Deep Super Resolution for Recovering Physiological Information from Videos

Daniel McDuff
Microsoft Research, Redmond, USA
damcduff@microsoft.com

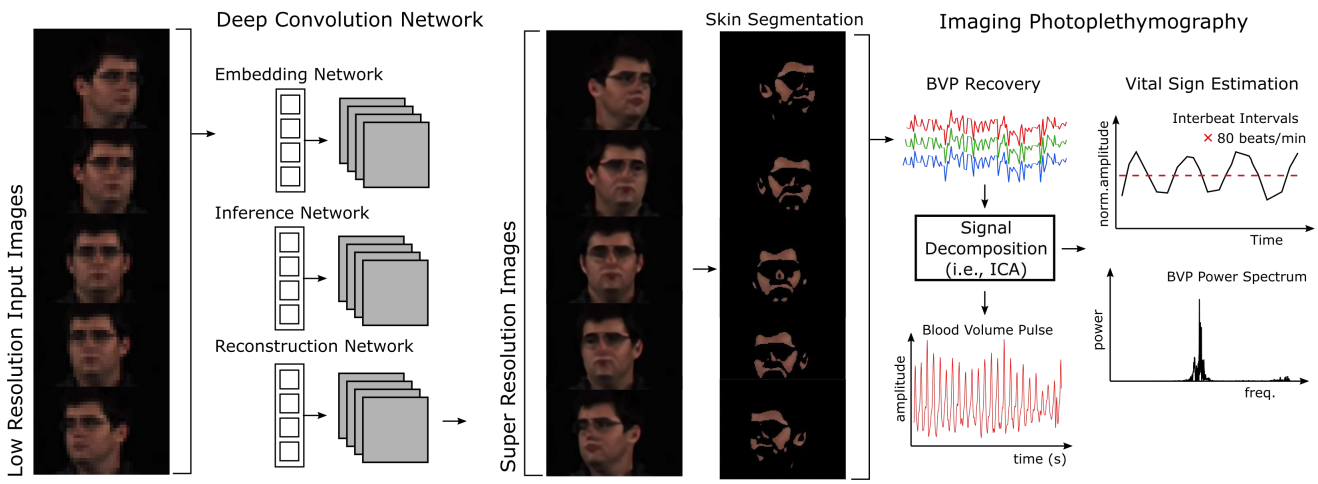

Figure 1: We present a system for recovering physiological information from videos using image super resolution. Our results show that a simple preprocessing step can considerably improve the accuracy of iPPG measurements. Furthermore, a deep learning-based super resolution model can outperform bilinear and bicubic image interpolation to provide further improvements in accuracy.

Abstract

Imaging photoplethysmography (iPPG) allows for remote measurement of vital signs from the human skin. In some applications the skin region of interest may only occupy a small number of pixels (e.g., if an individual is a large distance from the imager.) We present a novel pipeline for iPPG using an image super-resolution preprocessing step that can reduce the mean absolute error in heart rate prediction by over 30%. Furthermore, deep learning-based image super-resolution outperforms standard interpolation methods. Our method can be used in conjunction with any existing iPPG algorithm to estimate physiological parameters. It is particularly promising for analysis of low resolution and spatially compressed videos, where otherwise the pulse signal would be too weak.

1. Introduction

Video-based physiological measurement methods [16] present several advantages over traditional contact-based sensors (e.g., ECG, contact PPG). Contact devices often require obtrusive electrodes that can become uncomfortable with extended use and can be corrupted by body motions. These devices are often expensive and are not ubiquitously available. Video methods enable spatial analysis and visualization of blood flow, in addition to concomitant measurement of multiple people. Cameras are ubiquitous devices and the number of imagers in the world is continually increasing. Recent papers have demonstrated applications of video-based physiological measurement in ICUs [29], detecting sleep events such as apnea and tracking changes in cognitive load [18].

Photoplethysmography (PPG) is the measurement of the blood volume pulse (BVP) via light transmitted through, or reflected from the skin [1]. Traditional measurement involves contact sensors with dedicated light sources and customized imagers. Imaging photoplethysmography (iPPG) has developed as a method for capturing the BVP signal remotely using digital cameras and ambient light. Almost any digital camera (i.e., a webcam [21] or cellphone camera [11]) is sufficiently sensitive to capture the pulse signal when the subject is close to the device.

There are a number of methods for video compression that aim to reduce the bit rate of video data while retaining important visual information. However, video compression algorithms are not designed with the intention of preserving photoplethysmographic data. On the contrary compression algorithms often make assumptions that small differences in color values between pixels (intra-frame compression) or between frames (inter-frame compression) are not of high visual importance and discard them, influencing the underlying variations on which iPPG methods rely.

