

## Article

# Improving Accuracy of Contactless Respiratory Rate Estimation by Enhancing Thermal Sequences with Deep Neural Networks

Alicja Kwasniewska <sup>1,2,\*</sup> , Jacek Ruminski <sup>1,\*</sup>  and Maciej Szankin <sup>2,\*</sup> 

<sup>1</sup> Department of Biomedical Engineering, Faculty of Electronics, Telecommunications and Informatics, Gdansk University of Technology, Gabriela Narutowicza 11/12, 80-233 Gdansk, Poland

<sup>2</sup> Artificial Intelligence Products Group, Intel Corporation, 12220 Scripps Summit Dr, San Diego, CA 92131, USA

\* Correspondence: alicja.kwasniewska@pg.edu.pl (A.K.); jacek.ruminski@pg.edu.pl (J.R.); maciej.szankin@intel.com (M.S.); Tel.: +1-210-461-6553 (A.K.)

Received: 15 September 2019; Accepted: 9 October 2019; Published: 17 October 2019



**Featured Application:** The proposed Super Resolution Deep Neural Network allows for improving accuracy of Respiratory Rate (RR) estimation from extremely low resolution thermal sequences, i.e.,  $40 \times 30$  pixels. To the best of our knowledge deep learning hasn't been used for telemedicine use cases aimed at vital signs monitoring before. Thus, there are many potential applications where it can be useful, i.e., remote diagnostics using smart home platforms, long-distance vital signs monitoring in difficult to reach areas using cameras mounted on drones, monitoring of driver's and passengers' state of health in self-driving vehicles, emotions recognition from vital signs, or detecting unusual behaviors e.g., abnormal respiratory rate patterns at security checkpoints.

**Abstract:** Estimation of vital signs using image processing techniques have already been proved to have a potential for supporting remote medical diagnostics and replacing traditional measurements that usually require special hardware and electrodes placed on a body. In this paper, we further extend studies on contactless Respiratory Rate (RR) estimation from extremely low resolution thermal imagery by enhancing acquired sequences using Deep Neural Networks (DNN). To perform extensive benchmark evaluation, we acquired two thermal datasets using FLIR<sup>®</sup> cameras with a spatial resolution of  $80 \times 60$  and  $320 \times 240$  from 71 volunteers in total. In-depth analysis of the proposed Convolutional-based Super Resolution model showed that for images downsampled with a factor of 2 and then super-resolved using Deep Learning (DL) can lead to better RR estimation accuracy than from original high-resolution sequences. In addition, if an estimator based on a dominating peak in the frequency domain is used, SR can outperform original data for a down-scale factor of 4 and images as small as  $20 \times 15$  pixels. Our study also showed that RR estimation accuracy is better for super-resolved data than for images with color changes magnified using algorithms previously applied in the literature for enhancing vital signs patterns.

**Keywords:** respiratory rate; remote medical diagnostics; vital signs estimation; super resolution; deep learning; convolutional neural networks

## 1. Introduction

Significant technological progress in image processing algorithms and ability to improve perception of the world surrounding us using modern deep learning methods has led to invention and enablement of various applications that were previously much harder to implement, e.g., face

recognition [1,2], objects detection and segmentation [3–5] or image resolution enhancement [6]. Another factor that steers the direction of advances in new applications is the increased capabilities of many electronic digital devices. Medical diagnostics and reasoning systems also benefited from this progress, allowing for real-time vital signs analysis and tracking using standard cameras [7] or wearable devices [8].

The remote measurement of RR has many potential applications in medical diagnostics and screenings like monitoring of newborns or small children in incubators or hospital beds [9], monitoring of the severe acute respiratory syndrome (SARS) [10], support in emotion analysis [11], etc. The use of thermal cameras was proposed to analyze facial images and estimate RR based on a nasal heat flow and described in [9,12–17] and other papers. The typical estimation of RR requires a multi step procedure. First, a Region Of Interest (ROI) is detected for each thermal frame representing the source of thermal changes (due to respiration) in the area of nostrils or a mouth. The ROI can be specified manually or can be automatically detected (in a frame or for each frame) and tracked (between frames). In [18], the authors described a particle filter tracker driven by a probabilistic template function that was capable of adapting to abrupt positional and physiological changes. Other authors also proposed ROI tracking methods (e.g., [19,20]). In the next step, a single value is calculated to represent each ROI. A collection of such values forms a signal representing local temperature changes in time. Finally, a signal is filtered (e.g. removing high frequency components) and a frequency for dominated changes is calculated. It is assumed, that this frequency represents the respiratory rate.

