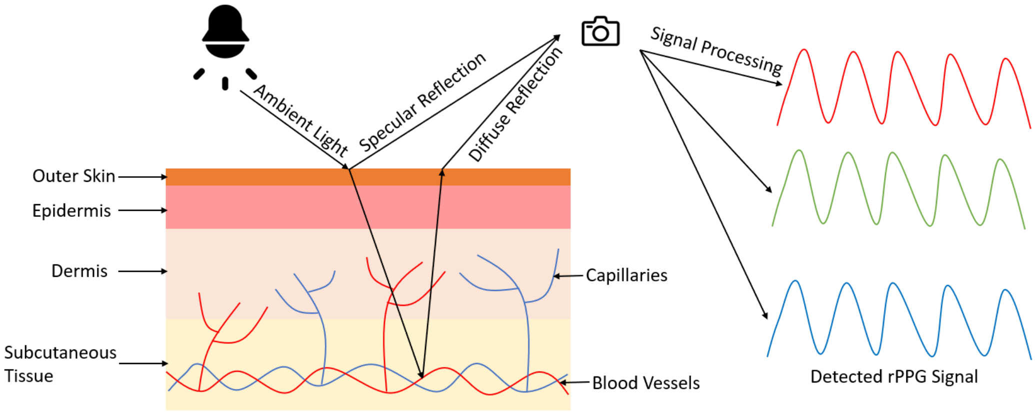

Figure 1. Principle of remote photoplethysmography (rPPG) based on the dichromatic reflection model (DRM). The digital camera captures the specular and diffuse reflection from ambient light. The specular reflection contains surface information that does not relate to physiological signals, while the diffuse reflection is modulated by blood flow. The rPPG signal can be obtained from further signal processing.

Verkruysse et al. [15] was the initial research that used a consumer-level camera with ambient light for measurement of rPPG signals. In their work, the green channel was found to contain the most significant PPG signal. Poh et al. [16] applied a blind source separation (BSS) technique, independent component analysis (ICA), on the recorded RGB color channels from a webcam to recover HR. Lewandoska et al. [17] applied a similar method, principal component analysis (PCA), which reduced the computational complexity while achieving a similar accuracy to ICA. However, these methods are subject to motion artifacts. To improve the motion robustness of the rPPG model, a chrominance-based approach (CHROM) was proposed [18]. In this approach, the dichromatic reflection model was used to describe the light reflected from the skin as specular and diffuse reflection components [19]. De Haan and van Leest [20] defined a blood-volume pulse vector, which represents the signature of blood volume change, to identify the subtle color changes due to the pulse from motion artifacts based on RGB measurement. Later, Wang et al. [21] proposed a data-driven algorithm, spatial subspace rotation (2SR), to estimate a spatial subspace of skin pixels and evaluate its temporal rotation to measure HR. Wang et al. [13] further proposed a plane-orthogonal-to-skin (POS) algorithm that defines a projection plane orthogonal to skin tone in the RGB space to extract the pulse signal. Further information about early conventional rPPG methods can be found in the following surveys [14,22,23].

Most conventional methods for remote HR measurement follow a similar framework as shown in Figure 2. Firstly, a digital camera captures a video recording of the subject. Next, a face detection algorithm, such as the Viola and Jones algorithm [24], is applied to obtain the bounding box coordinates of the subject's face. This is followed by selecting regions of interest (ROIs), such as the cheeks, to obtain an area that contains a strong signal. The pixels within the ROI(s) are used for rPPG signal extraction and HR is estimated by further post-processing, which typically involves frequency analysis and peak detection.

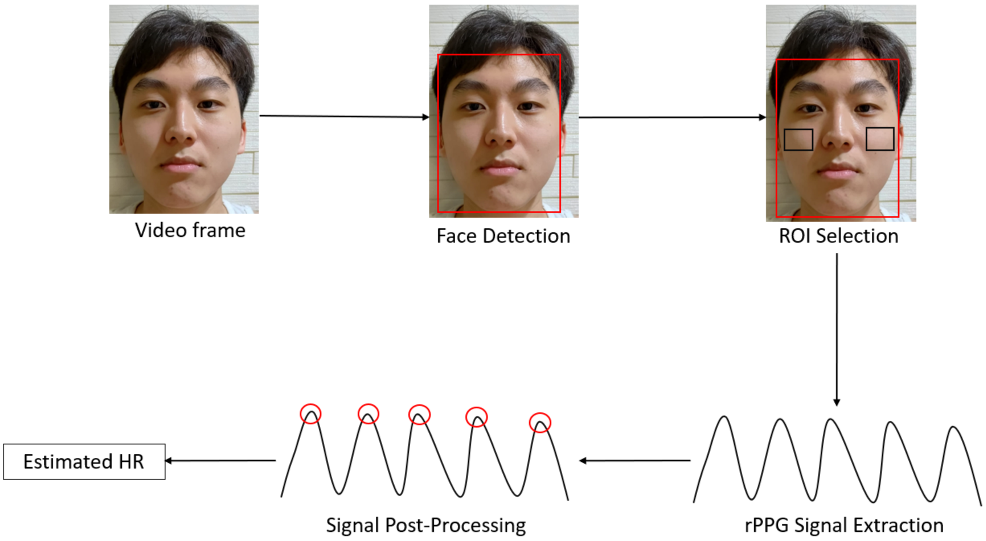

Figure 2. General framework of conventional methods for remote heart rate (HR) measurement. Face detection (e.g., Viola and Jones algorithm) is performed on the video frames, resulting in the red bounding box on the face. Next, regions of interest (ROIs) such as the cheeks marked by the black boxes are selected within the face box. The rPPG signal is extracted from the pixels within the ROIs. Lastly, post-processing techniques, such as frequency analysis (e.g., Fourier transform) and peak detection, are applied on the extracted signal to estimate HR.

As with many computer vision and signal processing applications, DL methods have shown promise in mapping the complex physiological processes for remote HR measurement. While many review papers have discussed the conventional techniques for non-contact physiological monitoring [10,12,14,23], there is limited emphasis on DL methods, despite their popularity in the research community. The number of research papers utilizing DL methods for remote HR measurement has increased year after year and is expected to grow continuously. Our paper aims to provide researchers with an extensive review of DL approaches for remote HR measurement and an improved understanding of their benefits and drawbacks.

In the following sections of this paper, we categorize DL approaches for remote HR measurement as end-to-end and hybrid DL methods. We proceed to classify them based on model architecture and critically analyze their methods. We then discuss the real-world applications that benefit from this technology and introduce some common resources, including toolboxes, datasets, and open challenges for researchers in this field. Finally, we analyze the current knowledge gaps and suggest future directions for research.

2. End-to-End Deep Learning Methods

In this section, we detail the end-to-end DL approaches for remote HR measurement. We classify a method as end-to-end if it takes in a series of video frames as input and directly outputs the HR without any intermediate steps. Since many DL methods are designed to output the rPPG signal, these are also grouped in the same category for subsequent analysis (Figure 3). As shown in Table 1, the methods are further classified based on the type of DL technique used. While end-to-end DL methods are indisputably great tools due to their straightforward model optimization process, they require enormous amounts of training data and are difficult to validate. More work needs to be done on the interpretation of such models for translation to clinical application [25].

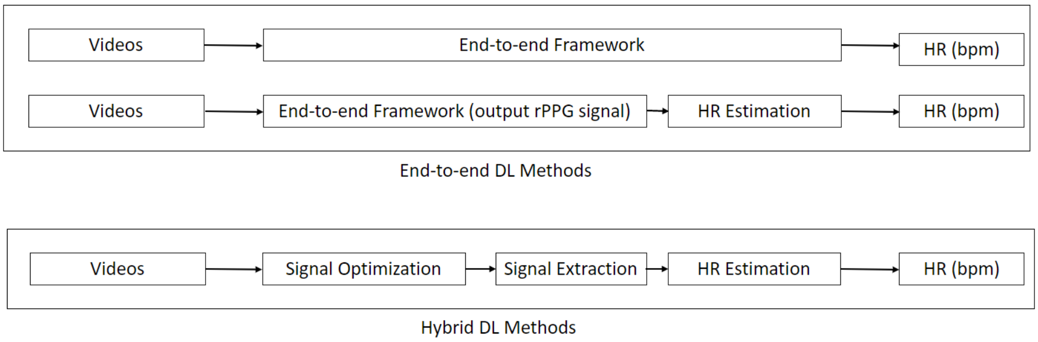

Figure 3. Schematic diagram of end-to-end deep learning (DL) methods and hybrid DL methods. End-to-end DL methods directly output the HR or rPPG signal with a single model, while hybrid DL methods utilize DL techniques at various stages.

Table 1. Summary of end-to-end deep learning (DL) methods for remote HR measurement.

Ref.	Year	2D CNN	3D CNN	2D CNN + RNN	NAS	Attention
[26]	2018	✓				
[27]	2018	✓				✓
[28]	2020	✓				✓
[29]	2019		✓	✓		
[30]	2019		✓			✓
[31]	2020		✓		✓	
[32]	2021		✓			✓
[33]	2021		✓			
[34]	2021		✓			✓
[35]	2019			✓		✓

2.1. 2D Convolutional Neural Network (2D CNN)

Špetlík et al. [26] proposed an end-to-end HR estimation approach, where the output of the model was a single scalar value of the predicted HR. HR-CNN is a two-step CNN that contains an extractor and an HR estimator. The 2D CNN extractor was trained to maximize the signal-to-noise ratio (SNR) in order to extract the rPPG signal from a sequence of video frames. Then, the extracted rPPG signal was fed into the HR estimator to output the predicted HR value, where the training process minimized the mean absolute error (MAE) between the predicted and ground truth HR. Špetlík et al. [26] claimed that their proposed method better addressed video compression artifacts, where most conventional rPPG signal extraction methods fail. They validated it on three public datasets, as well as proposed a new challenging dataset (ECG-Fitness) which contained different motions and lighting conditions.

DeepPhys [27] is a VGG-style 2D CNN that jointly trained a motion and appearance model (Figure 4). The motion model took the normalized difference between adjacent frames as an input motion representation; it is built on top of the dichromatic reflection model for modeling motions and color changes. The appearance model guided the motion model to learn motion representation through an attention mechanism. The network learned soft-attention masks from the original video frames and allocated higher weights to skin areas with stronger signals. This attention mechanism also enabled the visualization of the spatio-temporal distribution of physiological signals. With the motion representation and attention mechanism, Chen and McDuff [27] claimed that physiological signals under different lighting conditions can be better captured, being more robust to illumination changes and subject motion.

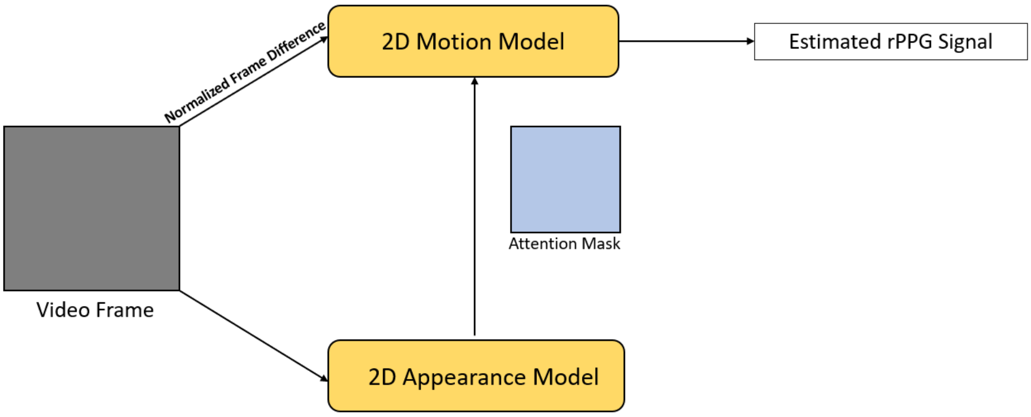

Figure 4. Architecture of DeepPhys [27].

MTTS-CAN [28] is an improvement built on top of DeepPhys [27]. MTTS-CAN captured temporal information through the introduction of a temporal shift module (TSM) [36]. TSM allowed information exchange among neighboring frames and avoided expensive 3D convolution operations by shifting chunks in the tensor along the temporal axis. In addition, the input of the appearance model was a frame obtained by performing averaging adjacent multiple frames rather than the original video frame. Furthermore, it estimated HR and RR simultaneously by using a multi-task variant. Since this network was completely based on 2D CNNs, it only took 6 ms per frame for on-device inference, which demonstrated its potential of being utilized in real time applications.

2.2. Spatio-Temporal Network—3D Convolutional Neural Network (3D CNN)

As 2D CNNs only take spatial information of video frames into account, researchers have proposed different 3D CNN frameworks to also make use of the temporal information contained in the videos. These so-called spatio-temporal networks (STNs) can provide a more comprehensive representation of the spatial and temporal information of the physiological signals in the video stream.

Three-dimensional CNN PhysNet [29] is an end-to-end STN aimed at locating the peak of every individual heartbeat (Figure 5). It is able to estimate both HR and HRV accurately, allowing more complicated applications, such as emotion recognition. It took the original RGB video frames as input and directly output the final rPPG signal. In addition, it utilized the negative Pearson correlation as the loss function in order to have higher trend similarity and fewer peak location errors.

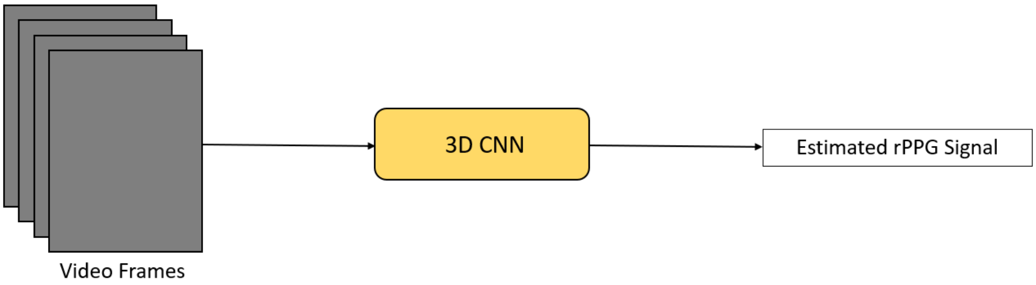

Figure 5. Architecture of 3D CNN PhysNet [29].

Yu et al. [30] proposed a two-stage end-to-end STN to not only estimate the rPPG signal but also to overcome the problem of highly compressed facial videos (Figure 6). Compressed facial videos were fed into a spatio-temporal video enhancement network (STVEN) to improve the quality of the videos while retaining as much information as possible. The enhanced videos were further fed into a spatio-temporal 3D CNN (rPPGNet) to extract the rPPG signal. Inside rPPGNet, an attention mechanism was applied to obtain dominant rPPG features from skin regions. rPPGNet is able to operate individually for rPPG signal extraction but can be trained jointly with STVEN to achieve better performance. Yu et al. [30] claimed that rPPGNet is able to recover better rPPG signals with curves and peak locations for accurate HR and HRV estimation.