Video compression severely impacts the performance of iPPG methods [15]. McDuff et al. showed how heart rate measurement accuracy decreased linearly with the video compression score. However, in many situations it is impractical to store raw video files. Such data require enormous amounts of storage; for example, each raw, uncompressed 5.5-minute video file collected in this study was 11.9 GiB in size. Collecting data from many subjects (n=25) resulted in over 250 GiB for the two hours of standard definition video. This can be restrictive for researchers who may want to collect and distribute video data for iPPG purposes.

It is possible for iPPG signals to be recovered over relatively long-distances (up to 50 m [2].) As the subject moves further from the camera the limiting factor will soon become the number of pixels a single face, or skin region of interest (ROI), will occupy in the image.

Can facial videos with small ROI pixel areas or spatially compressed images be used reliably for iPPG? We systematically test this and show how image super-resolution can effectively boost the signal-to-noise ratio. We believe this is the first example of attempting to recover the iPPG signal from small frames (41×30 pixels) in which the face/skin occupies at most a few hundred pixels. We present an iPPG pipeline for recovering the pulse signal. Our pipeline features an image super-resolution preprocessing step. Figure 1 shows an overview of our approach. In the remainder of the paper we: 1) provide background on iPPG and image-super resolution, 2) describe our method, 3) present experiments and results on a corpus of videos from 25 subjects, and 4) discuss implications and future work.

2. Background

2.1. Imaging Photoplethysmography

Initial work by Takano and Ohta [25] demonstrated that the pulse could be recovered from digital video recorded from the human face. The green color channel typically carries the strongest pulse information [25, 28]. Although no single pixel contains a strong signal, when many pixels are spatially averaged the pulse wave can be recovered. While the green signal can be used, in many real-world contexts more sophisticated signal processing is required. Poh et al. [21] proposed the use of Independent Component Analysis (ICA) to recover the pulse signal and enable a fully automated iPPG framework with automated face segmentation. HR and heart rate variability (HRV) metrics can be recovered from videos with small head motions using this method [22].

Methods inspired by optical models of the skin have helped advance the state-of-the art. The CHROM [5] method uses a linear combination of the chrominance signals and makes the assumption of a standardized skin color profile to white-balance the video frames. The Pulse Blood Vector (PBV) method [6] relies on characteristic blood volume changes in different regions of the frequency spectrum to weight the color channels. Basing calibration on a more advanced skin-tissue model offers the potential of allowing more accurate recovery of the pulse wave and potentially reduces the need for computationally complex algorithms [20, 14]. The plane orthogonal to the skin (POS) [30] algorithm assumes the presence of a pulsatile color space signal and posits that this will be orthogonal to the skin color space. Using this framework the authors were able to demonstrate good pulse wave recovery even in the presence of larger head motions. Normalized Least Mean Squares (NLMS) adaptive filtering has been applied to help combat the effects of motion and illumination changes [13]. A recent approach using self-adaptive matrix completion also found good results by selectively choosing the region of interest to analyze [27]. We show that a preprocessing step can be applied to iPPG algorithms to improve the accuracy, and show that this benefits multiple existing approaches.

2.2. Image Super Resolution

Super resolution is a class of techniques that enables the resolution of an image to be enhanced. Specifically, the aim is to recover missing high-frequency spatial information from surrounding pixels. With digital images this can be performed algorithmically. The simplest form of image resolution upscaling involves estimating missing information using nearest neighbor or interpolation functions based on surrounding pixels. The most common interpolation function is a bicubic that leverages the surrounding 4×4 pixel grid to infer the values of intermediate pixels and thus enable a higher resolution output image. However, this smoothly interpolates between pixel values and thus performs poorly at recovering high frequency spatial information.

Machine learning can be used to infer missing information more effectively. Some methods use statistical image priors to predict the values of "missing" pixels [24]. Neighbor embedding [3], sparse coding [31] and random forests [23] have all been applied to the problem. More recently, researchers have tended to focus on deep neural networks for super resolution tasks yielding state-of-the-art

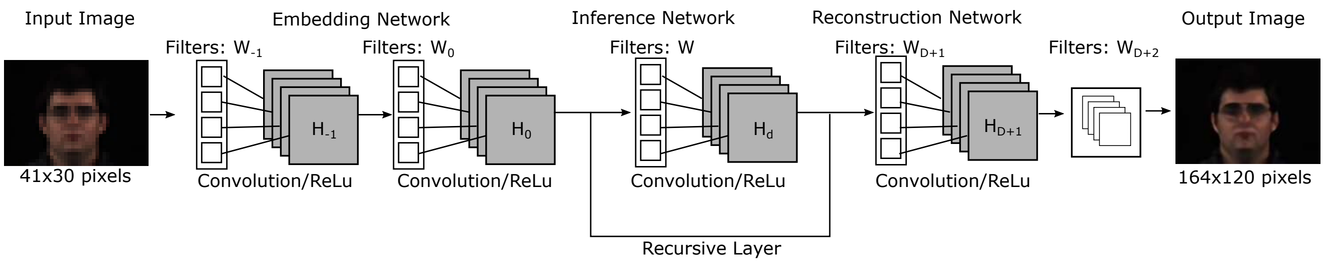

Figure 2: The architecture of the deeply-recursive convolutional network (DRCN) model for image super resolution [10]. The model features three networks, an embedding network, recursive inference network and a reconstruction network.

