Super-Resolution Convolutional Network for Image Quality Enhancement in Remote Photoplethysmography based Heart Rate Estimation

Smera Premkumar K[1], A. Angelopoulou[2], E. Kapetanios[3], T. Chaussalet[2] and Jude Hemanth D[1,*]

[1] Department of ECE, Karunya Institute of Technology and Sciences, Coimbatore, India
[2]School of Computer Science and Engineering, University of Westminster, London, UK
[3]School of Physics, Engineering & Computer Science, University of Hertfordshire, UK
*Corresponding Author: judehemanth@karunya.edu

Abstract. Heart rate (HR) is one of the important vital parameters of the human body and understanding this vital sign provides key insights into human wellness. Imaging photoplethysmography (iPPG) allows HR detection from video recordings and its unbeatable compliance over the state of art methods has made much attention among researchers. Since it is a camera-based technique, measurement accuracy depends on the quality of input images. In this paper, we present a pipeline for efficient measurement of HR that includes a learning-based super-resolution preprocessing step. This preprocessing image enhancement step has shown promising results on low-resolution input images and works better on iPPG algorithms. The experimental results verified the reliability of this method.

Keywords: Remote Photoplethysmography, Image Enhancement, Heart Rate (HR) Detection, Convolutional Neural Network

1. Introduction

Heart Rate Measurement (HRM) is a crucial physiological regulator of a person's total cardiac output. The far-flung heart rate tracking devices have led to substantial interest in Heart Rate(HR) as a prospective clinical diagnosis tool. The gold standard to analyze cardiac measurements is an Electrocardiogram (ECG) which measures the electrical activity of the heart through attached sensors(called electrodes) to the skin. Another favoured optical-based technique is Photoplethysmography (PPG) which detects the changes in blood volume pulse (BVP) via contact sensors attached to anatomical locations such as wrists, fingers, and toes. Commercial wearable devices such as fitness trackers, and smart watches make use of this principle where the sensor emits light to the skin and measures the reflected intensity due to optical absorption of blood.

These state-of-art methods of measuring cardiac activities either need physical contact in clinical settings or a sensor attached to the body and it is not suitable for long time

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_15

monitoring as they can cause discomfort, especially in neonates, and elderly care. In recent years, Imaging Photoplethysmography(iPPG) or Remote Photoplethysmography (rPPG) has been a prominent topic among researchers that measures HR from face images by tracking volumetric variations of blood circulation, while it is invisible to the human eye. Even though it is a progressive method of PPG technology, the iPPG method does not require any kind of physical contact. This relies on the signals obtained from the video streams that can predict not only heart rate but also other vital information of the body like heart rate variability, blood pressure etc. and thereby infers mental stress, variations in cardiovascular functions, quality of sleep, drowsiness.

Recently, Covid 19 Pandemic made a huge transition in healthcare from in-person care to telehealth. This practical application of the PPG principle can measure the vital parameters well-nigh possible from any mobile camera/webcam. It provides a static, contactless health monitoring of patients and no longer needs any clinical settings. Although appreciable progress has been made in rPPG methods, still a few challenges remain open such as motion, skin tone, compression etc. Since rPPG is a camera-based technique Compression artifact is an inevitable challenge, especially in real-time scenarios such as telehealth. Remote measurement of HR relies on very fine details of the input video and a small artifact can make a huge impact on the accuracy of results especially when it comes to processing low-resolution images/videos. This paper aims at a fast super-resolution network (FSRCNN) for the low-resolution video input which enhances the input frames that ensure vital sign measurement more efficiently. We present a pipeline for iPPG measurement with a preprocessing super-resolution step which provides an improved signal to noise ratio (SNR) and combat the effects of temporal compression of video to an extent.

2. Related Works

Photoplethysmography (PPG) is a non-invasive, optical technique that is used to detect volumetric changes of blood in the microvascular bed of tissue [1]. The possibility of non-contact physiological computation using a thermal camera has been introduced in [2] and [3] demonstrated that plethysmography signals could be measured from the human face from simple consumer-level camera recordings with ambient light conditions. Since then, a substantial number of researches have been conducted in remote Photoplethysmography.An ICA based algorithm has been explained in [4] as an optimal combination of raw signals in which the raw signals is separated into independent non-Gaussian channels. In this method, authors arbitrate that the second component produced after the ICA is considered as a periodic one and used for further processing.