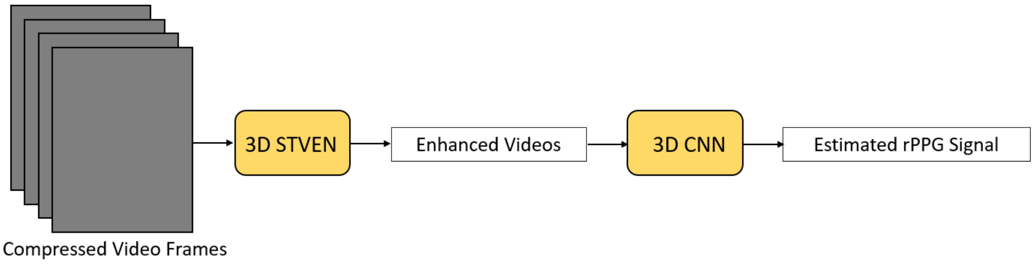

Figure 6. Architecture used in Yu et al. [30].

Yu et al. [31] utilized neural architecture search (NAS) to automatically find the best-suited backbone 3D CNN for rPPG signal extraction (Figure 7). In their research, a special 3D convolution operation, namely temporal difference convolution (TDC), was designed to help track the ROI and improve the robustness in the presence of motion and poor illumination. Then, NAS was performed based on two gradient-based NAS methods [37,38] in order to form a backbone network for rPPG signal extraction. Two data augmentation methods were proposed, as well, in order to prevent data scarcity.

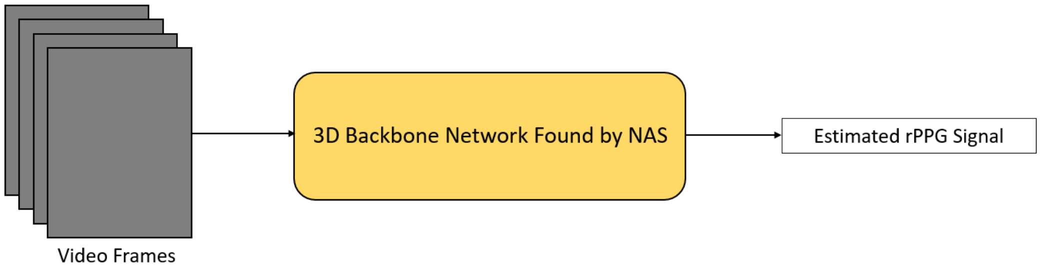

Figure 7. Architecture of AutoHR [31].

Hu et al. [32] designed a novel facial feature extraction method in order to avoid extracting redundant information from video segments and to enhance long-range video temporal modeling. A 3D CNN was used to extract facial features of the input video frames. Next, aggregation functions were applied to incorporate long-range spatio-temporal feature maps into short segment spatio-temporal feature maps. These feature maps were then fed into a signal extractor with several spatio-temporal convolution [39] to extract the rPPG signal. A spatio-temporal strip pooling method and an attention mechanism were further applied to the extracted rPPG signal to accommodate head movement and avoid ignoring important local information.

Zhang et al. [33] proposed an efficient multi-hierarchical convolutional network to perform estimation quickly, where only 15 s of face video was required for effectively reconstructing the rPPG signal and estimating HR. A three-layer 3D CNN was used to extract low-level facial feature maps from RGB face videos. These feature maps were passed to a spatio-temporal stack convolution module for deeper feature extraction and generation of a high-level feature map. Channel-wise feature extraction was then performed on the high-level feature map to produce a channel-wise feature map. A skin map was also generated based on low-level feature maps for emphasizing skin regions with stronger signals. Next, a weight mask was constructed by performing feature fusion on the skin map and the channel-wise feature map. Finally, the high-level feature map was multiplied by the weight mask by channels and was fed into a rPPG signal extractor.

ETA-rPPGNet [34] is another network aimed at dealing with the problem of extracting redundant video information (Figure 8). In this network, a time-domain segment subnet was designed to model the long-range temporal structure of the video. Split video segments were passed to different subspace networks of this subnet to extract facial features. Then, an attention mechanism was applied to learn important spatial features. Next, an aggregation function was used to aggregate the temporal context in order to cut down redundant video information and a feature map was obtained in each individual subspace network. These individual feature maps were concatenated and fed into the backbone network for rPPG signal extraction. Inside the backbone network, an attention module was also added for eliminating different noise (e.g., head movement, illumination variation). Finally, the extracted rPPG signal was further processed by a 1D convolution operation to model the correlation held in the local time domain effectively.

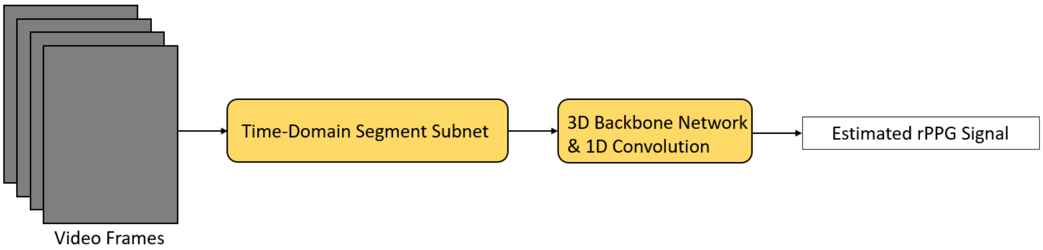
Video Frames
Time-Domain Segment Subnet
3D Backbone Network & 1D Convolution
Estimated rPPG Signal

Figure 8. Architecture of ETA-rPPGNet [34].

2.3. Spatio-Temporal Network—2D Convolutional Neural Network + Recurrent Neural Network (2D CNN + RNN)

Researchers have also designed another type of spatio-temporal network, which is the combination of 2D CNN for spatial information and RNN for temporal context.

In the same work of Reference [29], a different version of PhysNet, which combined a 2D CNN with different RNNs (LSTM, BiLSTM, ConvLSTM [40]) was proposed to compare the performance of 3D CNN-based PhysNet and RNN-based PhysNet and evaluate the performance of different RNNs (Figure 9). The input and output of the network remained the same as for the 3D CNN PhysNet. The input was firstly fed into a 2D CNN to extract spatial features of the RGB video frames; then, the RNN was used to propagate these spatial features in the temporal domain. In their research, 3D CNN-based PhysNet achieved a better performance than RNN-based PhysNet, and the BiLSTM variant had the worst performance, indicating the backward information flow of spatial features was not necessary. Table 2 shows the performance of different versions of PhysNet in terms of root mean square error (RMSE) and Pearson correlation coefficient (R) [29].

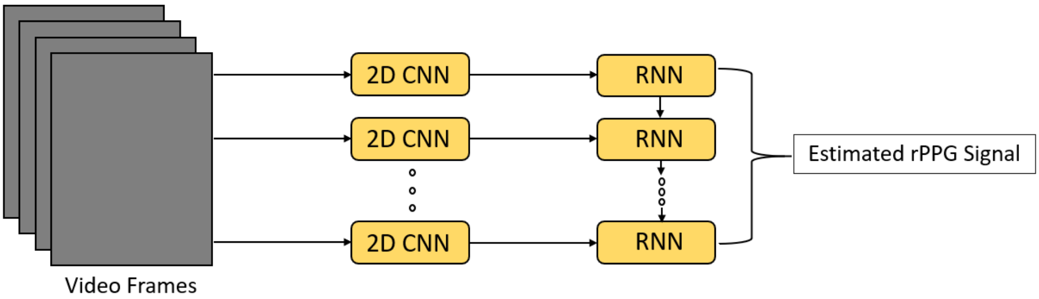

Figure 9. Architecture of RNN-based PhysNet [29].

Table 2. Performance of different versions of PhysNet [29] on the OBF [41] dataset. Root mean square error (RMSE) in beats per minute (bpm) and Pearson correlation coefficient (R) were used as the evaluation metrics.

3D CNN-Based	LSTM Variant	BiLSTM Variant	ConvLSTM Variant
RMSE = 2.048, R = 0.989	RMSE = 3.139, R = 0.975	RMSE = 4.595, R = 0.945	RMSE = 2.937, R = 0.977

In Reference [35], another combination of 2D CNN with a ConvLSTM network with attention mechanism was proposed for rPPG signal extraction. The 2D CNN part had a similar approach as DeepPhys [27], which consisted of a trunk branch and a mask branch. The trunk branch was used to extract spatial features from a sequence of face images, while the mask branch learned and generated attention masks and passed them to the trunk branch to guide feature extraction. These spatial features were then fed into a ConvLSTM network in order to make use of the temporal correlation held in video frames for rPPG signal extraction.

3. Hybrid Deep Learning Methods

In this section, we describe hybrid DL methods for remote HR measurement. For hybrid DL methods, DL techniques are only applied in some parts of the pipeline. We further indicate whether the methods are used for signal optimization, signal extraction, or HR estimation (Figure 3).

3.1. Deep Learning for Signal Optimization

In most existing remote HR measurement pipelines, the input is the original video recorded by a digital camera. Therefore, face detection or skin segmentation is needed to ignore irrelevant background information. Moreover, some specific skin regions, such as the cheeks, contain stronger signals and are usually selected as the ROI [42]. In this subsection, we describe these DL-based signal optimization methods to enable more effective signal extraction.

In Reference [43], a 2D CNN for skin detection was created and trained on a private video database. Both skin and non-skin region samples were manually segmented and treated as positive and negative samples, respectively. Conventional rPPG algorithms (ICA and PCA) were then performed on the detected skin region for evaluation. Tang et al. [43] suggested that low-cost cameras could capture rPPG signals with their method, which worked on single-channel input by choosing the RGB channel with the least noise under different conditions. This method could be combined with traditional rPPG methods in order to improve their performance. However, it utilized all the skin areas of the face for rPPG signal extraction, which may include unnecessary noise. Moreover, their method was only validated on a private dataset with yellow skin tones.

In Reference [44], a single-photon avalanche diode (SPAD) camera was used to record videos. This camera was able to work well in dark environments. The recorded frame was a low-resolution grayscale image. A 2D CNN encoder-decoder model took this as input and produced a single channel image with values between zero and one, representing the probability that the particular pixel was regarded as skin. In addition, a transfer learning approach was adopted in the training process due to the lack of data for this specific skin detection problem. The model was trained on a large dataset of unlabeled face images for colorization and then further trained on a skin mask dataset. Finally, a binary skin mask was obtained by thresholding and passed for signal extraction.

In Deep-HR [45], a receptive field block (RFB) network was utilized to detect the ROI as an object [46]. This network was trained on a private dataset with videos recorded in realistic settings to improve overall robustness. Furthermore, a generative adversarial network (GAN)-style module was designed to enhance the detected ROI. A CNN that learned the distribution of high-quality ROIs acted as a discriminator to supervise another deep encoder-decoder network that served as a generator to regenerate the detected ROI. This high-quality detected ROI was passed for subsequent signal extraction. The architecture used for signal optimization in Deep-HR [45] is illustrated in Figure 10.

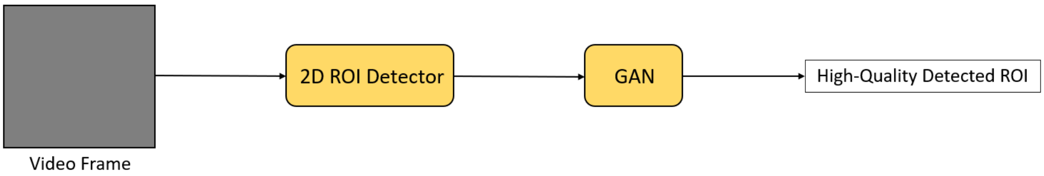

Figure 10. Architecture used in Deep-HR [45] for signal optimization.

3.2. *Deep Learning for Signal Extraction*

Signal extraction is the most important part in the remote HR measurement pipeline, and it is the leading focus in this research field. Its principal goal is to extract the rPPG signal from videos for HR estimation. In addition, refining the extracted rPPG signal for better HR estimation is a method to improve the estimation accuracy. Researchers have proposed many different DL methods for obtaining a high-quality rPPG signal, and we are going to categorize and describe them based on the type of neural network being used. Table 3 shows neural networks used in different DL-based signal extraction methods.

Table 3. Summary of hybrid DL methods for signal extraction in remote HR measurement pipeline.

Ref.	Year	LSTM	2D CNN	3D CNN	2D CNN + RNN	3D CNN + RNN	GAN
[47]	2019	✓					
[48]	2020	✓					
[45]	2020		✓				✓
[49]	2021		✓				
[50]	2019			✓			
[51]	2020			✓			
[52]	2020			✓			
[53]	2020			✓			
[54]	2020			✓			
[55]	2019				✓		
[56]	2020				✓		
[57]	2020				✓		
[58]	2021					✓	
[59]	2021						✓

3.2.1. Long Short-Term Memory (LSTM)

In References [47,48], a LSTM network was applied for signal filtering, improving the quality of the extracted rPPG signal. As the rPPG signal extracted by conventional methods may contain several noise, filtering the noise-contaminated rPPG signal is able to produce a noiseless rPPG signal for more accurate HR estimation. The LSTM network in Reference [47] was firstly trained on a large amount of synthetic data. Then, it was further trained on real data for model fine-tuning, enhancing its generalization ability. This method is able to effectively overcome the problem of data shortage. The architecture used for signal filtering in Reference [47] is shown in Figure 11.

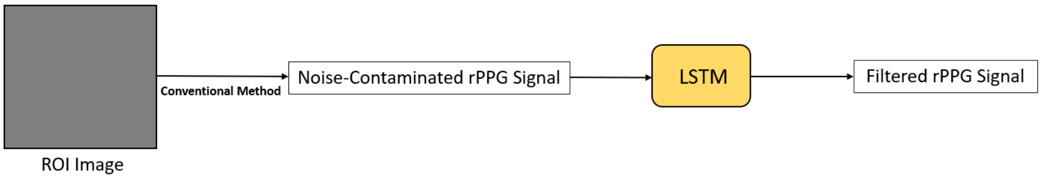

Figure 11. Architecture used in Bian et al. [47].

3.2.2. 2D Convolutional Neural Network (2D CNN)

In Deep-HR [45], a 2D CNN was learned to extract color information of the ROI pixels (Figure 12). Noise was further removed from the extracted information by using a GAN-style module. A discriminator that accesses high-quality rPPG signals was used to guide a generator to reconstruct a noiseless rPPG signal. This noise removing technique can be applied in other rPPG methods to improve the performance, as well.