results [4, 12, 7]. These end-to-end networks are often more computationally expensive than interpolation but can provide much more superior results. Most relevant to this work would be the success of super resolution on facial images. In our application, unlike much prior super resolution research, we are both concerned with the recovery of high frequency spatial information and the correct inference of subtle color information pertaining to the pulse signal.

Why might super-resolution aid in iPPG methods? First, a sharper image can lead to more accurate facial registration and/or skin segmentation. Second, a learning-based super resolution method may recover missing color information more accurately than a linear or bicubic interpolation and this will in turn improve the signal-to-noise ratio of the recovered pulse signal. However, to date there is no empirical evidence to support these hypotheses. Therefore, we perform a systematic analysis of image interpolation/super resolution methods in the context of iPPG signal recovery.

3. Method

Our proposed pipeline features two stages. 1) Preprocessing of the video frames to increase the spatial resolution. 2) An iPPG method to recover the pulse wave and heart rate from the resulting videos. The interpolation/super resolution step is added before the iPPG algorithm and could be combined with any existing physiological measurement pipeline. For this reason we believe it has wide applicability.

3.1. Super Resolution

We compared three interpolation/super-resolution methods in this analysis.

Bilinear: The interpolation considers four pixels and fits a linear function in two dimensions to infer the missing values. Each inferred pixel is a weighted average of the pixels in a 2×2 neighborhood used to fit the linear function.

Bicubic: The interpolation considers 16 pixels and fits a cubic function in two dimensions to infer the missing values. Each inferred pixel is a weighted average of the pixels in a 4×4 neighborhood used to fit the cubic function.

Deeply-Recursive Convolutional Network: We use a deeply-recursive convolutional network (DRCN) as proposed by Kim et al. [10] as a learning-based method to infer the missing values. This network consists of three subnetworks ($f(\mathbf{x}) = f_3(f_2(f_1(\mathbf{x})))$). Figure 2 shows the network architecture. The embedding network (f_1) acts as a preprocessing layer creating a set of features for inference. Where:

$$H_{-1} = max(0, W_{-1} * \mathbf{x} + b_{-1}) \quad (1)$$

$$H_0 = max(0, W_0 * H_{-1} + b_0) \quad (2)$$

$$f_1(\mathbf{x}) = H_0 \quad (3)$$

H_n denotes the hidden layer values. The weight, *W*, and bias, *b*, matrices are learned during training.

The inference network (f_2) solves the task of super-resolution recursively.

$$H_d = max(0, W * H_{d-1} + b) \quad (4)$$

Finally, the reconstruction network (f_3) transforms the high resolution features back into the original image space. Where:

$$H_{D+1} = max(0, W_{D+1} * H_D + b_{D+1}) \quad (5)$$

$$\hat{\mathbf{y}} = max(0, W_{D+2} * H_{D+1} + b_{D+2}) \quad (6)$$

$$f_3(H) = \hat{\mathbf{y}} \quad (7)$$

The network was trained on the dataset presented in [31]. It was not trained on similar data to the frames in our video dataset and thus the generalizability of our solution is high. The DRCN network has shown good performance on standard image datasets used for evaluating image super-resolution. However, in those cases the objective was to maximize the peak signal-to-noise ratio (PSNR). As described above, it is important that the super resolution algorithm does not interfere with the subtle pixel color differences between frames.

3.2. Imaging Photoplethysmography

In our experiments the images were first upsampled using bilinear interpolation, bicubic interpolation or the

DRCN super resolution model. Then we applied one of two iPPG algorithms to extract the blood volume pulse and heart rate. We intentionally used previously published methods for this stage of the analysis as we did not want to introduce other sources of variance.

In both cases the images were converted from RGB to YC_RC_B and skin detection was applied to segment the pixels of interest. A pixel was considered skin if the following criteria were satisfied:

$$\begin{aligned} Y &> 80 \\ C_B > 77 &\text{ and } C_B < 127 \\ C_R > 133 &\text{ and } C_R < 173 \end{aligned} \tag{8}$$

The resulting pixels were then spatially averaged for each frame to give three observation signals x_1, x_2 and x_3.