In recent years, new portable and cost-effective thermal cameras have been available. For example, FLIR® Lepton family cameras [21] are very small (e.g.,  $10.5 \times 11.7 \times 6.4$  mm, with an internal shutter) and cost less than 200USD. These features allow to think about wide application of thermal monitoring, e.g., to support remote diagnosis of elderly people at home (e.g., during a video talk or as a self-diagnostics). However, a spatial resolution of these small thermal cameras is as low as  $80 \times 60$  or  $160 \times 120$ . Small resolution of images could be a problem for detection of facial features or detection of a ROI representing respiration-related temperature changes. Different methods have been proposed in (a visible light spectrum) computer vision to improve low resolution images or to detect (and amplify) small, local changes in videos.

Subtle intensity variations introduced in thermal videos of a face due to respiratory activities can be enhanced using Eulerian Video Magnification (EVM [22]) or related algorithms [23,24]. This technique amplifies intensity differences within a particular frequency spectrum. This works well if the estimate respiratory rate (frequency) is known. Otherwise, noise and motion artefacts are highly amplified. Therefore, some researchers propose to magnify only selected segments within a video [25]. The EVM algorithm has been already successfully used for enhancing vital sign signals [26–28].

Recently, many different deep-learning algorithms for super resolution have been proposed. It has been proved that such algorithms can efficiently improve the presentation of details in the processed low-resolution (LR) visible light images. One of the first method in this area was SRCNN [29], which implemented a single Convolutional Neural Network (CNN) achieving the state-of-the-art restoration quality. Later, different Deep Neural Networks (DNNs) based solutions have been introduced to further improve the restoration quality (or perception). In [30] Kim et al. introduced a novel Deeply Recursive Convolutional Network (DRCN) model. It utilizes a skip connection correlating a LR input with a high resolution (HR) reference data and uses recursive supervision to minimize the exploding/vanishing gradients problem. Other improvements to SR include the application of residual mappings and gradient clipping (Deeply Recursive Residual Network (DRRN) [31]), the use of attention networks [32], the application of multi-scale residual hierarchical networks [33], etc. The very good results have also been obtained using SR algorithms based on generative networks [6,34–36].

DNN-based algorithms have also been applied to process thermal, infrared images. Cho et al. [37] proposed simple CNN-based models to classify different raw materials from short range thermal images. Other state-of-the-art CNN models were used in different thermal imaging applications,

e.g., SqueezeNet based Thermalnet [38], denoising CNN [39], etc. Some authors applied DNN models to improve a thermal image contrast. For example, in [40] authors proposed a conditional GAN-based network to address the thermal imaging contrast enhancements. The experiments performed on thermal images obtained for indoor and outdoor scenes showed much better results in comparison to traditional histogram equalization techniques and other image transformation methods. Only limited number of papers present the results of SR algorithms applied to thermal images. In [41] authors proposed a cascaded architecture of deep networks with multiple receptive fields. They investigated the large-scale resolution improvement ( $8\times$ ) using thermal images of indoor and outdoor scenes. However, the training dataset for the proposed cascaded deep networks was very small (100 images). In [42] authors applied the Comprehensive Sensing (CS) Theory to the SRCNN-like model structure to improve high frequency information and alleviate some fixed pattern noise present in the output from CS. However, the model was trained using 400 visible light images and tested only on 6 (outdoor) thermal images. In this paper, we combine the use of super resolution algorithms for thermal image sequences of faces to investigate a possible accuracy improvement in the respiratory rate estimation task.

Our work is probably the first attempt to the improvement of contactless RR estimation using SR DNN specifically designed for thermal data, since most of previous solutions focus on visible light images that have different characteristics than thermal images. Our model is characterized by a wider receptive field which allows for taking into account more distant relations between neighbouring facial areas due to the blurring effect in thermal images. Some attempt to this problem has been done by us in our earlier preliminary work [43], where we showed that the use of SR DNNs can improve the RR estimation accuracy. Yet, in this previous study we only analysed one RR estimator on a single dataset. To avoid being biased by a specific method and data, in this study, we extend our experiments to different RR estimators and evaluate the proposed solution on two thermal datasets collected by us for 71 volunteers.