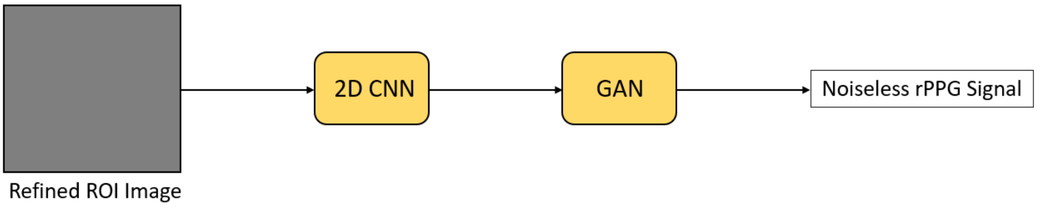

Figure 12. Architecture used in Deep-HR [45] for signal extraction.

MetaPhys [49] utilized a pretrained 2D CNN, namely TS-CAN, which is another version of MTTS-CAN [28] for signal extraction. The difference between them was the use of the multi-task variant so that TS-CAN could only estimate HR and RR one at a time, while MTTS-CAN could estimate HR and RR simultaneously. Furthermore, a meta-learning approach was proposed for better generalization of the model. Model-Agnostic Meta-Learning (MAML) [60] was utilized as the personalized parameter update schema to produce a general initialization so that fast adaptation could be performed when only a few training samples were available. In addition, both supervised and unsupervised training methods were evaluated on MetaPhys. Liu et al. [49] claimed that this approach can reduce bias due to skin tone, improving the robustness of the model.

3.2.3. Spatio-Temporal Network—3D Convolutional Neural Network (3D CNN)

In Reference [50], a 3D CNN was designed to extract features from unprocessed video streams, followed by a multilayer perceptron to regress HR. In the paper, a data augmentation method was also proposed for generating realistic videos effectively with synthetic rPPG signals. The synthetic rPPG signal was transformed to a video by using vector repetition. Noise was also added to the synthetic videos in order to make them realistic.

Siamese-rPPG [51] is a framework based on a Siamese 3D CNN (Figure 13). The idea behind this framework is different facial regions may suffer from different noise and have their own appearances. However, they should reflect more or less the same rPPG characteristics. Therefore, the forehead and cheek regions with more rPPG information were firstly selected as the ROI. Next, pixels in these two ROIs were passed to the forehead branch and the cheek branch for extraction, respectively; both were 3D CNNs with the same architecture. Weight sharing mechanism was also applied to these two branches so that, even if either the cheek or forehead region was contaminated with noise, the framework could use the other region for signal extraction, improving the overall robustness. After that, the outputs from these two branches were fused by an addition operation, followed by two 1D convolutional operations and an average pooling, to produce the predicted rPPG signal.

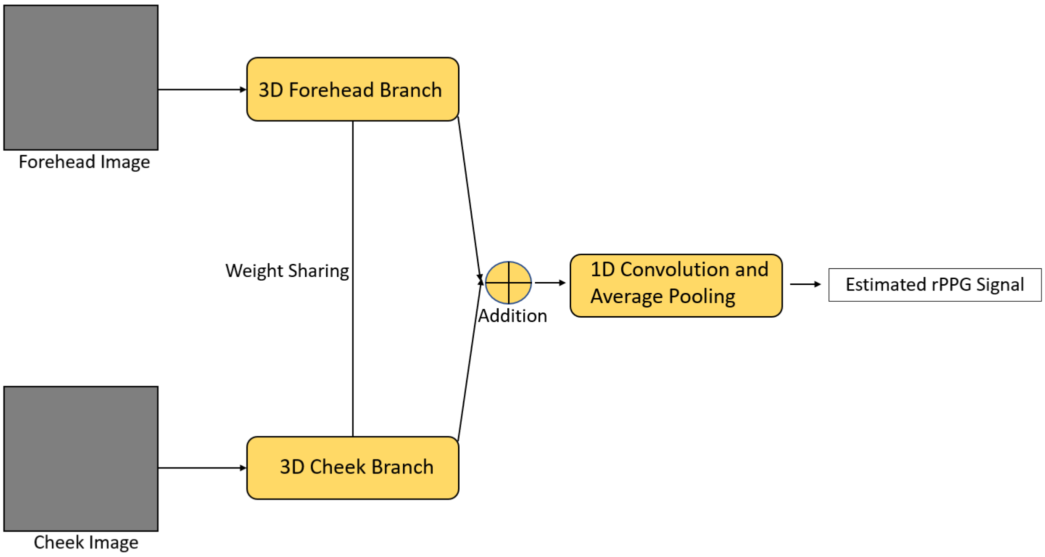

Figure 13. Architecture of Siamese-rPPG [51].

HeartTrack [52] utilized a 3D CNN with attention mechanism for signal extraction. In this 3D spatio-temporal attention network, a hard attention mechanism was used to help the network ignore unrelated background information and a soft attention mechanism was used to help the model filter out covered areas. The extracted signal was further fed into a 1D CNN for time series analysis. Synthetic data was also used in the training process in order to address the problem of inadequate real data.

In Reference [53], a multi-task framework was proposed for learning a rPPG signal extraction model and augmenting data simultaneously. There were a total of 3 main networks in this framework. The first one was a signal extractor that directly extracted the rPPG signal from the input facial videos. The second one was a reconstruction network for generating synthetic videos from real images. The third one was also a reconstruction network for generating synthetic videos from real videos. They were designed to support each other, and these two reconstruction networks could effectively handle the problem of insufficient training data and improve the overall robustness.

DeeprPPG [54] is a framework that can use different skin regions as the input for rPPG signal measurement, allowing customized ROI selection and wider applications. It took a skin region clip from the original video as the input and a spatio-temporal network was utilized to extract the rPPG signal. A spatio-temporal aggregation function was also proposed for easing the side effect of regions contaminated by different noise and improving the robustness of the model.

3.2.4. Spatio-Temporal Network—2D Convolutional Neural Network + Recurrent Neural Network (2D CNN + RNN)

In Reference [55], a two-stream approach was adopted for feature extraction and rPPG signal extraction. For the feature extraction stream, a 2D CNN with low-rank constraint loss function was proposed to force the network to learn synchronized spatial features from spatio-temporal maps, improving the robustness of face detection and ROI alignment errors. For the rPPG signal extraction stream, a 2D CNN was firstly used to extract the rPPG signal, and then the rPPG signal was further refined by a two-layer LSTM network. Lastly, the outputs from these two streams were concatenated for HR estimation.

In Reference [56], a 2D CNN was used to extract spatial features and local temporal information, and an LSTM network was utilized for extracting global temporal information held in consecutive frames. One fully connected layer was further applied to the output of the LSTM to estimate HR. This framework was able to overcome processing latency and update HR in about 1 s, showing the potential of being adopted in real-time HR monitoring.

Meta-rPPG [57] utilized a transductive meta-learner to take unlabeled data during deployment for self-supervised weight adjustment, allowing fast adaptation to different distribution of samples (Figure 14). In this framework, a ResNet-alike convolutional encoder was firstly used to extract latent features from a stream of face images. Next, these extracted features were passed to a BiLSTM network to model the temporal context, followed by a multilayer perceptron (MLP) for rPPG signal estimation. A synthetic gradient generator was also proposed for transductive learning. It was based on a shallow Hourglass network [61] and further applied to a few-shot learning framework in order to generate gradients for unlabeled data [62].

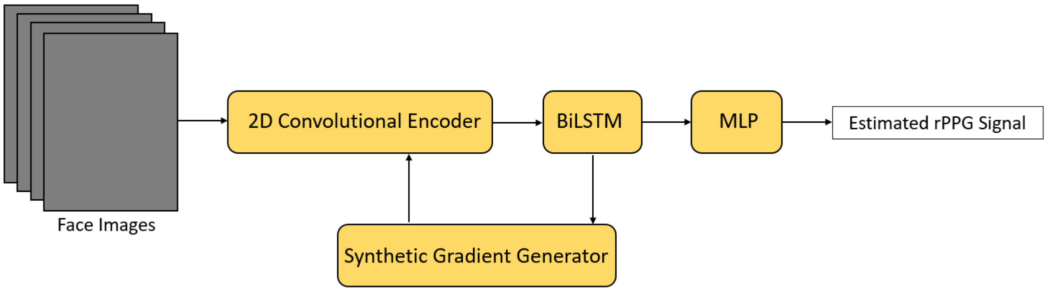

Figure 14. Architecture of Meta-rPPG [57].

3.2.5. 3D Convolutional Neural Network + Recurrent Neural Network (3D CNN + RNN)

PRNet [58] is a one-stage spatio-temporal framework for HR estimation from stationary videos (Figure 15). Firstly, a 3D CNN extractor was utilized to extract spatial features and capture local temporal features from the defined ROI. Next, the output feature map was further fed into an LSTM extractor for extracting global temporal features. Lastly, a fully connected layer was applied to estimate HR from the extracted feature map. Huang et al. [58] claimed that this framework is able to predict HR with only 60 frames of the video (2 s), while other remote HR estimation methods usually need 6–30 s of the video.

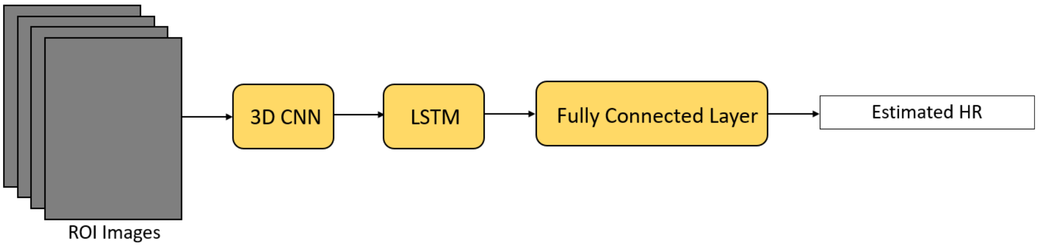

Figure 15. Architecture of PRNet [58].

3.2.6. Generative Adversarial Network (GAN)

PulseGAN [59] is a framework based on GAN to generate realistic rPPG signals (Figure 16). In the paper, a rough rPPG signal was firstly obtained by applying the CHROM algorithm on the defined ROI. Then, PulseGAN took this as input and generated a high-quality, realistic rPPG signal for performing HR estimation accurately. Moreover, the structure of PulseGAN was based on the conditional GAN approach [63]. The discriminator accessed the ground truth rPPG signal and guided the generator to map a rough rPPG signal extracted by CHROM to a final rPPG signal that is similar to the ground truth one. The rough rPPG signal was also set as a condition in the discriminator. Song et al. [59] mentioned that this framework can be combined with other conventional rPPG methods easily in order to improve the quality of the extracted rPPG signal, resulting in more accurate HR estimation.

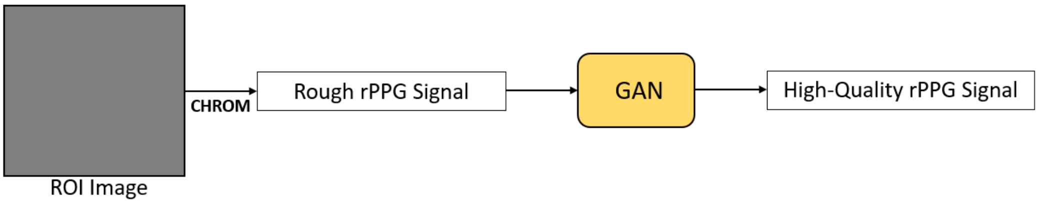

Figure 16. Architecture of PulseGAN [59].

3.3. Deep Learning for Heart Rate Estimation

Traditionally, the extracted rPPG signal can be filtered with a bandpass filter followed by frequency analysis or peak detection to estimate HR. However, HR estimation can also be classified as a regression problem and solved by DL methods. Moreover, different representations of the HR signal have been proposed for DL-based HR estimation (Table 4).

In Reference [64], the rPPG signal was extracted by conventional methods (e.g., ICA, PCA, CHROM), and short-time Fourier transform and bandpass filtering were applied to the extracted rPPG signal to obtain a frequency domain representation. This representation was further combined with the time domain signal to form a spectrum image, a kind of HR signal representation. Lastly, an HR estimator based on ResNet18 [65] pretrained with the ImageNet dataset was used to estimate HR from spectrum images directly. Based on this method, HR can be estimated accurately regardless of which conventional methods were used, since the HR estimator can learn features in spectrum images and directly map them into HR. The architecture used for HR estimation in Reference [64] is illustrated in Figure 17.

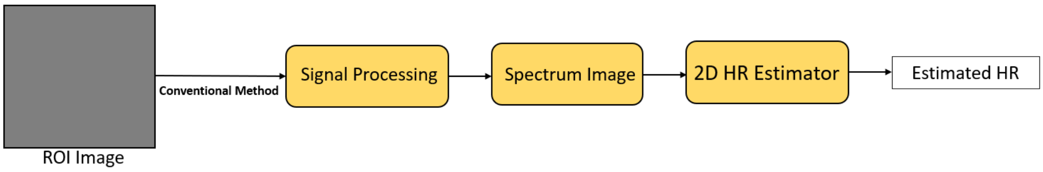

Figure 17. Architecture used in Yang et al. [64].

Table 4. Summary of all mentioned end-to-end and hybrid DL methods for remote HR measurement.