ICA Method [17]: In this case, we detrended the color signals based on a smoothness priors approach [26] with λ=1000. ICA was used to recover the sources from the observed color signals. A 6th-order Butterworth filter was applied to the resulting signals (cut-off frequencies of 0.7 and 2.5 Hz). The BVP signal was selected as the channel with the greatest frequency power in the range 0.7 and 2.5 Hz. From the recovered BVP signal, inter-beat intervals (IBIs) were detected using a peak detection algorithm [17]. The automatically identified pulse peaks were used to extract the heart rate and BVP signal-to-noise ratio (SNR) as explained below. The heart rate was defined as:

$$HR = \frac{60}{\overline{IBI}} \tag{9}$$

where $\overline{IBI}$ is the average inter-beat interval in seconds for a 30 second window.

POS Method [30]: To add additional weight to the results and demonstrate that the super-resolution preprocessing can benefit different iPPG algorithms we implemented the POS method presented by Wang et al. [30]. Our video framerate was 120 frames-per-second (FPS) and the window length was set to 192 frames. Within each window the spatially averaged color signals ($\mathbf{X} = [\mathbf{x}_1\ \mathbf{x}_2\ \mathbf{x}_3]$) were each normalized by dividing by their respective mean. Next, the normalized color matrix, $\hat{\mathbf{X}}$, was multiplied by the projection matrix *P* to give $\mathbf{Y}$, where:

$$P = \begin{pmatrix} 0 & 1 & -1 \\ -2 & 1 & 1 \end{pmatrix} \tag{10}$$

The window output was calculated as:

$$\mathbf{h} = \mathbf{Y_1} + \frac{\sigma(\mathbf{Y_1})}{\sigma(\mathbf{Y_2})}.\mathbf{Y_2} \tag{11}$$

Finally, the estimated BVP signal was constructed by adding the overlapping window output signals together for each 30 second segment of video. A 6th-order Butterworth filter was applied to the model outputs (cut-off frequencies of 0.7 and 2.5 Hz for HR). From the pulse signal we extracted the heart rate using an FFT. The HR was chosen from the peak with the greatest power in the frequency domain between 0.7 and 2.5 Hz.

4. Metrics

We extracted the HR estimates from a set of 10 non-overlapping 30 second windows for each 5.5-minute video (we discarded the first and last 15 seconds of each video). This resulted in 250 observation windows. We evaluate the heart rate estimates using several performance metrics. Mean absolute error (MAE):

$$\text{MAE} = \frac{\sum_{i=1}^{N} |\text{HR}_i - \text{HR}_i|}{N} \tag{12}$$

and root mean squared error (RMSE):

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\text{HR}_i - \text{HR}_i)^2}{N}} \tag{13}$$

Where N is the total number of observation windows (250). We also calculate the Pearson's Correlation Coefficient, r.

Finally, to capture a measure of the BVP signal quality, without relying on peak detection for the HR estimation, we calculate the blood volume pulse signal-to-noise (SNR) ratio. The BVP SNR was calculated according to the method proposed by De Haan et al. [5]. The gold-standard HR frequency was determined from the manually corrected ECG peaks.

$$SNR = 10\log_{10}\left(\frac{\sum_{f=30}^{240}((U_t(f)\hat{S}(f))^2}{\sum_{f=30}^{240}(1-U_t(f))\hat{S}(f))^2)}\right) \tag{14}$$

Where $\hat{S}$ is the power spectrum of the BVP signal (S), *f* is the frequency (in BPM) and $U_t(f)$ is a binary template that is 1 for the heart rate region from HR-6BPM to HR+6BPM and its first harmonic region from 2*HR-12BPM to 2*HR+12BPM, and 0 elsewhere.

5. Data

We used the dataset collected by Estepp et al. [8] for testing. Videos were recorded with a Basler Scout scA640-120gc GigE-standard, color camera, capturing 8-bit, 658x492 pixel images, 120 FPS. For the purposes of our experiments these videos were down-sampled (spatially compressed) to 41×30 pixels. Examples of an input frame at this resolution can be seen in Figure 3. The pixelation is clear in the frame at this size. To our knowledge this is the first work to attempt to recover the PPG signal from such a