In particular, our contributions are summarized as follows: (a) we analyze the effect of the super-resolution algorithm on two different datasets of thermal sequences (with original resolutions:  $80 \times 60$  and  $320 \times 240$ ), (b) we investigate if the processing of super-resolved thermal sequences allows for improving the RR estimation in comparison to RAW data processing, (c) we compare accuracy of the RR estimation from images super-resolved with our DL model to the results of sequences with magnified color changes using the Eulerian Video Magnification algorithm, (d) we analyze the role of thermal image bit-depths on the RR estimation accuracy and on image quality metrics (Peak Signal to Noise Ratio and Structural Similarity Index) for sequences with resolution degradation and enhancement.

The rest of the work is organized into following sections: Section 2 provides details about a data collection and pre-processing process, as well as resolution degradation and enhancement methods. In Section 3 we present achieved preliminary results of calculated image quality metrics and errors of RR estimation. These results are further discussed in Section 4. Readers may access the code and trained models used in our study at the link provided in Section 5. Finally, Section 6 concludes our work and specifies ideas for future improvements.

## 2. Methodology

### 2.1. Respiratory Rate Estimation

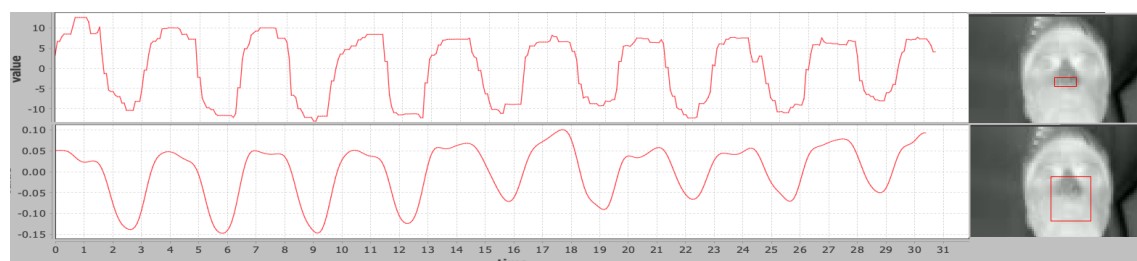
The contactless estimation of respiratory rate (RR) has revolutionized a conventional measurement of vital signs, usually obtained using wires and electrodes attached to the skin. Various studies have already proved that it's possible to accurately extract RR from sequences of a very small spatial resolution, e.g.,  $80 \times 60$  [16]. In this study, we evaluate two respiratory rate estimators proposed and described in details in [44]:  $eRR_{sp}$  and  $eRR_{as}$ . In the  $eRR_{sp}$  estimator we assume that the respiratory signal is dominant in the signal spectrum. Thus, the detection of the maximum value in the frequency

domain (the frequency value of the dominating peak) can be used for the RR estimation. The main drawback of this estimator is the fact that it always returns the result even for non-respiratory signals (such as noise) so it's advisable to use in together with other methods for reliable results. The second estimator utilizes the fact that the periodic signal and the auto-correlation sequence of it have the same cyclic characteristics. Calculation of the auto-correlation for following time lags and similar analysis as for *eRR<sub>sp</sub>* allows for estimating of RR values. Since this estimator uses the autocorrelation spectrum it is further referred to as *eRR<sub>as</sub>*.

Our primary contribution lies in applying RR estimators to thermal sequences previously enhanced with Convolutional Neural Network based Super Resolution (SR) model to verify if the accuracy of RR estimation can be improved with deep learning (DL). The DL model proposed by us is compared against the Eulerian Video Magnification (EVM) algorithm [22]. In addition, we also examine the influence of resolution degradation on the accuracy of RR estimation by down-scaling thermal sequences with bicubic interpolation.

In this work, we utilized short data segments (300–400 samples) for RR values extraction. The limited window size enables the use of contactless RR estimation in many practical applications due to the short acquisition and processing time. In addition it allows for reducing possible motion artifacts, as the person doesn't have to be still for a very long time.