Ref.	Year	End-to-End/Hybrid	Description
[26]	2018	End-to-End	End-to-end HR estimation with an extractor and an estimator
[27]	2018	End-to-End	Normalized frame difference as motion representation, attention mechanism was used to guide the motion model, visualization of spatio-temporal distribution of physiological signals
[43]	2018	Hybrid	2D CNN network for skin detection
[64]	2018	Hybrid	Spectrum images were used for HR estimation
[66]	2018	Hybrid	Spatio-temporal maps were used for HR estimation, transfer learning approach to deal with data shortage
[29]	2019	End-to-End	Compared 3D CNN-based and RNN-based spatio-temporal network, can estimate HR and HRV accurately
[30]	2019	End-to-End	Enhancing video quality to deal with highly compressed videos, can estimate HR and HRV accurately
[35]	2019	End-to-End	Attention mechanism was used to guide the trunk branch for signal extraction
[47]	2019	Hybrid	LSTM network for signal filtering, transfer learning approach to deal with data shortage
[50]	2019	Hybrid	3D CNN for signal extraction, data augmentation method for generating videos with synthetic rPPG signals, multilayer perceptron for HR estimation
[55]	2019	Hybrid	2D CNN-based two-stream approach for signal extraction, and LSTM network for signal refining
[67]	2019	Hybrid	Spatio-temporal maps were used for HR estimation, attention mechanism was applied to remove noise
[28]	2020	End-to-End	Temporal shift module to model temporal information, attention mechanism was applied to guide the motion model, able to estimate HR and RR simultaneously by one network
[31]	2020	End-to-End	Used NAS to find a well-suited network for HR estimation
[44]	2020	Hybrid	2D CNN encoder-decoder model for skin detection, transfer learning approach to deal with data shortage
[45]	2020	Hybrid	Two GAN-style modules to enhance the detected ROI and remove noise, 2D CNN for signal extraction
[48]	2020	Hybrid	LSTM network for signal filtering
[51]	2020	Hybrid	Siamese 3D CNN for signal extraction
[52]	2020	Hybrid	3D CNN with attention mechanism for signal extraction, feedforward neural network for HR estimation
[54]	2020	Hybrid	3D CNN that can take different skin regions for signal extraction
[56]	2020	Hybrid	2D CNN + LSTM spatio-temporal network for signal extraction
[57]	2020	Hybrid	2D CNN + BiLSTM spatio-temporal network for signal extraction, meta-learning approach for fast adaptation
[68]	2020	Hybrid	Spatio-temporal maps were used for HR estimation
[69]	2020	Hybrid	Spatio-temporal maps were used for HR estimation
[70]	2020	Hybrid	Spatio-temporal maps were used for HR estimation, transfer learning approach to deal with data shortage
[32]	2021	End-to-End	Avoid extracting redundant information from video segments, attention mechanism was applied to deal with different noise
[33]	2021	End-to-End	An efficient framework for performing HR estimation quickly
[34]	2021	End-to-End	Dealt with the problem of extracting redundant video information, attention mechanism was applied to learn important features and eliminate noise
[49]	2021	Hybrid	TS-CAN from another paper was utilized for signal extraction, meta-learning approach for fast adaptation
[53]	2021	Hybrid	Multi-task framework for simultaneous signal extraction and data augmentation
[58]	2021	Hybrid	3D CNN + LSTM spatio-temporal network for signal extraction
[59]	2021	Hybrid	GAN for generating high-quality rPPG signal from rough rPPG signal
[71]	2021	Hybrid	Spatio-temporal maps were used for HR estimation, NAS was used to find a CNN for mapping spatio-temporal maps into HR

Another type of HR signal representation is the spatio-temporal map (Figure 18) used for HR estimation in References [66–71]. Generally, an ROI selection step was involved in the construction of these spatio-temporal maps. Color information of the RGB channels of the ROI pixels was utilized and concatenated in temporal sequences, and placed into rows to form a spatio-temporal map. Finally, a neural network was used to estimate HR from spatio-temporal maps directly. This kind of HR signal representation can highlight the HR signal and suppress the information that is unrelated to the HR signal. In References [66,70], transfer learning was applied to pretrain the HR estimator with the ImageNet dataset to deal with insufficient data. In Reference [68], a combination of 2D CNN and gated recurrent unit (GRU) was used for HR estimation (Figure 19). In Reference [71], NAS was also utilized to find a lightweight and optimum CNN to estimate HR from spatio-temporal maps. In Reference [67], an attention module was added to mitigate the effect of different noise.

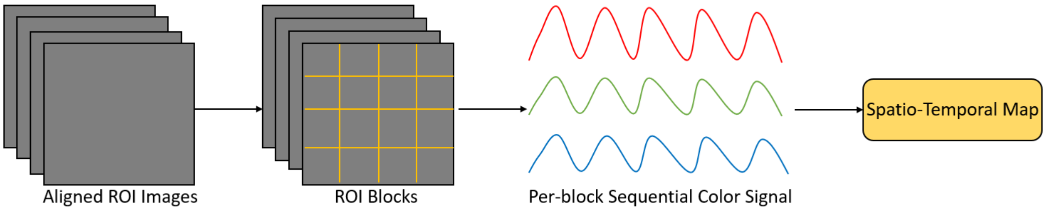

Figure 18. General procedure of constructing a spatio-temporal map. Firstly, the images are aligned and ROI selection is performed to obtain ROI images. Then, these ROI images are divided into several ROI blocks. Next, within each block, the average color value is calculated for each color channel. After that, the average color value of each channel at the same block but different frames are concatenated into temporal sequences. Finally, the temporal sequences of each block are placed into rows to form a spatio-temporal map.

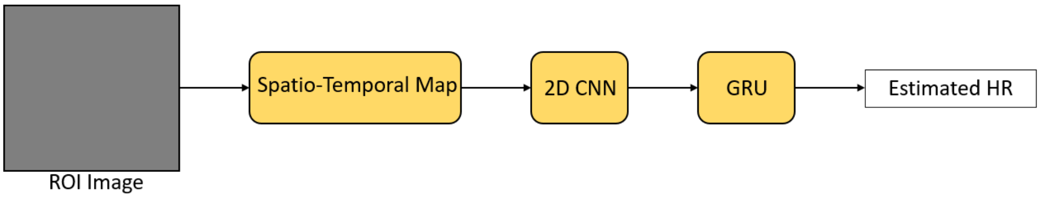

Figure 19. Architecture of RhythmNet [68].

HR estimation can be treated as a regression problem by using simple fully-connected layers or feedforward neural networks. In References [45,55,56,58], HR was regressed by fully-connected layers from the extracted rPPG signal. The architecture used for HR estimation in Reference [56] is shown in Figure 20. In References [50,52], feedforward neural networks were also utilized to estimate HR from the extracted features.

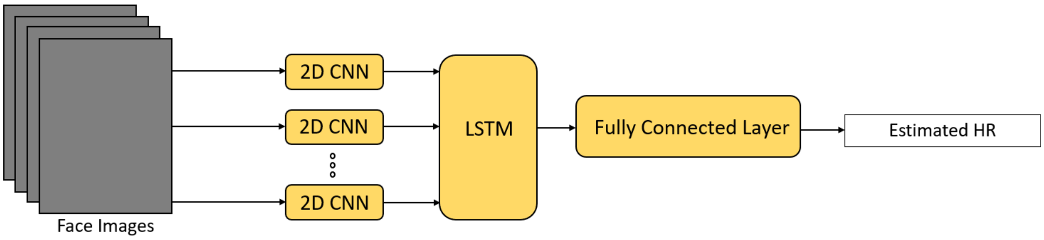

Figure 20. Architecture used in Huang et al. [56].

4. Applications

With further research and inevitable technological advances, remote health monitoring technology will undoubtedly play a vital role in many aspects. The utilization of contactless HR monitoring introduces benefits that existing contact-based PPG methods lack. In this section, we describe a few potential applications enabled by remote monitoring of physiological signals.

4.1. Affective Computing

Since rPPG technology can be integrated with consumer-level cameras, it has great potential for affective computing and human–computer interaction applications. Researchers have demonstrated the feasibility of rPPG-based methods for interpreting human affects, such as cognitive stress estimation [72,73], emotion recognition [29,74], engagement detection [75], and pain recognition [76–78]. These studies illustrate the capability of using rPPG technology beyond the medical domain.

4.2. Pandemic Control

With the current COVID-19 outbreak, the value of contactless HR monitoring has become very clear, particularly for screening the public. It has been reported that temperature screening alone is an insufficient indication for coronavirus infection [79,80]. Therefore, the accuracy of screening based on temperature decreases because asymptomatic infected patients have a temperature within the normal range [81,82]. Given this inadequacy, using HR as a criterion for COVID-19 screening was investigated. In References [83,84], it was shown that tachycardia(high HR) is also a symptom of COVID-19. Moreover, the relationship between atrial fibrillation (AF) and COVID-19 was observed in several studies [85–87], suggesting rPPG-based AF detection would be useful for discovering potential COVID-19 patients [88]. Meanwhile, during and since the pandemic, the use of wearable smart devices for measuring vital signs, such as HR, BP, and SpO2, have become widespread [89,90]. Such contact-based methods can be replaced by rPPG technology to provide convenience to users with precise screening and detection, resulting in more efficient and effective pandemic control.

4.3. Deepfake Detection

Recently, deepfake, a technology to produce high-synthetic videos with DL implementations, has attracted researchers' attention. Unfortunately, if not tragically, this technology has been used to generate fake news and hoax videos, posing threats to the society. For example, a high-quality video of the 44th President of the United States Barack Obama has been synthesized by using a DL approach [91], which shows him apparently making a speech that he never actually made. These fake videos are of such high quality such that they are indistinguishable to humans and even complicated computer vision algorithms [92–94]. As a result, deepfake detection methods need to be developed to encounter such problems. Currently, there have been few attempts in capturing abnormalities in biological signals, such as HR, as a means to detect deepfake videos [95,96].

4.4. Telehealth

Within the last few years, telehealth has become more popular all over the world, with more than half of the health care organizations in the U.S. making use of the service [97]. The integration of telehealth with rPPG technology provides various benefits to the users and society. For instance, users will experience a better daily workflow since the time required to travel to healthcare institutions for health-checkups and for doctor consultations will be reduced. Furthermore, the application of telehealth software intertwined with rPPG technology allows the user to measure their physiological signs and detect early symptoms of different health problems (e.g., atrial fibrillation) from any location by using a consumer-level device [88,98,99]. Deploying rPPG technology promotes social distancing and safety for those healthcare workers at the front lines. Furthermore, remote health monitoring can reduce the workload of hospitals and minimize the chance of spreading diseases since it encourages less physical contact with patients and fewer human resources are needed [100], which is especially vital during the midst of a pandemic.

4.5. Face Anti-Spoofing

Today, using biometric information of individuals for authentication is very common. One of the most common forms is facial recognition, which is based on the analysis of unique features of a person's face [101]. However, biometric presentation attacks can exist alongside the face authentication process. For example, attackers can source photos (photo attacks) or videos (replay attacks) of the person from social networking sites easily and present them to the authentication system [102]. Remote HR measurement technology can be incorporated to enhance the authentication system [103]. In References [104–108], rPPG-based face presentation attack detection approaches were developed, suggesting the potential of rPPG technology in the security industry.

4.6. Driving Condition Monitoring

In order to reduce the number of traffic accidents, rPPG technology can be adopted to monitor drivers and track a driver's physiological status. Most road accidents are caused by human factors including fatigue, drowsiness, and illness. Factors, such as disparity in oxygen levels, HR, and RR, may lead to non-specific health problems, which interfere with or degrade decision-making capabilities. This monitoring allows abnormal vital signs to be detected early, with alerts shown immediately so that drivers can adjust their behavior accordingly, avoiding accidents. There have been several attempts for monitoring drivers' physiological conditions using rPPG methods [109–118]. In References [112,116,118], a near-infrared camera was used instead of an RGB camera for monitoring. In References [113,115], neural networks were applied for physiological signal estimation.

4.7. Searching for Survivors during Natural Disasters

During natural disasters, such as earthquakes and fires, searching for survivors becomes a vital but extremely challenging task. Rescue teams must operate in extremely hazardous conditions, such as collapsed buildings. rPPG technology can be a potential way to reduce risk for search and rescue teams, and improve their efficiency. In References [119–121], an unmanned aerial vehicle (UAV) or a drone was used to capture videos, representing a more convenient, safe, and effective way to look for survivors. In Reference [122], research using a drone for multiple subject detection over a long-distance was conducted. This illustrates the potential of using controllable devices equipped with a camera combined with rPPG technology for searching for survivors.

4.8. Neonatal Monitoring

As neonates or infants have very sensitive and fragile skin, using contact-based methods to measure their health conditions is inappropriate. rPPG methods are one of the suitable candidates for long-term physiological status monitoring of newborns in neonatal intensive care units (NICU). Several studies have trialed rPPG methods for such monitoring [9,123–132]. In Reference [131], DL-based segmentation was utilized to reduce computational time, which brought it one step closer to real-time applications. In Reference [132], DL-based ROI detection was applied to handle pose variations and illumination changes, further improving the estimation accuracy. These examples indicate the promise of using rPPG technology for neonatal monitoring.

4.9. Fitness Tracking

During fitness training, having health monitors to keep track of the current physiological condition is an excellent way to prevent over-exercising and help to adjust the fitness process to an individual's real-time needs and condition. Contact-based methods, such as smartwatches or digital bracelets, for such monitoring can cause discomfort or pain during heavy exercise. rPPG technology can be utilized to provide simple remote fitness tracking. In References [20,133–136], rPPG methods in fitness training settings were studied. In References [20,133,136], motion artifact during exercise was the major focus. In References [134,135], a feedback control system was implemented, as well, for adjusting the speed of the treadmill automatically.

5. Resources

As remote physiological monitoring is an emerging field in computer vision and biomedical engineering, there are resources available for researchers to accelerate progress and ease the transition of newcomers. In this section, we detail some of the open-source toolboxes to help to implement related algorithms and most of the datasets that are commonly used for model training and benchmarking. Furthermore, open challenges in rPPG are also described to encourage various researchers to contribute to the field.

5.1. Toolboxes

iPhys [137] is an open-source toolbox written in MATLAB. It contains commonly used implementations in rPPG pipelines, such as face detection, ROI definition, and skin segmentation. It also includes four conventional rPPG methods for baseline comparison. Other plotting and signal quality calculation functions are provided, as well, for performance evaluation.

In Reference [138], the whole rPPG pipeline based on ICA and some utilities are written in MATLAB. Beginners can quickly run or even modify the provided script to evaluate the performance of the particular rPPG method.

The Python tool for Virtual Heart Rate (pyVHR) [139] is a recently developed Python package for heart rate estimation based on rPPG methods. In this package, 8 conventional rPPG methods are implemented and evaluated based on 5 datasets. Other frequently used pre-processing and post-processing techniques are provided, as well. Practitioners can extend the framework to evaluate their own algorithms on these 5 datasets.

5.2. Datasets

Performance evaluation is important for researchers to test whether their proposed methods are good enough when compared with other methods and able to solve existing challenges. For supervised methods, datasets are also crucial for proper training and achieving state-of-the-art performance. In this subsection, we detail most of the datasets that are commonly used for benchmarking and model training.

AFRL [140] is a dataset proposed by the United States Air Force Research Laboratory. It was aimed to evaluate the effect of head motion artifacts. During data acquisition, a multi-imager semicircular array (a total of 9 synchronized, visible spectrum imagers) centered on the imaged participant in a controlled light environment was used to record the participant's head motions during specific tasks. At the same time, electrocardiogram (ECG) and fingertip reflectance PPG were recorded as ground truth signals. The imaged participant was told to perform specific tasks, which included staying still, sweeping around the imagers with a pre-defined angle per second, and randomly re-orienting the head position to an imager. The background of the environment consisted of either a solid black fabric or a patterned, colored fabric.

COHFACE [141] is a publicly available dataset proposed by the Idiap Research Institute. The purpose of proposing this dataset was to allow researchers to evaluate their developed rPPG algorithms on a publicly available dataset so that comparisons between different algorithms could be conducted in a standard and principled manner. In this dataset, a conventional webcam was used to capture the full face of the participant in two different illumination settings (studio lighting and natural lighting) to evaluate the effect of illumination variation. Skin reflectance PPG and respiratory signal were recorded as ground truth signals. The only disadvantage of this dataset was the heavy compression, so noise artifact was unavoidably added.

MAHNOB-HCI [142] is a multimodal dataset that was originally recorded for emotion recognition and implicit tagging research. However, as ground truth signals, such as ECG and respiration amplitude, were recorded, it was also suitable for rPPG algorithm evaluation. Moreover, six cameras were used to capture different views (frontal view, profile view, wide angle, close ups) of the participant, which made this dataset useful for evaluating the algorithm when pose angle varied.