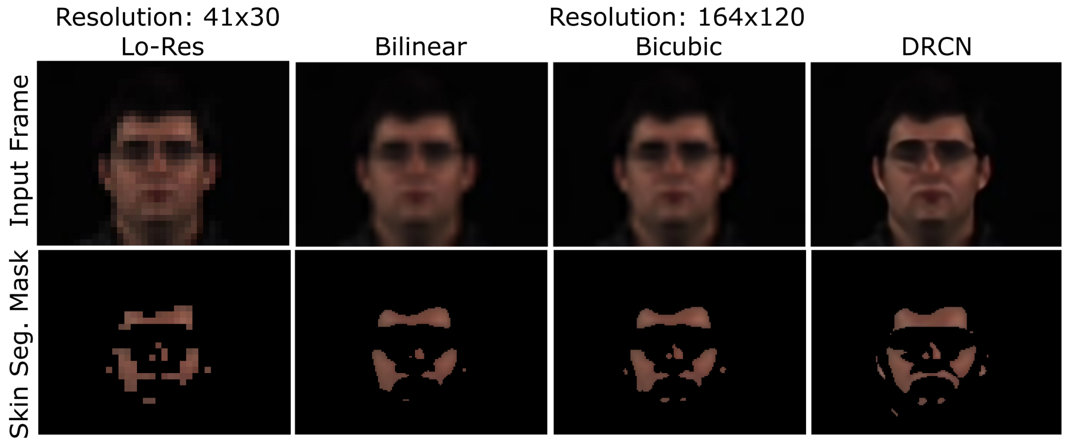

Figure 3: Examples of the low resolution (Lo-Res) input frames (41×30 pixels), bilinear and bicubic interpolated frames and the DRCN output frames (164×120 pixels). The segmented skin regions for each frame are shown below.

low resolution image patch. The camera was equipped with a 16 mm fixed focal length lens. Twenty-five participants (17 males) were recruited to participate for the study. Nine individuals were wearing glasses, eight had facial hair, and four were wearing makeup on their face and/or neck. The participants exhibited the following estimated Fitzpatrick Sun-Reactivity Skin Types [9]: I-1, II-13, III-10, IV-2, V-0. Gold-standard contact physiological signals were measured using a BioSemi ActiveTwo research-grade biopotential acquisition unit. The participants were recorded during a five-minute task. The task features head motion (rotation about the vertical axis) at a constant angular velocity of 10 degrees/sec. The five minute videos of the twenty-five participants resulted in over two hours of recordings.

6. Results

Using our video dataset we compared the performance of the set of image interpolation and super-resolution methods applied prior to the iPPG methods. In each case we upsampled all the frames by a factor of four. Alongside the deeply-recursive convolutional network (DRCN) we chose the following baseline methods: 1) no upsampling (original low resolution frame, 41×30 pixels), 2) bilinear interpolation (to 164×120 pixels), 3) bicubic interpolation (to 164×120 pixels). All the analysis was performed in MATLAB (Mathworks, Inc.) and the bilinear and bicubic interpolation methods were those implemented in the Computer Vision Toolbox. We present the results using each interpolation method and each of the two iPPG methods (ICA and POS). Table 1 shows the correlation coefficient, MAE, RMSE and BVP SNR for each method. The results obtained using the original high resolution (Hi-Res) frames are also shown. Figure 5 shows scatter plots of the camera HR and gold-standard contact HR for each case. All image interpolation/super-resolution methods (bilinear, bicubic, DRCN) improved the heart rate estimates. The DRCN method outperformed the bilinear and bicubic methods with a correlation of 0.87 and MAE and RMSE of 3.1 BPM and 4.8 BPM respectively when using the ICA iPPG method. A similar trend was observed when using the POS iPPG method.

To help establish whether the deeply-recursive super resolution method improves the skin segmentation and/or the color inference in the missing pixels leading to improved spatially averaged color signals (x_1, X_2, x_3) we performed a second experiment. We used the skin segmentation mask extracted from the DRCN output images and applied that to the bicubic interpolated images. We call this the DRCN-Bicubic hybrid. We compare these results with those from the Bicubic and DRCN methods separately. A diagram comparing the DRCN method with the DRCN-Bicubic hybrid is shown in Figure 4. The results are shown in Table 1. The DRCN-Bicubic hybrid method outperformed the Bicubic method showing that the DRCN super resolution does indeed improve the skin segmentation step. The DRCN method outperformed the DRCN-Bicubic hybrid method showing that the skin segmentation alone was not the only improvement, but that the pixel color values resulting from DRCN super resolution provide more accurate spatially averaged color signals.

7. Discussion

In many applications of iPPG it cannot be guaranteed that the skin ROI will occupy a large number of pixels. Reasons for this may be that the subject is a long distance from the imager, the imager has a low pixel density, or that the video has been spatially downsized (intra-frame compression) to reduce the storage volume required. However, to

iPPG Method	ICA				POS			
		HR		**BVP**		**HR**		**BVP**
Upsampling	Corr.	MAE	RMSE	SNR	Corr.	MAE	RMSE	SNR
Original (Hi-Res)	0.871	2.85	4.32	0.078	0.813	3.57	4.89	-0.008
None (Lo-Res)	0.753	4.41	6.79	-0.106	0.735	4.97	6.91	-0.079
Linear	0.803	3.56	5.45	**0.054**	0.805	3.54	5.05	-0.021
Bicubic	0.820	3.30	5.12	0.051	0.803	3.47	4.76	-0.016
DRCN-Bicubic	0.848	3.30	4.95	0.023	0.830	3.27	4.63	-0.026
DRCN	**0.866**	**3.08**	**4.82**	0.007	**0.840**	**3.12**	**4.45**	**-0.013**

Table 1: Summary of the overall results. Heart rate correlation (corr), mean absolute error (MAE), root mean squared error (RMSE) and blood volume pulse signal-to-noise ratio (SNR). The gold-standard HR measurement was taken from the manually corrected ECG signals.