At first, we analyzed the original sequences without spatial resolution modifications. The raw breathing signal was constructed by aggregating pixel values within the manually selected region of interest (ROI) using the average or the skewness operator. As previously verified in [44], if the skewness aggregation is applied, the extracted waveforms are practically insensitive to the size of the ROI, assuming the whole area of a nose is marked. On the other hand, the averaging operation applied to bigger regions leads to the smoothing effect of the respiratory-related pixel values changes, making the extraction of RR values almost impossible. Taking it into account, in case of the skewness operator, we selected the ROI which was covering the whole nose, while for the averaging it was covering a smaller area (e.g. nostrils only), as presented in Figure 1.



**Figure 1.** Comparison of the region of interest (ROI) selection for different aggregation operations used to obtain raw signals from which RR values were extracted. **Top:** average; **bottom:** skewness aggregation operator - the selected ROI and the extracted signal.

The obtained raw signals were filtered using moving average filter and the 4th-order high pass Butterworth filter (the cutoff frequency set to 0.125 Hz). Then, both RR estimators were applied in order to extract respiratory values. The same procedure was used for all other sequences, i.e., EVM, SR and bicubic without changing the location of the selected ROI to compare influence of resolution enhancement/degradation on the value of the estimated vital sign. In this way, we were able to evaluate how RR changes if a spatial resolution is decreased or improved with image processing techniques, as the area used for obtaining signals remained constant.

The Root Mean Square Error was calculated using a difference between values of RR computed with RR estimators and a reference RR value. The process of obtaining thermal sequences and reference respiratory signals is described in details in the following section.

## 2.2. Data Collection and Pre-Processing

In this section, we provide details about datasets used in this study, as well as algorithms used for resolution degradation and enhancement.

### 2.2.1. Datasets

The aim of this study is to evaluate face hallucination DL networks on images acquired in thermography, i.e., intensities of electromagnetic radiation in the range of 8–12  $\mu\text{m}$  (Long-Wave Infrared). In thermal imaging, electromagnetic radiation intensities are represented as digital values, usually of a 14-bit resolution. The final image is constructed by converting raw radiation values to temperature data by assigning color values to the digital ones.

The SC3000 database is a collection of thermal sequences recorded for 40 volunteers from a group of 21 females and 19 males of the age  $34.11 \pm 12$ . For data acquisition the FLIR® SC3000 thermal camera was used. To record sequences, the camera was placed on a tripod at the height of 112 cm from the ground and at the distance of 1.2 m from a face of the volunteer. The sequences were gathered over a period of 2 min each, during which volunteers were asked to look directly into the camera and stay possibly still. The model of the camera used for data collection is capable of recording temperature in a range from  $-20\text{ }^{\circ}\text{C}$  to  $+80\text{ }^{\circ}\text{C}$  in the  $320 \times 240$  pixels spatial resolution at 30 Frames Per Second (FPS). Data was captured in the noise reduction mode in the 14-bit radiometric format, which was then linearly down-scaled and saved as 8-bit grayscale PNG images, resulting in 3600 frames per person. In addition, to avoid contrast decrease by lossy data compression, we also generated 16-bit grayscale PNG images by up-scaling of the bit-depth to 16-bit float values (3600 frames per person). In order to obtain ground-truth measurements for respiratory rate, each volunteer was asked to bend the finger when exhaling air and to straighten it when inhaling in the way that it can be captured on the camera.

The second dataset used in our research, refereed thereafter as the Lepton dataset, consists of 31 1-min thermal sequences obtained from volunteers of the age  $26 \pm 8.1$ . For data capture, the FLIR® Lepton thermal camera was used. Due to the small form factor of the device (a circuit board is smaller than  $1\text{ cm}^2$ ), this camera module can be incorporated in embedded platforms, allowing for application in various medical oriented solutions [45,46]. The Lepton camera is equipped with a thermal sensor capable of recording data in the 14-bit dynamic range and the  $80 \times 60$  spatial resolution at 9 FPS. Acquired sequences were up-sampled in the post-processing step to 12 FPS in order to increase the number of samples. These frames were further used for our model training and it's crucial to provide deep neural networks with enough data to make correct predictions [47]. During the data capture the camera was placed at a distance of 0.5 m from a volunteer's face. Similarly to the SC3000 dataset, obtained raw data were linearly down-scaled and saved as 8-bit grayscale PNG images, resulting in 720 frames per person. Also, to represent data in the PNG image format while avoiding loss of information, we generated a second set of images (720 frames per person) with values represented as 16-bit float numbers. The reference respiratory rate measurements were collected using the respiratory monitor belt (Vernier RMB), which strapped around the chest can record the pressure associated with the expansion and contraction of the chest during breathing.