With the emergence of the deep learning end to end method, extensive opportunities are opening up for performing tasks more efficiently in a better way. Chen et.al [5] was introduced the first end to end Learning Model ‘DeepPhys’ which is based on a Convolutional Attention Network (CAN) and enables spatiotemporal visualization of signals. This paper proposed a skin reflection model that is exceptionally robust in different illumination conditions. Subsequently, many learning-based methods have been proposed in the literature for HR estimation which includes illustrated ETA-

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_15

rPPGNet illustrated by [6] ,meta phys model demonstrated by [7] and the neural architecture, AutoHR proposed [8].

Although appreciable progress has been made in rPPG methods in the last few years, still a few challenges remain open such as motion robustness, illumination, and skin tone compression artifacts etc [9]. It should be noted that the compression artifacts are also an inevitable challenge in rPPG environment, but only a few approaches are there in literature to succeed in dealing with compression artifacts. Most video datasets are captured as raw images to avoid lossy compression and it demands immense storage. But it is a major challenge in real-time scenarios such as telemedicine, as it requires large memory requirements. Since rPPG measurement leverages very fine details of input video, compression makes a huge impact on the accuracy and robustness of physiological signal measurement. A significant research gap can be seen in literature to incorporate compression of low-resolution input videos into consideration. McDuff et al. [10] explained the impacts of video compression and Song. et al [11] have shown the resolution of the input videos affects the quality of output measurement.

3. Methods

We present a preprocessing step that can improve the accuracy of the output physiological signal even if the input image is with low resolution. The motivation behind this work is the great success of learning based super-resolution networks on face images/videos. The process of enhancing a low resolution (LR) image is called super-resolution (SR) while an interpolation method takes the weighted average of the neighboring pixels. A neural network can hallucinate details based on some prior information it collects from a large set of images and the details are then added to the image to create an SR image. An enhanced, sharper image can lead to reliable skin segmentation and thereby missing colour signal information can be extracted more efficiently.

3.1 Super Resolution Preprocessing

We use a Fast SRCNN proposed by C.Dong et.al [12] as the preprocessing enhancement network. We compared the Fast SRCNN with interpolation methods and normal SR convolutional network to check the computational complexity and speed. The main advantage of Fast SRCNN is that it directly accepts a low-resolution input and no longer needs a bicubic interpolation step. If we are resizing or moving an image, the new locations we have got will not necessarily match up with the previous ones. Bilinear interprets the missing values by using linear interpolation between the values whereas a bicubic interpolation takes 16 pixels(4x4) into the account. We have tried a super-resolution network proposed by c.Dong [13] for the preprocessing step. This method uses a bicubic interpolation preprocessing step to up sample the LR image and then a convolution operation is performed to improve image quality.

We found that this method has high computational complexity and low speed. To overcome this, we have tried FSRCNN, the advanced version of this meth-

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_15

od. and it works better in our work. Fast SRCNN capable of producing a super scaled image directly from the LR image. As shown in figure 2, the FSRCNN has four steps before a deconvolution process which include feature extraction, shrinking, non-linear mapping and expanding process. The feature extraction process replaces bicubic interpolation with 5x5 convolutions and a feature map reduction has happened in shrinking. Then 3x3 multiple layers are applied followed by a 1x1 convolution expansion. It is used to reduce the number of parameters and it could help to speed up the network. PReLU is used as an activation function. The network was pretrained and this network shows faster performance on standard image datasets for image SR Then using a 9x9 filter, a high-resolution image is reconstructed. This network shows higher image quality as well as a shorter run time when compared to interpolation and SRCNN. The network architecture is illustrated in Figure 1.

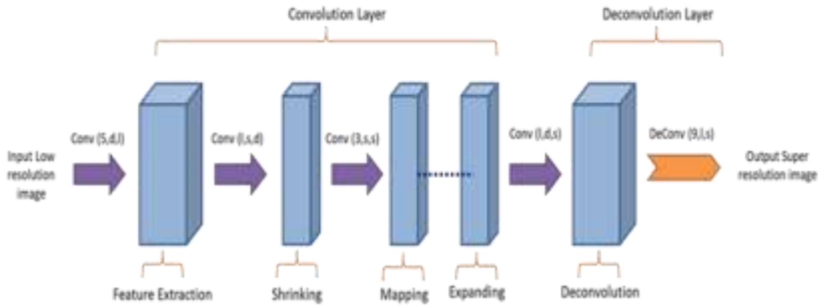

Figure 1: Architecture of fast super-resolution convolutional neural network

We can represent an FSRCNN network as FSRCNN (d,s,m). The computational complexity can be calculated as