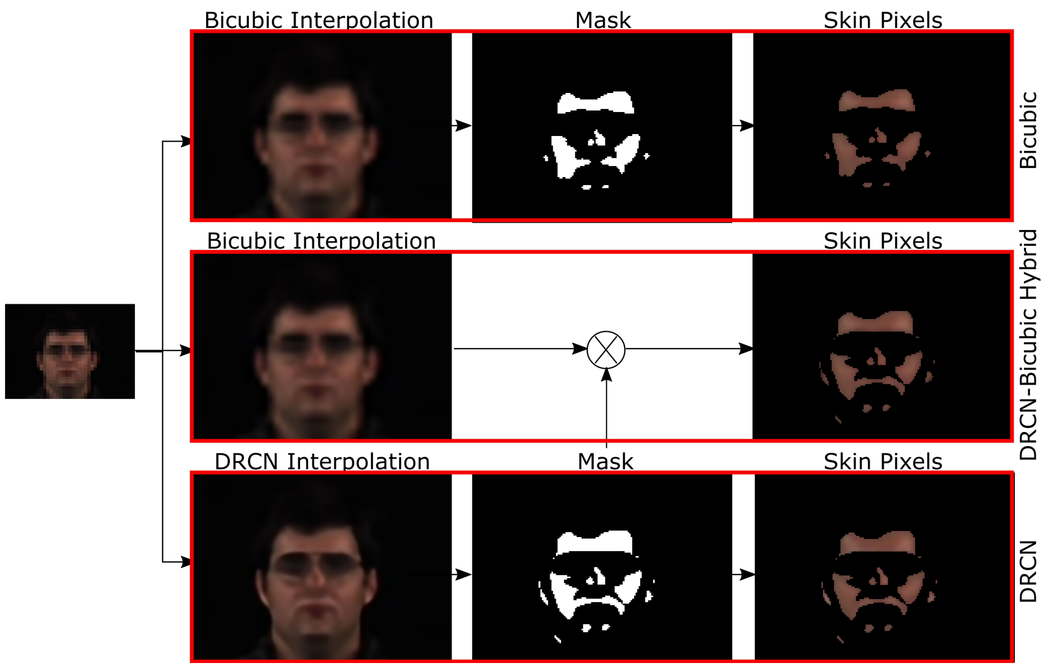

Figure 4: To test whether the improvements using the DRCN preprocessing were a result of improved skin segmentation or color estimation we used the skin segmentation mask extracted from the DRCN output images and applied that to the Bicubic interpolated images (DRCN-Bicubic hybrid). We compare this to using the Bicubic or DRCN upsampling alone. Our results showed that DRCN outperformed both the Bicubic and DRCN-Bicubic hybrid methods illustrating that the DRCN method does improve skin segmentation and color estimation.

date most work has focused on analysis of standard or high definition video in which the face occupies a large proportion of the frame and thus there are a high number of skin pixels (typically thousands).

Our results show that if frames are interpolated before iPPG analysis the accuracy of the recovered physiological measurements can be improved considerably. All the interpolation/super resolution methods we tried (bilinear, bicubic and DRCN) reduced the MAE and RMSE of the HR predictions compared to when using the low resolution (Lo-Res) frames and improved the correlation with the gold-standard contact measurements. The MAE was reduced by 19.3%, 20.6% and 30.2% using the bilinear, bicubic and DRCN methods respectively. Thus, the deeply-recursive convolutional network reduced error by a further 10% compared to the traditional interpolation approaches. Similar improvements were seen when using both iPPG methods (ICA and POS) giving confidence that our preprocessing algorithm is not limited to one iPPG approach.