MMSE-HR [143] is another multimodal dataset proposed for facial expression analysis. Some vital signs, such as BP, RR, and HR, were also recorded, making this dataset appropriate for testing rPPG algorithms. Furthermore, subjects from different races (Black, White, Asian, Hispanic/Latino) participated in the data acquisition, so researchers were able to evaluate their proposed methods against different skin tones.

OBF [41] is a large dataset made by the University of Oulu in Finland specifically for remote physiological signal measurement. Aside from healthy subjects in this dataset, patients with atrial fibrillation (AF) also participated in data collection in order to validate rPPG methods for clinical applications, such as diagnosing cardiac diseases. In addition, there were two different recording states, one for healthy participants and one for AF patients. Healthy participants were recorded in a resting state and a post-exercise state (5-min exercise). AF patients were recorded before and after cardioversion treatment.

PURE [144] is a dataset proposed for examining head motion artifacts in rPPG methods in more detail. During data acquisition, participants were told to perform six different tasks (holding steady, talking, slow translation, fast translation, small rotation, medium rotation) in order to introduce different kinds of head motion. Naturally changing illumination (daylight with clouds through a large window) was used for recording, as well.

UBFC-RPPG [145] is another dataset proposed mainly for rPPG algorithm evaluation. The data recording was conducted indoors with indoor illumination and slight changes in sunlight. One special aspect of the recording is that participants were told to play a time-sensitive mathematical game. Its purpose was to augment the HR of participants and hence simulate a real-life human-computer interaction scenario for evaluation.

VIPL-HR [146] is a large-scale multimodal dataset created for remote pulse estimation research. In this dataset, various face variations due to head motion (stable, large motion, talking), illumination changes (lab, dark, bright), and acquisition diversity (smartphone, webcam, RGB-D camera) were introduced in order to test the overall robustness of the proposed algorithm. The dataset was compressed with different codecs (MJPG, FMP4, DIVX, PIM1, X264) in order to retain the completeness of the signals as much as possible, while being convenient for public access at the same time.

A summary of the mentioned datasets is provided in Table 5. Moreover, Table 6 illustrates the performance of all mentioned DL methods on these common datasets. The evaluation metrics in Table 6 include root mean square error (RMSE) in bpm, mean absolute error (MAE) in bpm, Pearson correlation coefficient (R), and signal-to-noise ratio (SNR) in decibels (dB).

Table 5. Summary of common datasets for remote physiological monitoring.

Dataset	Subjects	Description
AFRL [140]	25	9 RGB cameras with 120 fps, resolution is 658 × 492, ECG, PPG, RR are recorded
COHFACE [141]	40	1 RGB webcam with 20 fps, resolution is 640 × 480, BVP, RR are recorded
MAHNOB-HCI [142]	27	1 RGB camera with 60 fps, 5 monochrome cameras with 60 fps, both resolution are 780 × 580, ECG, RR are recorded
MMSE-HR [143]	140	1 3D stereo imaging sensor with 25 fps, 1 2D video sensor with 25 fps, 1 thermal sensor with 25 fps, RGB sensor resolution is 1040 × 1392, thermal sensor resolution is 640 × 480, HR, RR, BP are recorded
OBF [41]	106 (6 with atrial fibrillation)	1 RGB camera with 60 fps, 1 NIR camera with 30 fps, RGB camera resolution is 1920 × 1080, NIR camera resolution is 640 × 480, ECG, BVP, RR are recorded
PURE [144]	10	1 RGB camera with 30 fps, resolution is 640 × 480, HR, SpO2, PPG are recorded
UBFC-RPPG [145]	42	1 RGB webcam with 30 fps, resolution is 640 × 480, HR, PPG are recorded
VIPL-HR [146]	107	1 RGB webcam with 25 fps, 1 RGB-D camera with 30 fps, 1 smartphone camera with 30 fps, RGB webcam resolution is 960 × 720, RGB-D NIR camera resolution is 640 × 480, RGB-D RGB camera resolution is 1920 × 1080, smartphone camera resolution is 1920 × 1080, HR, SpO2, BVP are recorded

Table 6. Performance of all mentioned end-to-end and hybrid DL methods for HR measurement on commonly used datasets listed in Table 5. Refs. [43,44,56] are not included here as they are evaluated on their own private datasets.

Methods	AFRL	COHFACE	MAHNOB-HCI	MMSE-HR	OBF	PURE	UBFC-RPPG	VIPL-HR
[26]	X	RMSE = 10.78 MAE = 8.10 R = 0.29	RMSE = 9.24 MAE = 7.25 R = 0.51	X	X	RMSE = 2.37 MAE = 1.84 R = 0.98	X	X
[27]	MAE = 2.45 SNR = 4.65	X	MAE = 4.57 SNR = −8.98	X	X	X	X	X
[64]	X	X	RMSE = 4.26 R = 0.81	X	X	X	X	X
[66]	X	X	RMSE = 4.49	RMSE = 6.83	X	X	X	X
[29]	X	X	RMSE = 7.88 MAE = 5.96 R = 0.76	X	RMSE = 1.812 R = 0.992	X	X	X
[30]	X	X	RMSE = 5.93 MAE = 4.03 R = 0.88	X	RMSE = 1.8 R = 0.992	X	X	X
[35]	X	RMSE = 11.88 MAE = 7.31 R = 0.36 SNR = −1.93	X	X	X	RMSE = 1.58 MAE = 0.88 R = 0.99 SNR = 9.18	X X X X	X X X X
[47]	X	X	X	RMSE = 3.187 MAE = 4.35 R = 0.8254	X	X	X	X
[50]	X	X	X	X	X	X	RMSE = 8.64 MAE = 5.45	X X
[55]	X	RMSE = 9.96 MAE = 8.09 R = 0.40	X	X	X	X	X	X
[67]	X	X	X	RMSE = 10.10 R = 0.64	X	X	X	RMSE = 7.99 MAE = 5.40 R = 0.66

Table 6. *Cont.*

Methods	AFRL	COHFACE	MAHNOB-HCI	MMSE-HR	OBF	PURE	UBFC-RPPG	VIPL-HR
[28]	RMSE = 3.72 **MAE = 1.45** R = 0.94 **SNR = 8.64**	X	X	RMSE = 5.66 MAE = 3.00 **R = 0.92** SNR = 2.37	X	X	X	X
[31]	X	X	RMSE = 5.10 MAE = 3.78 R = 0.86	RMSE = 5.87 R = 0.89	X	X	X	RMSE = 8.68 MAE = 5.68 R = 0.72
[45]	X	X	RMSE = 3.41 R = 0.92	X	X	X	X	X
[48]	X	X	X	**MAE = 1.31** **SNR = 9.44**	X X	X X	X X	X X
[51]	X	**RMSE = 1.29** MAE = 0.70 R = 0.73	X	X	X	RMSE = 1.56 MAE = 0.51 R = 0.83	**RMSE = 0.97** MAE = 0.48	X
[52]	X	X	X	X	X	X	RMSE = 3.368 MAE = 2.412 **R = 0.983**	X
[54]	X	RMSE = 7.06 MAE = 3.07 **R = 0.86**	RMSE = 6.26 MAE = 4.81 R = 0.79	X	X	**RMSE = 0.43** **MAE = 0.28** R = 0.999	X	X
[57]	X	X	RMSE = 3.68 MAE = 3.01 R = 0.85	X	X	X	RMSE = 7.42 MAE = 5.97 R = 0.53	X
[68]	X	X	RMSE = 3.99 R = 0.87	RMSE = 5.49 R = 0.84	X	X	X	RMSE = 8.14 MAE = 5.30 R = 0.76
[69]	X	X	X	RMSE = 6.04 R = 0.84	**RMSE = 1.26** **R = 0.996**	X	X	**RMSE = 7.97** **MAE = 5.02** **R = 0.796**
[70]	X	X	**RMSE = 3.23** **MAE = 1.53** **R = 0.97**	X	X	X	X	X

Table 6. *Cont.*

Methods	AFRL	COHFACE	MAHNOB-HCI	MMSE-HR	OBF	PURE	UBFC-RPPG	VIPL-HR
[32]	X	RMSE = 7.52 MAE = 5.19 R = 0.68	X	X	X	RMSE = 1.21 MAE = 0.74 **R = 1.00**	X	X
[33]	X	RMSE = 9.50 MAE = 5.57 R = 0.75	X	X	X	X	RMSE = 3.82 MAE = 2.15 R = 0.97	X
[34]	X	RMSE = 6.65 MAE = 4.67 R = 0.77	X	RMSE = 5.84 R = 0.85	X	RMSE = 0.77 MAE = 0.34 R = 0.99	RMSE = 3.97 MAE = 1.46 R = 0.93	X
[49]	X	X	X	**RMSE = 3.12** MAE = 1.87 R = 0.89	X	X	RMSE = 3.12 MAE = 2.46 R = 0.96	X
[53]	X	RMSE = 1.65 **MAE = 0.68** R = 0.72	X	X	X	RMSE = 1.07 MAE = 0.40 R = 0.92	RMSE = 2.09 **MAE = 0.47**	X
[58]	X	X	RMSE = 6.42 MAE = 5.01	X	X	X	RMSE = 7.24 MAE = 5.29	X
[59]	X	X	RMSE = 6.53 MAE = 4.15 R = 0.71	X	X	RMSE = 4.29 MAE = 2.28 R = 0.99	RMSE = 2.10 MAE = 1.19 R = 0.98	X
[71]	X	X	X	X	X	RMSE = 2.02 MAE = 1.65 R = 0.99	X	RMSE = 8.01 MAE = 5.12 R = 0.79

5.3. Open Challenge on Remote Physiological Signal Sensing

Creating an open challenge on a specific machine learning task is a common way in the field of machine learning to encourage people to participate and solve a particular problem using DL methods. One of the most famous open challenges is the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [147]. This challenge has been running annually for 8 years (2010–2017), and its focuses are object recognition, object detection, and image classification. Many DL methods have been proposed for this task, and this competition has definitely boosted research interest in this field, allowing rapid development in DL-based computer vision. An open challenge on remote physiological signal sensing was also organized in 2020, namely Remote Physiological Signal Sensing (RePSS 2020) [148]. In this challenge, the focus was measuring the average HR from color facial videos. The VIPL-HR-V2 dataset, which is the second version of VIPL-HR [146], and the OBF dataset [41] were used for model training and testing. The RePSS 2021 is also currently running, and its focus was changed to measure inter-beat-interval (IBI) curve and RR. This open challenge can have the same optimistic effect as ILSVRC, encouraging people to participate and engage in this research field.

6. Research Gaps

During the last few decades, many methods for ascertaining remote HR measurements have been proposed. This has attracted much attention, and increasing numbers of researchers are engaged in this exciting area. In this section, we discuss some of the research gaps in order to suggest some possible future directions in this field.

6.1. Influencing Factors

The performance of remote HR measurement based on rPPG is influenced by many factors, such as illumination changes, motion artifacts, skin-tone variations, and video compression [12,14,22,23,149]. There are several methods proposed for handling these challenges. For example, the utilization of different HR signal representations, such as spectrum images [64] and spatio-temporal maps [66–71], as well as the use of attention mechanism [27,28,30,32,34,35,49,52,67], can deal with illumination variations and motion noise. STVEN [30] was designed to improve the robustness of HR measurement under video compression. Meta-learning approaches with fast adaptation to uncommon samples [49,57] are suitable to deal with skin-tone variations. Additional work is needed to better understand and quantify the effects these influencing factors have on remote physiological measurement. More importantly, new methods should provide insight into how these challenges are handled from a technical and biophysical perspective, rather than just evaluating their performance on a dataset that contains the influencing factors.

6.2. Measuring Other Vital Signs

Undoubtedly, HR is a very important physiological indicator to indicate the current health condition of a person. Researchers in this field are mainly interested in estimating HR, followed by RR. However, other vital signs are also important [150,151]. For example, BP is useful in detecting some cardiovascular diseases, such as hypertension, while SpO2 can reflect the health of the cardiorespiratory system by showing if a person has an adequate supply of oxygen. At the same time, these vital signs are associated to COVID-19, and they are useful for COVID-19 diagnosing, as well [152,153].There are relatively fewer studies that attempt to estimate BP [154–156] and SpO2 [157–159] remotely when compared HR and RR. There are still many research opportunities in other vital signs.

6.3. Datasets

Datasets are increasingly important for evaluating new proposed methods whether demonstrating success in addressing specific problems or increasing the effectiveness of previously proposed methods. For DL methods, datasets are even more important as they are used for training in supervised methods, as well. The performance of the supervised methods are greatly affected by the training datasets. Currently, most of the existing publicly available datasets focus on two major challenges in rPPG methods only, that is, motion artifacts and illumination variations [12,23]. Other challenges, such as skin-tone variations [22,160,161], multiple persons detection [14,122], and long distance estimations [14,162], need to be overcome, as well, if the methods are to be, ultimately, robust and highly applicable in the real-world, replacing all contact-based methods. Moreover, the subjects in these datasets are mainly adult participants. Datasets with newborns as participants are also desirable for evaluating rPPG methods. As a result, more comprehensive, high diversity and high quality datasets are needed to fully evaluate the robustness of any new proposed method and allow comprehensive training in supervised methods. Such datasets are extremely beneficial to the research community.

6.4. Performance on Different Heart Rate Ranges

According to performance results of RePSS 2020 [148], the top 3 teams were able to achieve a significantly better performance on the middle HR level, where HR ranges from 77 to 90 bpm, followed by the low HR level (less than 70 bpm). Performance at the high HR level (more than 90 bpm) was the worst. This is a challenge that absolutely needs to be addressed in order to be accurate enough to be applied in real-world applications because these significantly lower or higher HRs are showing specific health problems. Moreover, this result indicates that using common metrics, such as mean absolute error (MAE), root mean square error (RMSE), signal-to-noise ratio (SNR), and Pearson correlation coefficient (R), to evaluate rPPG methods may not be effective enough. Evaluation on a wider range of HR levels is required in order to comprehensively test the robustness of the proposed method.

6.5. Understanding of Deep Learning-Based Methods

The advantage of using CNN in rPPG technology is that a good result can be obtained without very deep understanding or analysis of the specific problem; the disadvantage is that this DL method is a black box, which means we do not have a full understanding of why such a result is obtained. The lack of understanding of how CNN-based methods work on rPPG technology may be a barrier to further development of this technology and evaluation of these DL methods. Reference [25] is a work that focused on the understanding of CNN-based rPPG methods, rather than proposing a new model with state-of-the-art performance. Several experiments were performed to explore the CNN-based rPPG signal extraction and improve the understanding of this approach. In the paper, some important observations have been made. For example, it showed that the CNN for rPPG signal extraction is actually learning information related to the PPG signal but not the motion-induced intensity changes [27]. In addition, the CNN training is affected by the physiological delay between the video data and the reference finger oximeter. Researchers should direct their attention to more studies that focus on the understanding of DL-based rPPG methods in order to gain valuable insights and further improve the performance of these DL approaches.