$$O\{(25d + sd + 9ms^2 + ds + 81d)S_{LR}\} = O\{(9ms2 + 2sd + 106d)S_{LR}\}$$

We exclude the parameters of PReLU, which introduce a negligible computational cost. To compare the cost function, we use the mean square error (MSE). The optimization objective is represented as

$$\min \theta \ \frac{1}{n} \sum_{i=1}^{n} || F(Y_s^i; \theta) - Xi\|_2^2 ,$$

where Y_s^i and Xi is the ith LR and HR sub-image pair in the training data and $F(Y_s^i; \theta)$ is the network output for Y_s^i with parameters θ. Using a standard backpropagation, parameters are optimized using stochastic gradient descent.

To compare the efficiency of FSRCNN, the LR images are first up sampled using bicubic interpolation, SRCNN and then FSRCNN. From our experiments, it is evident that FSRCNN uses small filter sizes and a deeper network, and it has better PSNR (image quality) and less computational complexity compared to the other two methods. Because of its efficiency in training and testing over upscaling factors, it could be beneficial in real-time scenarios on generic CPU.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_15

4. Results & Discussions

The results of PSNR (dB) and test time have been verified on two publicly available datasets, MAHNOB-HCI and VicarVision. We use a pre-trained SRCNN network and used two public benchmark datasets for testing. The inference time is tested with the C++ implementation on an intel core i5 processor. The qualitative results based on image quality (PSNR) and test time are listed in Table 1.

Table 1: Test results

Test Data Set	Scaling Factor	Bicubic Interpolation PSNR/Time	SRCNN PSNR/Time(sec on CPU)	SRCNN PSNR /Time (sec on CPU)
MAHNOB-HCI	2	30.14/ -	32.35/1.3	33.25/.098
	3	26.45/ -	29.26/1.3	29.43/ .074
VICARVISION	2	29.81/ -	31.53/1.8	31.82/0.088
	3	26.41/ -	28.47/1.7	30.43/0.097

We use a pre-trained model for an upscaling factor in advance. During testing convolution operations are performed once and then, the image is up sampled to different sizes and the corresponding deconvolution layer is processed.

Figure2. Visualization of enhanced images

From the results, it is evident that the noticeable improvement in image quality achieves high running speed over interpolation and SRCNN methods. An example of visualization of image enhancement using interpolation and SRCNN can see in Figure 2. It has fast training and testing speed across different upscaling factors and this model can be adapted for real-time video SR.

4.1 HR Estimation

To test the reliability of this SR network in iPPG environment, we used two previously published iPPG methods to calculate the HR. To achieve this, the resulting pixels from the SRCNN network were spatially averaged for each frame to obtain the test signals for HR measurement. We used ICA method explained in [4] and the POS

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_15

method for PPG signal recovery. Spatially averaged colour signals were normalized and multiplied by the projection matrix P. A Butterworth filter is then applied to the model output. To extract the heart rate from the corresponding pulse signal, we used FFT, and HR was chosen from the peak with the greatest power. The HR was calculated using the average interbeat signal interval in seconds for a 30 second window.

Table 2: The effectiveness of the FSRCNN method on iPPG algorithms in terms of PSNR and RMSE

Method	ICA		POS	
	RMSE	SNR	RMSE	SNR
Bicubic	5.79	0.067	4.12	-0.043
SRCNN	5.21	0.012	5.12	-0.028
FSRCNN	4.74	0.027	4.31	-0.015

We perform HR estimation 30 seconds windows for each video. To check the quality, We used performance metric root mean square error (RMSE).

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N} \left| HR_i - \widehat{HR}_i \right|^2}{N}}$$

Where N is the total number of observation windows. HR_i is the ith measurement and $\widehat{HR}_i$ is the corresponding prediction. We evaluate the heart rate estimate using performance metrics RMSE and the blood volume pulse SNR.