Spatial upsampling of images, whether learned or not, helps with iPPG analysis as the skin region of interest can be segmented more accurately, reducing noise from background pixels, clothing or hair. We have shown examples of frames generated using each method to illustrate this (see

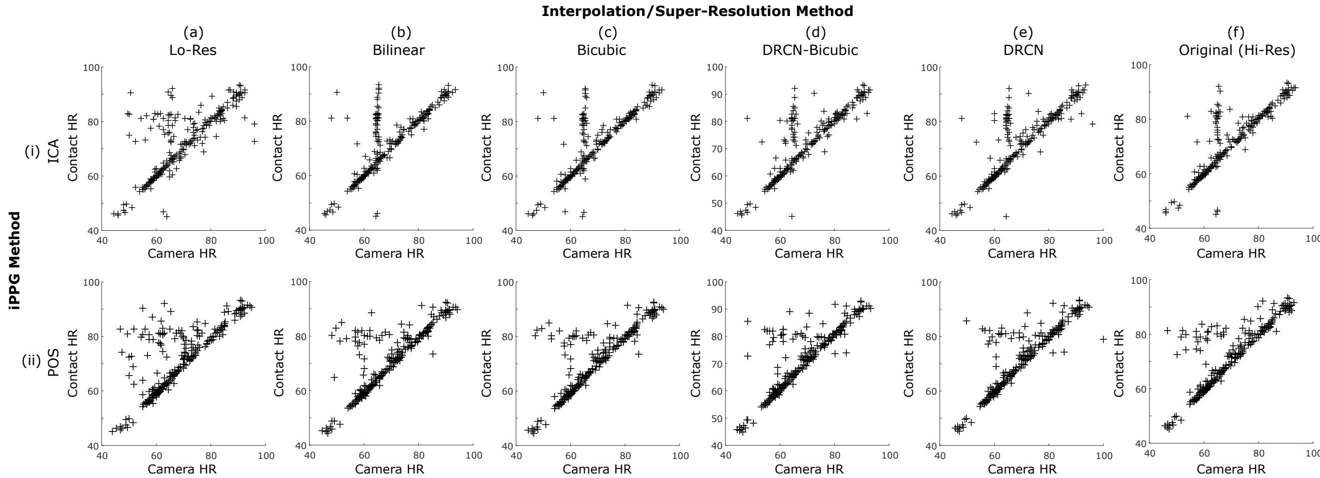

Figure 5: Scatter plots of the camera HR and gold-standard contact ECG HR. From left to right: Using different preprocessing interpolation/super-resolution methods a) None, b) Bilinear, c) Bicubic, d) DRCN-Bicubic, e) DRCN, f) the original image. From top to bottom: Using different iPPG methods i) ICA, ii) POS.

Figure 3). We questioned whether the improvement of the DRCN method over traditional interpolation was due to better skin segmentation alone or improved color inference in the skin pixels resulting in more accurate spatially averaged color signals (i.e., with a higher BVP signal-to-noise ratio). Our results suggest that DRCN provides benefits in both cases, even when the DRCN model was not trained on similar images to those in our dataset.

While the DRCN super resolution network provides the best results in terms of accuracy, the model does have a high computational complexity relative to bilinear or bicubic interpolation. Our results have shown that simple interpolation methods can still improve the accuracy of iPPG measurements considerably (reducing error by 20%). Bicubic interpolation is a simple and fast preprocessing step that could be used with any existing iPPG method and potentially used in real-time applications and on resource constrained devices (e.g., VR/AR headsets [19] and smartphones [11]).

8. Conclusions

In this work we present a new pipeline for non-contact measurement of physiological information incorporating a deep super resolution preprocessing step. This method presents new possibilities for extracting vital signs from compressed or low-resolution videos. Our results show that heart rate measurements and the BVP SNR can be improved using our method. Input videos of 41×30 pixel resolution were upsampled by a factor of four prior to iPPG processing. The low resolution input frames gave a MAE in HR estimates of 4.41 BPM. Adding a super resolution step reduced the error by over 30% to 3.08 BPM. The super resolution method improves both skin segmentation and color signal recovery. We used a pretrained deeply-recursive convolutional network model. The model was trained on a completely independent set of images that did not resemble our data. This gives high confidence about the generalizability of our result to other iPPG datasets. In addition, we showed that preprocessing improved results using two different iPPG algorithms. There might be the potential for additional performance gains if the network was trained on face images more closely resembling our data. Future work will consider if video super-resolution can also be used to combat the effects of inter-frame (temporal) compression within a video. If this were successful it would be very useful as inter-frame compression algorithms are particularly problematic for iPPG methods [15].

References

[1] J. Allen. Photoplethysmography and its application in clinical physiological measurement. *Physiological measurement*, 28(3):R1, 2007.

[2] E. B. Blackford, J. R. Estepp, A. M. Piasecki, M. A. Bowers, and S. L. Klosterman. Long-range non-contact imaging photoplethysmography: cardiac pulse wave sensing at a distance. In *Optical Diagnostics and Sensing XVI: Toward Point-of-Care Diagnostics*, volume 9715, page 971512. International Society for Optics and Photonics, 2016.

[3] H. Chang, D.-Y. Yeung, and Y. Xiong. Super-resolution through neighbor embedding. In *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, volume 1, pages I–I. IEEE, 2004.