Both datasets were captured in a laboratory room at the ambient temperature of 23 to 27  $^{\circ}\text{C}$ . Every participant, who took part in the data collection process, was informed about the details of the experiment and the signed informed consent was obtained from each volunteer. The experiments were performed with agreements to rules specified by the regional, institutional Bioethical Commission.

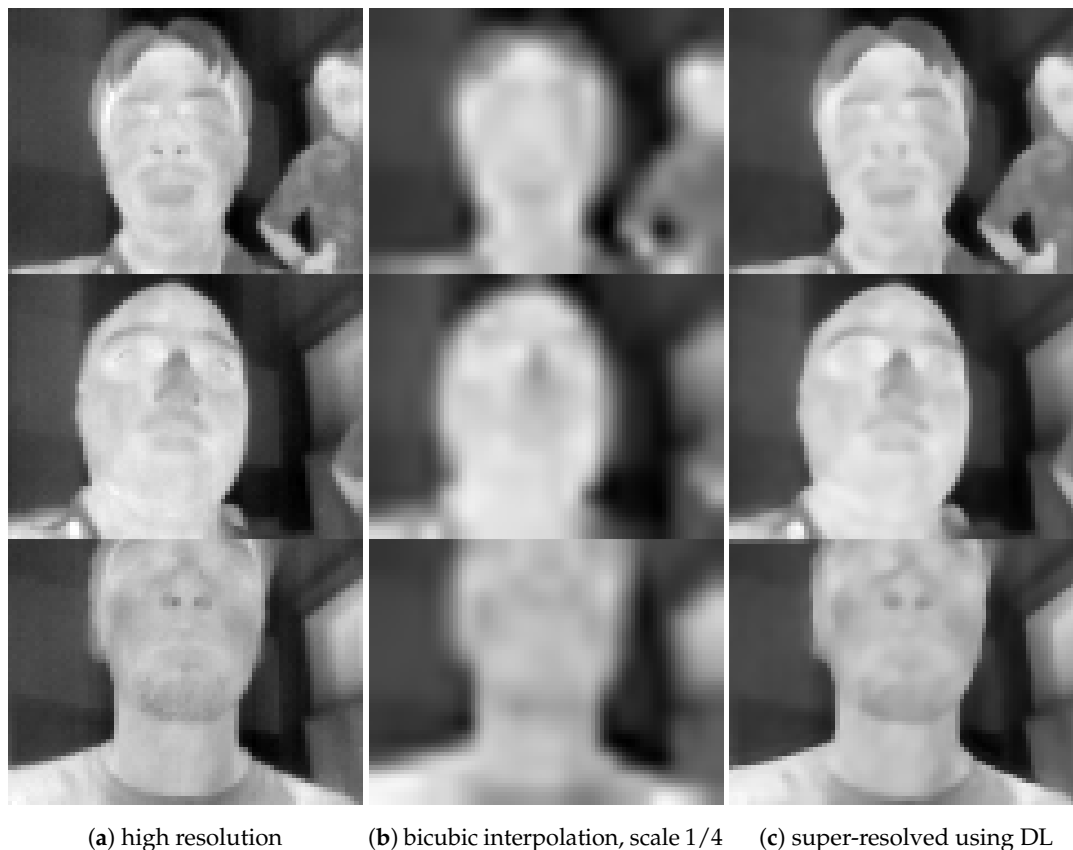
### 2.2.2. Image Resolution Degradation and Enhancement

The objective of Super Resolution (SR) methods is to restore an output as similar as possible to the original high resolution (HR) input by feeding to the network the low resolution (LR) image. Usually, we don't have access to two perfectly synchronized cameras with same parameters but different spatial resolutions: higher and lower. Thus, the common approach is to create a synthetic LR image using HR