The SRCNN network was trained using a dataset that is pre-trained using a different dataset and the evaluation is conducted using two publicly available datasets – VicarVision and MAHNOB-HCI. We compared the performance of bicubic interpolation, SRCNN and FSRCNN methods. We used 30seconds of each video to measure the Heart Rate. We have chosen o.6 seconds time interval. Table 2 shows the results using RMSE and SNR for each method. The FSRCNN method outperforms the bicubic interpolation method. The FSRCNN has reduced error compared to the bicubic interpolation method. The interpolation and FSRCNN method have reduced RMSE when compared to the low-resolution frames. Despite its reduced error, the FSRCNN network has less computational complexity and high speed compared to other interpolation methods.

5. Conclusion

In this paper, we present a pipeline with a Fast Super-resolution CNN based image enhancement method as preprocessing step for HR measurement from face videos. From the results, the proposed method has more accuracy than other interpolation methods. We have observed that this network can provide better results for low resolution and compressed input frames. This method can help to combine the advantage

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_15

of super-resolution with the iPPG measurements. In our future work, we would like to investigate the possibilities of advancements of remote methods using neural models to alleviate the challenges in rPPG measurement.

References

[1] Allen, J. (2007). Photoplethysmography and its application in clinical physiological measurement. Physiological Measurement, 28(3). https://doi.org/10.1088/0967-3334/28/3/R01

[2] Pavlidis, I., Dowdall, J., Sun, N., Puri, C., Fei, J., & Garbey, M. (2007). Interacting with human physiology. Computer Vision andImageUnderstanding, https://doi.org/10.1016/j.cviu.2006.11.018

[3] W. Verkruysse, L. O. Svaasand, and J. S. Nelson (2008).Remote plethysmographic imaging using ambient light. Opt. Express, vol. 16, no. 26, pp. 21 434–21 445

[4] Poh, M. Z., McDuff, D. J., & Picard, R. W. (2010). Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. Optics express, 18(10), 10762–10774. https://doi.org/10.1364/OE.18.010762

[5] Chen, W., & McDuff, D. (2018). DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks.. https://doi.org/10.1007/978-3-030-01216-8_22

[6] Hu, M., Qian, F., Guo, D., Wang, X., He, L., & Ren, F. (2021). ETA-rPPGNet: Effective Time-Domain Attention Network for Remote Heart Rate Measurement. IEEE Transactions on Instrumentation and Measurement, 70. https://doi.org/10.1109/TIM.2021.3058983

[7] Liu, X., Jiang, Z., Fromm, J., Xu, X., Patel, S., & McDuff, D. (2021). MetaPhys: Few-shot adaptation for non-contact physiological measurement. In ACM CHIL 2021 - Proceedings of the 2021 ACM Conference on Health, Inference, and Learning (Vol. 1, Issue 1). Association for Computing Machinery. https://doi.org/10.1145/3450439.3451870

[8] Yu, Z., Li, X., Niu, X., Shi, J., & Zhao, G. (2020). AutoHR: A Strong End-to-End Baseline for Remote Heart Rate Measurement with Neural Searching. IEEE Signal Processing Letters, 27, 1245–1249. https://doi.org/10.1109/LSP.2020.3007086

[9] Hoffman, W. F. C., &Lakens, D. (2021, June 25). Addressing Reproducibility Issues in Remote Photoplethysmography (rPPG) Research: An Investigation of Current Challenges and Release of a Public Algorithm Benchmarking Dataset. https://doi.org/10.17605/OSF.IO/XJF7U

[10] D. J. McDuff, E. B. Blackford and J. R. Estepp, "The Impact of Video Compression on Remote Cardiac Pulse Measurement Using Imaging Photoplethysmography," 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), 2017, pp. 63-70, doi: 10.1109/FG.2017.17.

[11] Song, R., Zhang, S., Cheng, J., Li, C., & Chen, X. (2020). New insights on super-high resolution for video-based heart rate estimation with a semi-blind source separation method. Computers in Biology and Medicine, 116(Complete). https://doi.org/10.1016/j.compbiomed.2019.103535

[12] C. Dong, C. C. Loy, K. He and X. Tang, "Image Super-Resolution Using Deep Convolutional Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295-307, 1 Feb. 2016, doi: 10.1109/TPAMI.2015.2439281.

[13] Dong, C., Loy, C.C., & Tang, X. (2016). Accelerating the Super-Resolution Convolutional Neural Network. ECCV.

ICCS Camera Ready Version 2022
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-08757-8_15