[4] R. Dahl, M. Norouzi, and J. Shlens. Pixel recursive super resolution. *arXiv preprint arXiv:1702.00783*, 2017.

[5] G. de Haan and V. Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.

[6] G. de Haan and A. van Leest. Improved motion robustness of remote-ppg by using the blood volume pulse signature. *Physiological measurement*, 35(9):1913, 2014.

[7] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[8] J. R. Estepp, E. B. Blackford, and C. M. Meier. Recovering pulse rate during motion artifact with a multi-imager array for non-contact imaging photoplethysmography. In *Systems, Man and Cybernetics (SMC), 2014 IEEE International Conference on*, pages 1462–1469. IEEE, 2014.

[9] T. B. Fitzpatrick. The validity and practicality of sun-reactive skin types i through vi. *Archives of dermatology*, 124(6):869–871, 1988.

[10] J. Kim, J. Kwon Lee, and K. Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016.

[11] S. Kwon, H. Kim, and K. S. Park. Validation of heart rate extraction using video imaging on a built-in camera system of a smartphone. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 2174–2177. IEEE, 2012.

[12] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint*, 2016.

[13] M. H. Li, A. Yadollahi, and B. Taati. A non-contact vision-based system for respiratory rate estimation. In *Engineering in Medicine and Biology Society (EMBC), 2014 Annual International Conference of the IEEE*, pages 2119–2122. IEEE, 2014.

[14] D. McDuff, E. Blackford, J. Estepp, and I. Nishidate. A fast non-contact imaging photoplethysmography method using a tissue-like model. In *SPIE BiOS*. International Society for Optics and Photonics, 2018.

[15] D. McDuff, E. B. Blackford, and J. R. Estepp. The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography. In *Automatic Face and Gesture Recognition (FG), 2015 11th IEEE International Conference and Workshops on*, pages 1–8. IEEE, 2017.

[16] D. McDuff, J. R. Estepp, A. M. Piasecki, and E. B. Blackford. A survey of remote optical photoplethysmographic imaging methods. In *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6398–6404. IEEE, 2015.

[17] D. McDuff, S. Gontarek, and R. Picard. Improvements in remote cardio-pulmonary measurement using a five band digital camera. *IEEE Transactions on Biomedical Engineering*, 61(10):2593 – 2601, 2014.

[18] D. McDuff, S. Gontarek, and R. Picard. Remote measurement of cognitive stress via heart rate variability. In *Engineering in Medicine and Biology Society (EMBC), 2014 Annual International Conference of the IEEE*, pages 2957–2960. IEEE, 2014.

[19] D. McDuff, C. Hurter, and M. Gonzalez-Franco. Pulse and vital sign measurement in mixed reality using a hololens. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, page 34. ACM, 2017.

[20] I. Nishidate, N. Tanaka, T. Kawase, T. Maeda, T. Yuasa, Y. Aizu, T. Yuasa, and K. Niizeki. Noninvasive imaging of human skin hemodynamics using a digital red-green-blue camera. *Journal of biomedical optics*, 16(8):086012, 2011.

[21] M.-Z. Poh, D. McDuff, and R. W. Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics Express*, 18(10):10762–10774, 2010.

[22] M.-Z. Poh, D. McDuff, and R. W. Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE Transactions on Biomedical Engineering*, 58(1):7–11, 2011.

[23] S. Schulter, C. Leistner, and H. Bischof. Fast and accurate image upscaling with super-resolution forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3791–3799, 2015.

[24] J. Sun, Z. Xu, and H.-Y. Shum. Image super-resolution using gradient profile prior. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[25] C. Takano and Y. Ohta. Heart rate measurement based on a time-lapse image. *Medical engineering & physics*, 29(8):853–857, 2007.

[26] M. P. Tarvainen, P. O. Ranta-Aho, P. A. Karjalainen, et al. An advanced detrending method with application to hrv analysis. *IEEE Transactions on Biomedical Engineering*, 49(2):172–175, 2002.

[27] S. Tulyakov, X. Alameda-Pineda, E. Ricci, L. Yin, J. F. Cohn, and N. Sebe. Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2396–2404, 2016.

[28] W. Verkruysse, L. O. Svaasand, and J. S. Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008.

[29] M. Villarroel, A. Guazzi, J. Jorge, S. Davis, P. Watkinson, G. Green, A. Shenvi, K. McCormick, and L. Tarassenko. Continuous non-contact vital sign monitoring in neonatal intensive care unit. *Healthcare technology letters*, 1(3):87, 2014.

[30] W. Wang, A. Den Brinker, S. Stuijk, and G. De Haan. Algorithmic Principles of Remote-PPG. *IEEE Transactions on Biomedical Engineering*, PP(99):1–12, 2016.

[31] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010.