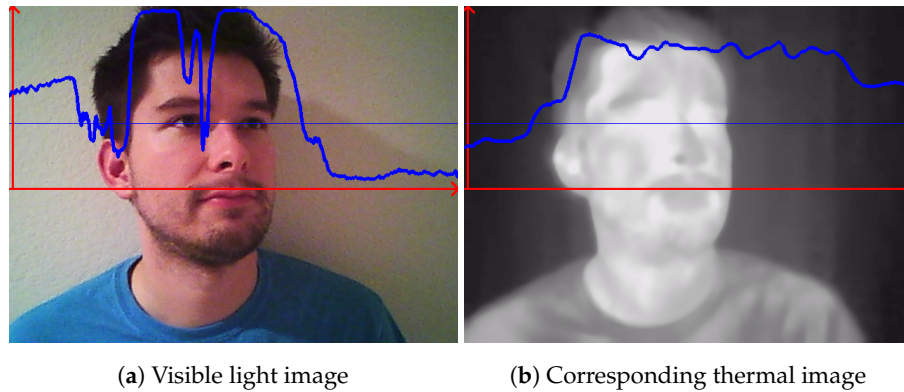
data. It is often achieved by down-scaling and then up-scaling of the HR image with inverse scales, i.e.,  $1/2$  and then  $2$ . In this way, the image has the same size but one can observe the effect of resolution degradation, as presented in Figure 2a,b.

In the case of Single Image Super Resolution (SISR) models that use convolutional filters, generated LR images are fed to the Neural Network (NN) that sequentially applies the stack of filters with weights adjusted during model training to restore high resolution features. If a stride parameter of convolutions is set to 1 and no pooling is used, the restored output has the same size as the LR input, but important components e.g., facial features are restored as shown in Figure 2c.

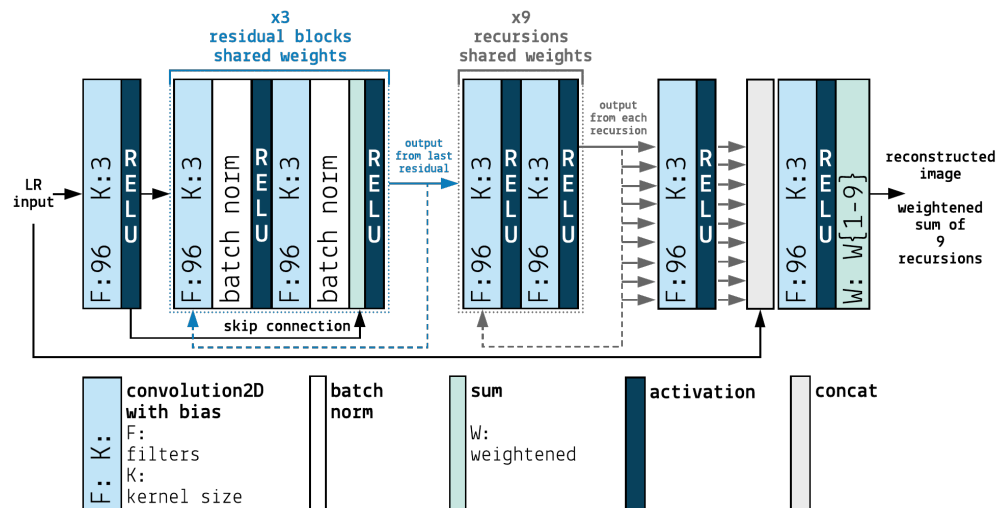


**Figure 2.** Examples of corresponding images from the Lepton 8-bit dataset.

The majority of DL solutions aimed at generating hallucinated face representations have been focused on visible light images [30,31,48]. Yet, it is very important to notice that thermal data have different characteristics than images acquired in the visible light. Due to the heat flow in objects, features present in thermal images are more blurred and a distance between specific facial areas is bigger, leading to a lower contrast and smoother color changes between them (see Figure 3). Hence, motivated by this finding, we designed our custom convolutional-based SR NN, which allows for widening of the receptive field without introducing new parameters. In this way, we show improvement of features restoration as the model gets more details about distant dependencies between interesting components. The architecture of the introduced SR NN is presented in Figure 4.



**Figure 3.** Comparison of pixel values differences (blue plot) at the eye level. Thermal and visible light image taken simultaneously using the FLIR® One Gen 2 camera. It can be observed that the thermal image is characterized by smoother changes and lower differences between adjacent pixels values than the visible light image. The figure presents one of the authors who has agreed to publish his photography.



**Figure 4.** Architecture of the proposed Convolutional-based Super Resolution model used in the study for enhancing thermal sequences from which respiratory rates were extracted.