7. Conclusions

In recent years, many methods for remote HR measurement have been proposed. Due to rapid development in the area of machine learning, DL methods have shown significant promise in this field. In this paper, we have provided a comprehensive review on most of the existing recent DL-based methods for remote HR estimation. We have further categorized these methods into end-to-end and hybrid DL methods, and grouped them based on the type of neural network being used. We then described some potential applications that can be achieved by using rPPG technology. Next, some rPPG resources, such as toolboxes, datasets, and open challenges, have been detailed in order to help accelerate research. Lastly, we have discussed some of the current research gaps in this field to shed some light on future areas and directions in this exciting field.

As remote physiological measurement establishes itself as an emerging research field, we suggest more work should focus on addressing different influencing factors and estimating other vital signs, which will assist in bridging the gap for real-world applications. Furthermore, high-quality and diverse datasets are crucial for proper benchmarking and analysis of different methods and the future development of more complex DL models and architectures. Last but not least, the understanding of different DL-based approaches is critical, especially when integrating these networks for high-stakes applications, such as healthcare diagnostics.

Author Contributions: C.-H.C. and K.-L.W. performed the literature review and prepared the draft manuscript. J.-W.C., T.-T.C. and R.H.Y.S. reviewed and edited the manuscript. All authors have read and agreed to the published version of the manuscript.

Funding: This research is partially funded by the ITC TSSSU program and HKUST UROP.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: We are grateful for the support from the Innovation and Technology Commission of Hong Kong and Undergraduate Research Opportunities Program of the Hong Kong University of Science and Technology.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Jeong, I.; Finkelstein, J. Introducing Contactless Blood Pressure Assessment Using a High Speed Video Camera. *J. Med. Syst.* **2016**, *40*, 1–10. [CrossRef]
2. Bal, U. Non-contact estimation of heart rate and oxygen saturation using ambient light. *Biomed. Opt. Express* **2015**, *6*, 86–97. [CrossRef] [PubMed]
3. Massaroni, C.; Nicolò, A.; Sacchetti, M.; Schena, E. Contactless Methods For Measuring Respiratory Rate: A Review. *IEEE Sens. J.* **2021**, *21*, 12821–12839. [CrossRef]
4. Iozzia, L.; Cerina, L.; Mainardi, L. Relationships between heart-rate variability and pulse-rate variability obtained from video-PPG signal using ZCA. *Physiol. Meas.* **2016**, *37*, 1934–1944. [CrossRef] [PubMed]
5. Scalise, L. Non contact heart monitoring. *Adv. Electrocardiogr. Methods Anal.* **2012**, *4*, 81–106.
6. Shao, D.; Liu, C.; Tsow, F. Noncontact Physiological Measurement Using a Camera: A Technical Review and Future Directions. *ACS Sens.* **2021**, *6*, 321–334. [CrossRef] [PubMed]
7. Sun, Y.; Thakor, N. Photoplethysmography Revisited: From Contact to Noncontact, From Point to Imaging. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 463–477. [CrossRef] [PubMed]
8. Swinehart, D.F. The Beer-Lambert Law. *J. Chem. Educ.* **1962**, *39*, 333,
9. Aarts, L.; Jeanne, V.; Cleary, J.P.; Lieber, C.; Nelson, J.; Bambang-Oetomo, S.; Verkruysse, W. Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit—A pilot study. *Early Hum. Dev.* **2013**, *89*. [CrossRef]
10. Rouast, P.; Adam, M.; Chiong, R.; Cornforth, D.; Lux, E. Remote heart rate measurement using low-cost RGB face video: A technical literature review. *Front. Comput. Sci.* **2018**, *12*, 858–872. [CrossRef]
11. Maurya, L.; Kaur, P.; Chawla, D.; Mahapatra, P. Non-contact breathing rate monitoring in newborns: A review. *Comput. Biol. Med.* **2021**, *132*, 104321. [CrossRef]
12. McDuff, D.J.; Estepp, J.R.; Piasecki, A.M.; Blackford, E.B. A survey of remote optical photoplethysmographic imaging methods. In Proceedings of the 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Milan, Italy, 25–29 August 2015; pp. 6398–6404. [CrossRef]
13. Wang, W.; den Brinker, A.C.; Stuijk, S.; de Haan, G. Algorithmic Principles of Remote PPG. *IEEE Trans. Biomed. Eng.* **2017**, *64*, 1479–1491. [CrossRef]
14. Khanam, F.T.Z.; Al-Naji, A.; Chahl, J. Remote Monitoring of Vital Signs in Diverse Non-Clinical and Clinical Scenarios Using Computer Vision Systems: A Review. *Appl. Sci.* **2019**, *9*, 4474. [CrossRef]
15. Verkruysse, W.; Svaasand, L.O.; Nelson, J.S. Remote plethysmographic imaging using ambient light. *Opt. Express* **2008**, *16*, 21434–21445. [CrossRef]
16. Poh, M.Z.; McDuff, D.J.; Picard, R.W. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Opt. Express* **2010**, *18*, 10762–10774. [CrossRef]

17. Lewandowska, M.; Rumiński, J.; Kocejko, T.; Nowak, J. Measuring pulse rate with a webcam—A non-contact method for evaluating cardiac activity. In Proceedings of the 2011 Federated Conference on Computer Science and Information Systems (FedCSIS), Szczecin, Poland, 19–21 September 2011; pp. 405–410.
18. de Haan, G.; Jeanne, V. Robust Pulse Rate From Chrominance-Based rPPG. *IEEE Trans. Biomed. Eng.* **2013**, *60*, 2878–2886. [CrossRef]
19. Tominaga, S. Dichromatic reflection models for a variety of materials. *Color Res. Appl.* **1994**, *19*, 277–285. [CrossRef]
20. de Haan, G.; van Leest, A. Improved motion robustness of remote-PPG by using the blood volume pulse signature. *Physiol. Meas.* **2014**, *35*, 1913–1926. [CrossRef] [PubMed]
21. Wang, W.; Stuijk, S.; de Haan, G. A Novel Algorithm for Remote Photoplethysmography: Spatial Subspace Rotation. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 1974–1984. [CrossRef] [PubMed]
22. Dasari, A.; Arul Prakash, S.K.; Jeni, L.; Tucker, C. Evaluation of biases in remote photoplethysmography methods. *Npj Digit. Med.* **2021**, *4*. [CrossRef]
23. Chen, X.; Cheng, J.; Song, R.; Liu, Y.; Ward, R.; Wang, Z.J. Video-Based Heart Rate Measurement: Recent Advances and Future Prospects. *IEEE Trans. Instrum. Meas.* **2019**, *68*, 3600–3615. [CrossRef]
24. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, Kauai, HI, USA, 8–14 December 2001; Volume 1, p. I. [CrossRef]
25. Zhan, Q.; Wang, W.; de Haan, G. Analysis of CNN-based remote-PPG to understand limitations and sensitivities. *Biomed. Opt. Express* **2020**, *11*, 1268–1283. [CrossRef] [PubMed]
26. Spetlik, R.; Franc, V.; Cech, J.; Matas, J. Visual Heart Rate Estimation with Convolutional Neural Network. In Proceedings of the British Machine Vision Conference (BMVC), Newcastle, UK, 2–6 September 2018.
27. Chen, W.; McDuff, D. DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
28. Liu, X.; Fromm, J.; Patel, S.; McDuff, D. Multi-Task Temporal Shift Attention Networks for On-Device Contactless Vitals Measurement. *arXiv* **2021**, arXiv:2006.03790v2.
29. Yu, Z.; Li, X.; Zhao, G. Remote Photoplethysmograph Signal Measurement from Facial Videos Using Spatio-Temporal Networks. In Proceedings of the British Machine Vision Conference (BMVC), Cardiff, UK, 9–12 September 2019.
30. Yu, Z.; Peng, W.; Li, X.; Hong, X.; Zhao, G. Remote Heart Rate Measurement From Highly Compressed Facial Videos: An End-to-End Deep Learning Solution with Video Enhancement. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27–28 October 2019; pp. 151–160. [CrossRef]
31. Yu, Z.; Li, X.; Niu, X.; Shi, J.; Zhao, G. AutoHR: A Strong End-to-End Baseline for Remote Heart Rate Measurement with Neural Searching. *IEEE Signal Process. Lett.* **2020**, *27*, 1245–1249. [CrossRef]
32. Hu, M.; Qian, F.; Wang, X.; He, L.; Guo, D.; Ren, F. Robust Heart Rate Estimation with Spatial-Temporal Attention Network from Facial Videos. *IEEE Trans. Cogn. Dev. Syst.* **2021**, 1. [CrossRef]
33. Zhang, P.; Li, B.; Peng, J.; Jiang, W. Multi-hierarchical Convolutional Network for Efficient Remote Photoplethysmograph Signal and Heart Rate Estimation from Face Video Clips. *arXiv* **2021**, arXiv:2104.02260
34. Hu, M.; Qian, F.; Guo, D.; Wang, X.; He, L.; Ren, F. ETA-rPPGNet: Effective Time-Domain Attention Network for Remote Heart Rate Measurement. *IEEE Trans. Instrum. Meas.* **2021**, *70*, 1–12. [CrossRef]
35. Hu, M.; Guo, D.; Wang, X.; Ge, P.; Chu, Q. A Novel Spatial-Temporal Convolutional Neural Network for Remote Photoplethysmography. In Proceedings of the 2019 12th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Suzhou, China, 19–21 October 2019; pp. 1–6. [CrossRef]
36. Lin, J.; Gan, C.; Han, S. TSM: Temporal Shift Module for Efficient Video Understanding. In Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019.
37. Liu, H.; Simonyan, K.; Yang, Y. DARTS: Differentiable Architecture Search. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
38. Xu, Y.; Xie, L.; Zhang, X.; Chen, X.; Qi, G.J.; Tian, Q.; Xiong, H. PC-DARTS: Partial Channel Connections for Memory-Efficient Architecture Search. In Proceedings of the International Conference on Learning Representations, New Orleans, LA, USA, 6–9 May 2019.
39. Qiu, Z.; Yao, T.; Mei, T. Learning Spatio-Temporal Representation with Pseudo-3D Residual Networks. In Proceedings of the 2017 IEEE International Conference on Computer Vision (ICCV), Venice, Italy, 22–29 October 2017; pp. 5534–5542. [CrossRef]
40. Shi, X.; Chen, Z.; Wang, H.; Yeung, D.Y.; Wong, W.K.; Woo, W.C. Convolutional LSTM Network: A Machine Learning Approach for Precipitation Nowcasting. In *Proceedings of the 28th International Conference on Neural Information Processing Systems—Volume 1*; NIPS'15; MIT Press: Cambridge, MA, USA, 2015; pp. 802–810.
41. Li, X.; Alikhani, I.; Shi, J.; Seppanen, T.; Junttila, J.; Majamaa-Voltti, K.; Tulppo, M.; Zhao, G. The OBF Database: A Large Face Video Database for Remote Physiological Signal Measurement and Atrial Fibrillation Detection. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 242–249. [CrossRef]

42. Lempe, G.; Zaunseder, S.; Wirthgen, T.; Zipser, S.; Malberg, H. ROI Selection for Remote Photoplethysmography. In *Bildverarbeitung für die Medizin 2013*; Meinzer, H.P., Deserno, T.M., Handels, H., Tolxdorff, T., Eds.; Springer: Berlin/Heidelberg, Germany, 2013; pp. 99–103.
43. Tang, C.; Lu, J.; Liu, J. Non-contact Heart Rate Monitoring by Combining Convolutional Neural Network Skin Detection and Remote Photoplethysmography via a Low-Cost Camera. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 1309–1315. [CrossRef]
44. Paracchini, M.; Marcon, M.; Villa, F.; Zappa, F.; Tubaro, S. Biometric Signals Estimation Using Single Photon Camera and Deep Learning. *Sensors* **2020**, *20*, 6102. [CrossRef]
45. Sabokrou, M.; Pourreza, M.; Li, X.; Fathy, M.; Zhao, G. Deep-HR: Fast Heart Rate Estimation from Face Video Under Realistic Conditions. *arXiv* **2020**, arXiv:2002.04821.
46. Liu, S.; Huang, D.; Wang, Y. Receptive Field Block Net for Accurate and Fast Object Detection. In *Computer Vision—ECCV 2018*; Ferrari, V., Hebert, M., Sminchisescu, C., Weiss, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 404–419.
47. Bian, M.; Peng, B.; Wang, W.; Dong, J. An Accurate LSTM Based Video Heart Rate Estimation Method. In *Pattern Recognition and Computer Vision*; Lin, Z., Wang, L., Yang, J., Shi, G., Tan, T., Zheng, N., Chen, X., Zhang, Y., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 409–417.
48. Botina-Monsalve, D.; Benezeth, Y.; Macwan, R.; Pierrart, P.; Parra, F.; Nakamura, K.; Gomez, R.; Miteran, J. Long Short-Term Memory Deep-Filter in Remote Photoplethysmography. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 1242–1249. [CrossRef]
49. Liu, X.; Jiang, Z.; Fromm, J.; Xu, X.; Patel, S.; McDuff, D. MetaPhys: Few-Shot Adaptation for Non-Contact Physiological Measurement. In *Proceedings of the Conference on Health, Inference, and Learning*; Association for Computing Machinery: New York, NY, USA, 2021; pp. 154–163.
50. Bousefsaf, F.; Pruski, A.; Maaoui, C. 3D Convolutional Neural Networks for Remote Pulse Rate Measurement and Mapping from Facial Video. *Appl. Sci.* **2019**, *9*, 4364. [CrossRef]
51. Tsou, Y.Y.; Lee, Y.A.; Hsu, C.T.; Chang, S.H. Siamese-RPPG Network: Remote Photoplethysmography Signal Estimation from Face Videos. In *Proceedings of the 35th Annual ACM Symposium on Applied Computing*; SAC'20; Association for Computing Machinery: New York, NY, USA, 2020; pp. 2066–2073. [CrossRef]
52. Perepelkina, O.; Artemyev, M.; Churikova, M.; Grinenko, M. HeartTrack: Convolutional neural network for remote video-based heart rate monitoring. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 1163–1171. [CrossRef]
53. Tsou, Y.Y.; Lee, Y.A.; Hsu, C.T. Multi-task Learning for Simultaneous Video Generation and Remote Photoplethysmography Estimation. In *Computer Vision—ACCV 2020*; Ishikawa, H., Liu, C.L., Pajdla, T., Shi, J., Eds.; Springer International Publishing: Cham, Switzerland, 2021; pp. 392–407.
54. Liu, S.Q.; Yuen, P.C. A General Remote Photoplethysmography Estimator with Spatiotemporal Convolutional Network. In Proceedings of the 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020), Buenos Aires, Argentina, 18–22 May 2020; pp. 481–488. [CrossRef]
55. Wang, Z.K.; Kao, Y.; Hsu, C.T. Vision-Based Heart Rate Estimation Via A Two-Stream CNN. In Proceedings of the 2019 IEEE International Conference on Image Processing (ICIP), Taipei, Taiwan, 22–25 September 2019; pp. 3327–3331. [CrossRef]
56. Huang, B.; Chang, C.M.; Lin, C.L.; Chen, W.; Juang, C.F.; Wu, X. Visual Heart Rate Estimation from Facial Video Based on CNN. In Proceedings of the 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA), Kristiansand, Norway, 9–13 November 2020; pp. 1658–1662. [CrossRef]
57. Lee, E.; Chen, E.; Lee, C.Y. Meta-rPPG: Remote Heart Rate Estimation Using a Transductive Meta-learner. In *Computer Vision—ECCV 2020*; Vedaldi, A., Bischof, H., Brox, T., Frahm, J.M., Eds.; Springer International Publishing: Cham, Switzerland, 2020; pp. 392–409.
58. Huang, B.; Lin, C.L.; Chen, W.; Juang, C.F.; Wu, X. A novel one-stage framework for visual pulse rate estimation using deep neural networks. *Biomed. Signal Process. Control* **2021**, *66*, 102387. [CrossRef]
59. Song, R.; Chen, H.; Cheng, J.; Li, C.; Liu, Y.; Chen, X. PulseGAN: Learning to Generate Realistic Pulse Waveforms in Remote Photoplethysmography. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1373–1384. [CrossRef]
60. Finn, C.; Abbeel, P.; Levine, S. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In Proceedings of the 34th International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 1126–1135. [CrossRef]
61. Newell, A.; Yang, K.; Deng, J. Stacked Hourglass Networks for Human Pose Estimation. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 483–499.
62. Hu, S.X.; Moreno, P.G.; Xiao, Y.; Shen, X.; Obozinski, G.; Lawrence, N.; Damianou, A. Empirical Bayes Transductive Meta-Learning with Synthetic Gradients. In Proceedings of the International Conference on Learning Representations, Glasgow, UK, 23–28 August 2020.
63. Mirza, M.; Osindero, S. Conditional Generative Adversarial Nets. *arXiv* **2014**, arXiv:1411.1784
64. Yang, W.; Li, X.; Zhang, B. Heart Rate Estimation from Facial Videos Based on Convolutional Neural Network. In Proceedings of the 2018 International Conference on Network Infrastructure and Digital Content (IC-NIDC), Guiyang, China, 22–24 August 2018; pp. 45–49. [CrossRef]

65. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778. [CrossRef]
66. Niu, X.; Han, H.; Shan, S.; Chen, X. SynRhythm: Learning a Deep Heart Rate Estimator from General to Specific. In Proceedings of the 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, China, 20–24 August 2018; pp. 3580–3585. [CrossRef]
67. Niu, X.; Zhao, X.; Han, H.; Das, A.; Dantcheva, A.; Shan, S.; Chen, X. Robust Remote Heart Rate Estimation from Face Utilizing Spatial-temporal Attention. In Proceedings of the 2019 14th IEEE International Conference on Automatic Face Gesture Recognition (FG 2019), Lille, France, 14–18 May 2019; pp. 1–8. [CrossRef]
68. Niu, X.; Shan, S.; Han, H.; Chen, X. RhythmNet: End-to-End Heart Rate Estimation From Face via Spatial-Temporal Representation. *IEEE Trans. Image Process.* **2020**, *29*, 2409–2423. [CrossRef] [PubMed]
69. Niu, X.; Yu, Z.; Han, H.; Li, X.; Shan, S.; Zhao, G. Video-based remote physiological measurement via cross-verified feature disentangling. In *European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 295–310.
70. Song, R.; Zhang, S.; Li, C.; Zhang, Y.; Cheng, J.; Chen, X. Heart Rate Estimation From Facial Videos Using a Spatiotemporal Representation with Convolutional Neural Networks. *IEEE Trans. Instrum. Meas.* **2020**, *69*, 7411–7421. [CrossRef]
71. Lu, H.; Han, H. NAS-HR: Neural architecture search for heart rate estimation from face videos. *Virtual Real. Intell. Hardw.* **2021**, *3*, 33–42. [CrossRef]
72. Meziatisabour, R.; Benezeth, Y.; De Oliveira, P.; Chappe, J.; Yang, F. UBFC-Phys: A Multimodal Database For Psychophysiological Studies of Social Stress. *IEEE Trans. Affect. Comput.* **2021**, 1. [CrossRef]
73. McDuff, D.; Gontarek, S.; Picard, R. Remote measurement of cognitive stress via heart rate variability. In Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 2957–2960. [CrossRef]
74. Gupta, P.; Bhowmick, B.; Pal, A. Exploring the Feasibility of Face Video Based Instantaneous Heart-Rate for Micro-Expression Spotting. In Proceedings of the 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Salt Lake City, UT, USA, 18–22 June 2018; pp. 1316–1323. [CrossRef]
75. Monkaresi, H.; Bosch, N.; Calvo, R.A.; D'Mello, S.K. Automated Detection of Engagement Using Video-Based Estimation of Facial Expressions and Heart Rate. *IEEE Trans. Affect. Comput.* **2017**, *8*, 15–28. [CrossRef]
76. Kessler, V.; Thiam, P.; Amirian, M.; Schwenker, F. Pain recognition with camera photoplethysmography. In Proceedings of the 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA), Montreal, QC, Canada, 28 November–1 December 2017; pp. 1–5. [CrossRef]
77. Huang, D.; Feng, X.; Zhang, H.; Yu, Z.; Peng, J.; Zhao, G.; Xia, Z. Spatio-Temporal Pain Estimation Network with Measuring Pseudo Heart Rate Gain. *IEEE Trans. Multimed.* **2021**, 1. [CrossRef]
78. Yang, R.; Guan, Z.; Yu, Z.; Zhao, G.; Feng, X.; Peng, J. Non-contact Pain Recognition from Video Sequences with Remote Physiological Measurements Prediction. *arXiv* **2021**, arXiv:2105.08822.
79. Mitra, B.; Luckhoff, C.; Mitchell, R.D.; O'Reilly, G.M.; Smit, D.V.; Cameron, P.A. Temperature screening has negligible value for control of COVID-19. *Emerg. Med. Australas.* **2020**, *32*, 867–869.
80. Vilke, G.M.; Brennan, J.J.; Cronin, A.O.; Castillo, E.M. Clinical Features of Patients with COVID-19: Is Temperature Screening Useful? *J. Emerg. Med.* **2020**, *59*, 952–956. [CrossRef] [PubMed]
81. Stave, G.M.; Smith, S.E.; Hymel, P.A.; Heron, R.J.L. Worksite Temperature Screening for COVID-19. *J. Occup. Environ. Med.* **2021**, 63, 638–641.
82. Lippi, G.; Mattiuzzi, C.; Henry, B. Is Body Temperature Mass Screening a Reliable and Safe Option for Preventing COVID-19 Spread? *SSRN Electron. J.* **2021**. [CrossRef]
83. Natarajan, A.; Su, H.W.; Heneghan, C. Assessment of physiological signs associated with COVID-19 measured using wearable devices. *medRxiv* **2020**,
84. Pavri, B.; Kloo, J.; Farzad, D.; Riley, J. Behavior of the PR Interval with Increasing Heart Rate in Patients with COVID-19. *Heart Rhythm* **2020**, *17*. [CrossRef]
85. Gawałko, M.; Kapłon-Cieślicka, A.; Hohl, M.; Dobrev, D.; Linz, D. COVID-19 associated atrial fibrillation: Incidence, putative mechanisms and potential clinical implications. *Int. J. Cardiol. Heart Vasc.* **2020**, *30*, 100631. [CrossRef] [PubMed]
86. Stone, E.; Kiat, H.; McLachlan, C.S. Atrial fibrillation in COVID-19: A review of possible mechanisms. *FASEB J.* **2020**, *34*, 11347–11354. [CrossRef]
87. Schnaubelt, S.; Breyer, M.K.; Siller-Matula, J.; Domanovits, H. Atrial fibrillation: A risk factor for unfavourable outcome in COVID-19? A case report. *Eur. Heart J. Case Rep.* **2020**, *4*, 1–6. [CrossRef]
88. Shi, J.; Alikhani, I.; Li, X.; Yu, Z.; Seppänen, T.; Zhao, G. Atrial Fibrillation Detection From Face Videos by Fusing Subtle Variations. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *30*, 2781–2795. [CrossRef]
89. Zhu, G.; Li, J.; Meng, Z.; Yu, Y.; Li, Y.; Tang, X.; Dong, Y.; Sun, G.; Zhou, R.; Wang, H.; et al. Learning from Large-Scale Wearable Device Data for Predicting the Epidemic Trend of COVID-19. *Discret. Dyn. Nat. Soc.* **2020**, *2020*, 1–8. [CrossRef]
90. Mishra, T.; Wang, M.; Metwally, A.A.; Bogu, G.K.; Brooks, A.W.; Bahmani, A.; Alavi, A.; Celli, A.; Higgs, E.; Dagan-Rosenfeld, O.; et al. Pre-symptomatic detection of COVID-19 from smartwatch data. *Nat. Biomed. Eng.* **2020**, *4*, 1208–1220. [CrossRef] [PubMed]
91. Suwajanakorn, S.; Seitz, S.M.; Kemelmacher-Shlizerman, I. Synthesizing Obama: Learning Lip Sync from Audio. *ACM Trans. Graph.* **2017**, *36*. [CrossRef]

92. Nguyen, T.; Nguyen, C.; Nguyen, D.; Nguyen, D.; Nahavandi, S. Deep Learning for Deepfakes Creation and Detection. *arXiv* **2019**, arXiv:1909.11573.
93. Korshunov, P.; Marcel, S. DeepFakes: A New Threat to Face Recognition? Assessment and Detection. *arXiv* **2018**, arXiv:1812.08685.
94. Westerlund, M. The Emergence of Deepfake Technology: A Review. *Technol. Innov. Manag. Rev.* **2019**, *9*, 40–53. [CrossRef]
95. Fernandes, S.; Raj, S.; Ortiz, E.; Vintila, I.; Salter, M.; Urosevic, G.; Jha, S. Predicting Heart Rate Variations of Deepfake Videos using Neural ODE. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea, 27–28 October 2019; pp. 1721–1729. [CrossRef]
96. Ciftci, U.; Demir, I.; Yin, L. FakeCatcher: Detection of Synthetic Portrait Videos using Biological Signals. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, 1. [CrossRef] [PubMed]
97. Tuckson, R.V.; Edmunds, M.; Hodgkins, M.L. Telehealth. *N. Engl. J. Med.* **2017**, *377*, 1585–1592. [CrossRef]
98. Song, R.; Li, J.; Wang, M.; Cheng, J.; Li, C.; Chen, X. Remote Photoplethysmography with an EEMD-MCCA Method Robust Against Spatially Uneven Illuminations. *IEEE Sens. J.* **2021**, *21*, 13484–13494. [CrossRef]
99. Zhao, F.; Li, M.; Qian, Y.; Tsien, J.Z. Remote Measurements of Heart and Respiration Rates for Telemedicine. *PLoS ONE* **2013**, *8*, e71384. [CrossRef]
100. Zhou, X.; Snoswell, C.L.; Harding, L.E.; Bambling, M.; Edirippulige, S.; Bai, X.; Smith, A.C. The Role of Telehealth in Reducing the Mental Health Burden from COVID-19. *Telemed. E-Health* **2020**, *26*, 377–379. [CrossRef]
101. Alsaadi, I. Physiological Biometric Authentication Systems, Advantages, Disadvantages And Future Development: A Review. *Int. J. Sci. Technol. Res.* **2015**, *4*, 285–289.
102. Kumar, S.; Singh, S.; Kumar, J. A comparative study on face spoofing attacks. In Proceedings of the 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, India, 5–6 May 2017; pp. 1104–1108. [CrossRef]
103. Yu, Z.; Qin, Y.; Li, X.; Zhao, C.; Lei, Z.; Zhao, G. Deep learning for face anti-spoofing: A survey. *arXiv* **2021**, arXiv:2106.14948.
104. Liu, S.Q.; Lan, X.; Yuen, P.C. Remote Photoplethysmography Correspondence Feature for 3D Mask Face Presentation Attack Detection. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
105. Li, X.; Komulainen, J.; Zhao, G.; Yuen, P.C.; Pietikäinen, M. Generalized face anti-spoofing by detecting pulse from face videos. In *Proceedings of the 2016 23rd International Conference on Pattern Recognition (ICPR)*; IEEE: New York, NY, USA, 2016; pp. 4244–4249.
106. Liu, S.; Yuen, P.C.; Zhang, S.; Zhao, G. 3D Mask Face Anti-spoofing with Remote Photoplethysmography. In *Computer Vision—ECCV 2016*; Leibe, B., Matas, J., Sebe, N., Welling, M., Eds.; Springer International Publishing: Cham, Switzerland, 2016; pp. 85–100.
107. Yu, Z.; Li, X.; Wang, P.; Zhao, G. TransRPPG: Remote Photoplethysmography Transformer for 3D Mask Face Presentation Attack Detection. *arXiv* **2021**, arXiv:2104.07419.
108. Lin, B.; Li, X.; Yu, Z.; Zhao, G. Face liveness detection by rppg features and contextual patch-based cnn. In Proceedings of the 2019 3rd International Conference on Biometric Engineering and Applications, Stockholm, Sweden, 29–31 May 2019; pp. 61–68.
109. Kuo, J.; Koppel, S.; Charlton, J.L.; Rudin-Brown, C.M. Evaluation of a video-based measure of driver heart rate. *J. Saf. Res.* **2015**, *54*, 55–59. [CrossRef] [PubMed]
110. Zhang, Q.; Xu, G.Q.; Wang, M.; Zhou, Y.; Feng, W. Webcam based non-contact real-time monitoring for the physiological parameters of drivers. In Proceedings of the 4th Annual IEEE International Conference on Cyber Technology in Automation, Control and Intelligent, Hong Kong, China, 4–7 June 2014; pp. 648–652. [CrossRef]
111. Lee, K.; Han, D.; Ko, H. Video Analytic Based Health Monitoring for Driver in Moving Vehicle by Extracting Effective Heart Rate Inducing Features. *J. Adv. Transp.* **2018**, *2018*, 1–9. [CrossRef]
112. Zhang, Q.; Zhou, Y.; Song, S.; Liang, G.; Ni, H. Heart Rate Extraction Based on Near-Infrared Camera: Towards Driver State Monitoring. *IEEE Access* **2018**, *6*, 33076–33087. [CrossRef]
113. Wu, B.F.; Chu, Y.W.; Huang, P.W.; Chung, M.L. Neural Network Based Luminance Variation Resistant Remote-Photoplethysmography for Driver's Heart Rate Monitoring. *IEEE Access* **2019**, *7*, 57210–57225. [CrossRef]
114. Huang, P.W.; Wu, B.J.; Wu, B.F. A Heart Rate Monitoring Framework for Real-World Drivers Using Remote Photoplethysmography. *IEEE J. Biomed. Health Inform.* **2021**, *25*, 1397–1408. [CrossRef] [PubMed]
115. Tsai, Y.C.; Lai, P.W.; Huang, P.W.; Lin, T.M.; Wu, B.F. Vision-Based Instant Measurement System for Driver Fatigue Monitoring. *IEEE Access* **2020**, *8*, 67342–67353. [CrossRef]
116. Magdalena Nowara, E.; Marks, T.K.; Mansour, H.; Veeraraghavan, A. SparsePPG: Towards Driver Monitoring Using Camera-Based Vital Signs Estimation in Near-Infrared. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Salt Lake City, UT, USA, 18–22 June 2018.
117. Wu, B.F.; Chu, Y.W.; Huang, P.W.; Chung, M.L.; Lin, T.M. A Motion Robust Remote-PPG Approach to Driver's Health State Monitoring. In *Computer Vision—ACCV 2016 Workshops*; Chen, C.S., Lu, J., Ma, K.K., Eds.; Springer International Publishing: Cham, Switzerland, 2017; pp. 463–476.
118. Hernandez-Ortega, J.; Nagae, S.; Fierrez, J.; Morales, A. Quality-Based Pulse Estimation from NIR Face Video with Application to Driver Monitoring. In *Pattern Recognition and Image Analysis*; Morales, A., Fierrez, J., Sánchez, J.S., Ribeiro, B., Eds.; Springer International Publishing: Cham, Switzerland, 2019; pp. 108–119.
119. Al-Naji, A.A.; Perera, A.; Chahl, J. Remote measurement of cardiopulmonary signal using an unmanned aerial vehicle. In Proceedings of the IOP Conference Series: Materials Science and Engineering, Bangkok, Thailand, 24–26 February 2018; Volume 405, p. 012001. [CrossRef]