The main difference of our network comparing to previous solutions is the use of residual blocks for building representation of the input image in a form of embeddings. Residual blocks have been already proved to ease the model optimization process, allowing for a significant depth increase without the risk of over-fitting [49]. After building the representation of the image, it is mapped to another vector representation once a non-linear activation function is applied. Intuitively, we may treat this mapping as the process of generating high-resolution patches from input representations, as suggested by the pioneer work in the area of applying DL to the SR task [29]. At this step, similarly to [48], we supervise the recursion to minimize the risk of vanishing gradient problem (in Figure 4 it is marked with all 9 gray outputs from recursions being passed to the final convolution operation).

In addition, it is usually advisable to correlate the LR image with the reconstructed output, as in general they are very similar except the lack of detailed features in the LR input. This correlation is realized in our model by the skip connection that concatenates the input and the restored outputs in the concat layer.

The network is updated using the Mean Square Error loss, defined as:

$$L(\theta) = \frac{1}{N} \sum_{i=1}^N \|Y_i - \hat{Y}_i\|^2 \quad (1)$$

where  $N$  is the number of images in the mini-batch used for gradient descent network updating,  $Y_i$  is the original HR image and  $\hat{Y}_i$  is the reconstructed output, calculated as the weighted sum of all recursions:

$$\hat{Y} = \frac{1}{D} \sum_{d=1}^D w^{(d)} (F_{rec}^{(d)} + X) \quad (2)$$

where  $D$  is the number of recursions,  $w^{(d)}$  is the weight associated with the output of the ( $d$ th) recursion  $F_{rec}^{(d)}$  and  $X$  is the LR input.

Since our model utilizes residual blocks in the *embedding* subnetwork and uses *supervised recursions*, we named it DRESNet - Deep Residual Embedding and Supervised-recursions. The weights of all residual and recursive blocks are shared, what allows for a significant depth increase while keeping the number of parameters constant. The final configuration of DRESNet was experimentally verified by us in [43], where we tested configurations with different number of residual and recursive blocks. In-depth analysis showed that the best results are achieved for 3 residuals in the *embedding* subnetwork and 9 recursions in the *mapping* subnetwork. Hence, this configuration was used in further experiments.

For the network training, we utilized data from first 15 volunteers of both the SC3000 and the Lepton dataset. Due to memory limits of the training hardware, only 4 images per volunteer from the SC3000 database and 20 images per volunteer from the Lepton database were randomly selected. We were able to select more images from the Lepton dataset, because each image was 4 times smaller than images in the SC3000 dataset. As a result, we built four sets: SC3000-8 with 60 8-bit images, SC3000-16 with 60 16-bit images, Lepton-8 with 300 8-bit images and Lepton-16 with 300 16-bit images. The proposed SR NN was trained on each of these sets separately, 10% of data was used for validation. As previously mentioned, to train the network we have to feed LR data and compare the reconstructed output to the original HR image. Thus, at first we generated synthetic LR inputs by down-scaling and then up-scaling original HR images with inverse scales. In this study, we performed experiments using 2 settings: scale 1/2 and 2; scale 1/4 and 4.

Hyperparameters for model training were tuned using the random search approach and final values were set as:  $41 \times 41$  patch size, patch stride of 21, Adam optimizer, momentum 0.9. In this study we used the weight decay approach with a decrease of 0.0001 and the learning rate decay approach with an initial value of the learning rate set to  $10^{-2}$  and a decrease of 0.1 every 5 subsequent epochs, for which a validation error hasn't decreased. Trained SR models were used to enhance all remaining thermal sequences (25 8-bit and 25 16-bit sequences from the SC3000 dataset; 16 8-bit and 16 16-bit sequences from the Lepton dataset) and verify how it affects accuracy of the RR estimation.

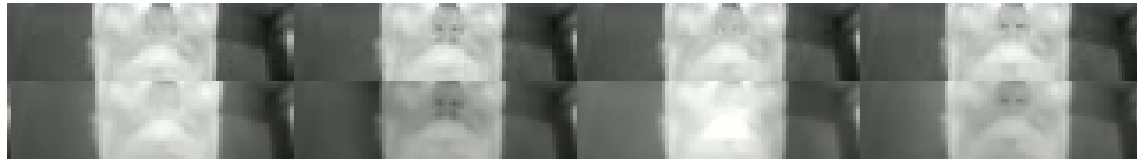
### 2.2.3. Color Changes Magnification

Since our study focuses on improving accuracy of the RR estimation by the image resolution enhancement, we compare the proposed DL-based solution with other existing algorithms aimed at enhancing breathing signals. Precisely, the robustness of the SR CNN network is evaluated together with the Eulerian Video Magnification (EVM) algorithm [22] on both datasets collected by us. EVM allows for visualizing information which is invisible to a naked eye, e.g., the blood flow or chest movements. It is achieved by spatial decomposition of video frames and temporal filtering. As a result, we get a signal that represents these hidden changes. To reveal them, the signal is magnified and added back to the input video.