120. Al-Naji, A.A.; Perera, A.; Chahl, J. Remote monitoring of cardiorespiratory signals from a hovering unmanned aerial vehicle. *Biomed. Eng. Online* **2017**, *16*. [CrossRef]
121. Al-Naji, A.; Perera, A.G.; Mohammed, S.L.; Chahl, J. Life Signs Detector Using a Drone in Disaster Zones. *Remote Sens.* **2019**, *11*, 2441. [CrossRef]
122. Al-Naji, A.; Chahl, J. Remote Optical Cardiopulmonary Signal Extraction with Noise Artifact Removal, Multiple Subject Detection Long-Distance. *IEEE Access* **2018**, *6*, 11573–11595. [CrossRef]
123. Klaessens, J.H.; van den Born, M.; van der Veen, A.; van de Kraats, J.S.; van den Dungen, F.A.; Verdaasdonk, R.M. Development of a baby friendly non-contact method for measuring vital signs: First results of clinical measurements in an open incubator at a neonatal intensive care unit. In *Advanced Biomedical and Clinical Diagnostic Systems XII*; Vo-Dinh, T., Mahadevan-Jansen, A., Grundfest, W.S., Eds.; SPIE: Bellingham, WA, USA, 2014; pp. 257–263.
124. Villarroel, M.; Guazzi, A.; Jorge, J.; Davis, S.; Watkinson, P.; Green, G.; Shenvi, A.; McCormick, K.; Tarassenko, L. Continuous non-contact vital sign monitoring in neonatal intensive care unit. *Healthc. Technol. Lett.* **2014**, *1*, 87–91. [CrossRef] [PubMed]
125. Scalise, L.; Bernacchia, N.; Ercoli, I.; Marchionni, P. Heart rate measurement in neonatal patients using a webcamera. In Proceedings of the 2012 IEEE International Symposium on Medical Measurements and Applications Proceedings, Budapest, Hungary, 18–19 May 2012; pp. 1–4. [CrossRef]
126. Cobos-Torres, J.C.; Abderrahim, M.; Martínez-Orgado, J. Non-Contact, Simple Neonatal Monitoring by Photoplethysmography. *Sensors* **2018**, *18*, 4362. [CrossRef] [PubMed]
127. Gibson, K.; Al-Naji, A.; Fleet, J.A.; Steen, M.; Chahl, J.; Huynh, J.; Morris, S. Noncontact Heart and Respiratory Rate Monitoring of Preterm Infants Based on a Computer Vision System: Protocol for a Method Comparison Study. *JMIR Res. Protoc.* **2019**, *8*, e13400. [CrossRef] [PubMed]
128. Mestha, L.K.; Kyal, S.; Xu, B.; Lewis, L.E.; Kumar, V. Towards continuous monitoring of pulse rate in neonatal intensive care unit with a webcam. In Proceedings of the 2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society, Chicago, IL, USA, 26–30 August 2014; pp. 3817–3820. [CrossRef]
129. van Gastel, M.; Balmaekers, B.; Oetomo, S.B.; Verkruysse, W. Near-continuous non-contact cardiac pulse monitoring in a neonatal intensive care unit in near darkness. In *Proceedings of the Optical Diagnostics and Sensing XVIII: Toward Point-of-Care Diagnostics*; Coté, G.L., Ed.; SPIE: Bellingham, WA, USA, 2018; pp. 230–238.
130. Malafaya, D.; Domingues, S.; Oliveira, H.P. Domain Adaptation for Heart Rate Extraction in the Neonatal Intensive Care Unit. In Proceedings of the 2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), Seoul, Korea, 16–19 December 2020; pp. 1082–1086. [CrossRef]
131. Antink, C.H.; Ferreira, J.C.M.; Paul, M.; Lyra, S.; Heimann, K.; Karthik, S.; Joseph, J.; Jayaraman, K.; Orlikowsky, T.; Sivaprakasam, M.; et al. Fast body part segmentation and tracking of neonatal video data using deep learning. *Med. Biol. Eng. Comput.* **2020**, *58*, 3049–3061. [CrossRef]
132. Chaichulee, S.; Villarroel, M.; Jorge, J.; Arteta, C.; McCormick, K.; Zisserman, A.; Tarassenko, L. Cardio-respiratory signal extraction from video camera data for continuous non-contact vital sign monitoring using deep learning. *Physiol. Meas.* **2019**, *40*, 115001. [CrossRef]
133. Wang, W.; den Brinker, A.C.; Stuijk, S.; de Haan, G. Robust heart rate from fitness videos. *Physiol. Meas.* **2017**, *38*, 1023–1044. [CrossRef]
134. Chang, C.M.; Hung, C.C.; Zhao, C.; Lin, C.L.; Hsu, B.Y. Learning-based Remote Photoplethysmography for Physiological Signal Feedback Control in Fitness Training. In Proceedings of the 2020 15th IEEE Conference on Industrial Electronics and Applications (ICIEA), Kristiansand, Norway, 9–13 November 2020; pp. 1663–1668. [CrossRef]
135. Zhao, C.; Lin, C.L.; Chen, W.; Chen, M.K.; Wang, J. Visual heart rate estimation and negative feedback control for fitness exercise. *Biomed. Signal Process. Control* **2020**, *56*, 101680. [CrossRef]
136. Xie, K.; Fu, C.H.; Liang, H.; Hong, H.; Zhu, X. Non-contact Heart Rate Monitoring for Intensive Exercise Based on Singular Spectrum Analysis. In Proceedings of the 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 28–30 March 2019; pp. 228–233. [CrossRef]
137. McDuff, D.; Blackford, E. iPhys: An Open Non-Contact Imaging-Based Physiological Measurement Toolbox. In Proceedings of the 2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Berlin, Germany, 23–27 July 2019; pp. 6521–6524. [CrossRef]
138. van der Kooij, K.M.; Naber, M. An open-source remote heart rate imaging method with practical apparatus and algorithms. *Behav. Res. Methods* **2019**, *51*, 2106–2119. [CrossRef]
139. Boccignone, G.; Conte, D.; Cuculo, V.; D'Amelio, A.; Grossi, G.; Lanzarotti, R. An Open Framework for Remote-PPG Methods and Their Assessment. *IEEE Access* **2020**, *8*, 216083–216103. [CrossRef]
140. Estepp, J.R.; Blackford, E.B.; Meier, C.M. Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography. In Proceedings of the 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC), San Diego, CA, USA, 5–8 October 2014; pp. 1462–1469. [CrossRef]
141. Heusch, G.; Anjos, A.; Marcel, S. A Reproducible Study on Remote Heart Rate Measurement. *arXiv* **2017**, arXiv:1709.00962.
142. Soleymani, M.; Lichtenauer, J.; Pun, T.; Pantic, M. A Multimodal Database for Affect Recognition and Implicit Tagging. *IEEE Trans. Affect. Comput.* **2012**, *3*, 42–55. [CrossRef]

143. Zhang, Z.; Girard, J.M.; Wu, Y.; Zhang, X.; Liu, P.; Ciftci, U.; Canavan, S.; Reale, M.; Horowitz, A.; Yang, H.; et al. Multimodal Spontaneous Emotion Corpus for Human Behavior Analysis. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 3438–3446. [CrossRef]
144. Stricker, R.; Müller, S.; Gross, H.M. Non-contact video-based pulse rate measurement on a mobile service robot. In Proceedings of the 23rd IEEE International Symposium on Robot and Human Interactive Communication, Edinburgh, UK, 25–29 August 2014; pp. 1056–1062. [CrossRef]
145. Bobbia, S.; Macwan, R.; Benezeth, Y.; Mansouri, A.; Dubois, J. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognit. Lett.* **2019**, *124*, 82–90. [CrossRef]
146. Niu, X.; Han, H.; Shan, S.; Chen, X. VIPL-HR: A Multi-modal Database for Pulse Estimation from Less-constrained Face Video. In *Asian Conference on Computer Vision 2018;* Jawahar, C., Li, H., Mori, G., Schindler, K., Eds.; Springer International Publishing: Cham, Switzerland, 2018; pp. 562–576.
147. Russakovsky, O.; Deng, J.; Su, H.; Krause, J.; Satheesh, S.; Ma, S.; Huang, Z.; Karpathy, A.; Khosla, A.; Bernstein, M.; et al. Imagenet large scale visual recognition challenge. *Int. J. Comput. Vis.* **2015**, *115*, 211–252. [CrossRef]
148. Li, X.; Han, H.; Lu, H.; Niu, X.; Yu, Z.; Dantcheva, A.; Zhao, G.; Shan, S. The 1st Challenge on Remote Physiological Signal Sensing (RePSS). In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Seattle, WA, USA, 14–19 June 2020; pp. 1274–1281. [CrossRef]
149. Nowara, E.M.; McDuff, D.; Veeraraghavan, A. Systematic analysis of video-based pulse measurement from compressed videos. *Biomed. Opt. Express* **2021**, *12*, 494–508. [CrossRef] [PubMed]
150. Kenzaka, T.; Okayama, M.; Kuroki, S.; Fukui, M.; Yahata, S.; Hayashi, H.; Kitao, A.; Sugiyama, D.; Kajii, E.; Hashimoto, M. Importance of vital signs to the early diagnosis and severity of sepsis: Association between vital signs and sequential organ failure assessment score in patients with sepsis. *Intern. Med.* **2012**, *51*, 871–876. [CrossRef]
151. Chalari, E.; Intas, G.; Stergiannis, P.; Vezyridis, P.; Paraskevas, V.; Fildissis, G. The importance of vital signs in the triage of injured patients. *Crit. Care Nurs. Q.* **2012**, *35*, 292–298. [CrossRef]
152. Manta, C.; Jain, S.S.; Coravos, A.; Mendelsohn, D.; Izmailova, E.S. An Evaluation of Biometric Monitoring Technologies for Vital Signs in the Era of COVID-19. *Clin. Transl. Sci.* **2020**, *13*, 1034–1044. [CrossRef]
153. Pimentel, M.A.; Redfern, O.C.; Hatch, R.; Young, J.D.; Tarassenko, L.; Watkinson, P.J. Trajectories of vital signs in patients with COVID-19. *Resuscitation* **2020**, *156*, 99–106. [CrossRef] [PubMed]
154. Djeldjli, D.; Bousefsaf, F.; Maaoui, C.; Bereksi-Reguig, F.; Pruski, A. Remote estimation of pulse wave features related to arterial stiffness and blood pressure using a camera. *Biomed. Signal Process. Control* **2021**, *64*, 102242. [CrossRef]
155. Luo, H.; Yang, D.; Barszczyk, A.; Vempala, N.; Wei, J.; Wu, S.J.; Zheng, P.P.; Fu, G.; Lee, K.; Feng, Z.P. Smartphone-based blood pressure measurement using transdermal optical imaging technology. *Circ. Cardiovasc. Imaging* **2019**, *12*, e008857. [CrossRef] [PubMed]
156. Fan, X.; Ye, Q.; Yang, X.; Choudhury, S.D. Robust blood pressure estimation using an RGB camera. *J. Ambient Intell. Humaniz. Comput.* **2020**, *11*, 4329–4336. [CrossRef]
157. Casalino, G.; Castellano, G.; Zaza, G. A mHealth solution for contact-less self-monitoring of blood oxygen saturation. In Proceedings of the 2020 IEEE Symposium on Computers and Communications (ISCC), Rennes, France, 7–10 July 2020; pp. 1–7. [CrossRef]
158. Shao, D.; Liu, C.; Tsow, F.; Yang, Y.; Du, Z.; Iriya, R.; Yu, H.; Tao, N. Noncontact Monitoring of Blood Oxygen Saturation Using Camera and Dual-Wavelength Imaging System. *IEEE Trans. Biomed. Eng.* **2016**, *63*, 1091–1098. [CrossRef] [PubMed]
159. Kong, L.; Zhao, Y.; Dong, L.; Jian, Y.; Jin, X.; Li, B.; Feng, Y.; Liu, M.; Liu, X.; Wu, H. Non-contact detection of oxygen saturation based on visible light imaging device using ambient light. *Opt. Express* **2013**, *21*, 17464–17471. [CrossRef] [PubMed]
160. Nowara, E.M.; McDuff, D.; Veeraraghavan, A. A meta-analysis of the impact of skin tone and gender on non-contact photoplethysmography measurements. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 284–285.
161. Hassan, M.; Malik, A.; Fofi, D.; Karasfi, B.; Meriaudeau, F. Towards health monitoring using remote heart rate measurement using digital camera: A feasibility study. *Measurement* **2020**, *149*, 106804. [CrossRef]
162. Blackford, E.B.; Estepp, J.R. Measurements of pulse rate using long-range imaging photoplethysmography and sunlight illumination outdoors. In Proceedings of the Optical Diagnostics and Sensing XVII: Toward Point-of-Care Diagnostics, San Diego, CA, USA, 6–10 August 2017; Volume 10072, p. 100720S.