In our work, the filtering frequency range (a parameter of EVM) was set to 0.16–0.33 Hz, assuming that a breathing rate of a healthy adult varies between 10–20 breaths per minute (bpm). The amplification factor of 20 was applied for signal magnification, as previously verified in [11].



All thermal sequences used for testing (i.e., sequences from volunteers 16th–31st for the Lepton and 16th–40th for the SC3000 dataset) were enhanced with EVM. Examples of frames from consecutive 2-sec windows (every 25th frame) with magnified color changes are presented in Figure 5. Frames were cropped to the nostril area, where the breathing signal is the most visible due to temperature differences between inhaled and exhaled air.



**Figure 5.** Example of frames extracted from the Lepton sequences enhanced with EVM algorithm. **Top:** original sequence, **Bottom:** enhanced sequence.

### 3. Results

Table 1 presents metrics used to verify the reliability of the proposed CNN-based SR model. Specifically, we calculated the Peak Signal to Noise Ratio (PSNR) and the Structural Similarity Index Metric (SSIM), often used for evaluation of SR solutions. In Table 2 we collected values of RR estimated from thermal image sequences. In both tables we present results for data after resolution degradation and enhancement. Down-scaling and then up-scaling of the image with inverse scales, i.e., 1/2 and 2 resulted in resolution degradation, producing the synthesized low-resolution (LR) output of the same size as the input image. In case of SR, after generating LR images they were enhanced with the proposed DL model. Enhancing of breathing signal was additionally performed by applying EVM—the color magnification algorithm. Results of RR estimation from sequences enhanced with EVM are also available in Table 2. Results of RR above 50 were treated as errors and not taken into account for RMSE calculation.

**Table 1.** Peak Signal to Noise Ratio (PSNR) and Structural Similarity Index Metric (SSIM) for sequences downsampled with bicubic interpolation and then enhanced with the proposed SR DL model using scale of 2 and 4. Best result for each dataset was marked in **bold** (separately for 8 and 16-bit data).

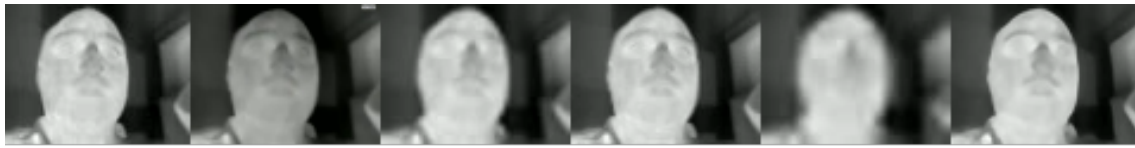
Scaling Factor	Dataset	Bit Resolution	Evaluation Metrics on the Test Set	
			PSNR	SSIM
1/2 and 2	Lepton Bicubic	8 bits	41.82 ± 0.55	0.96 ± 0.01
		16 bits	63.82 ± 0.73	0.99 ± 0.01
	Lepton SR	8 bits	<b>43.21 ± 0.33</b>	<b>0.97 ± 0.01</b>
		16 bits	<b>72.92 ± 4.47</b>	<b>0.99 ± 0.01</b>
1/4 and 4	Lepton Bicubic	8 bits	39.91 ± 0.46	0.81 ± 0.11
		16 bits	61.31 ± 0.69	0.99 ± 0.01
	Lepton SR	8 bits	42.18 ± 0.24	0.95 ± 0.01
		16 bits	67.34 ± 5.83	0.99 ± 0.02
1/2 and 2	SC3000 Bicubic	8 bits	42.69 ± 3.36	0.81 ± 0.22
		16 bits	68.98 ± 1.02	0.99 ± 0.01
	SC3000 SR	8 bits	43.61 ± 0.18	0.96 ± 0.01
		16 bits	<b>70.05 ± 0.91</b>	<b>0.99 ± 0.01</b>
1/4 and 4	SC3000 Bicubic	8 bits	41.36 ± 2.30	0.79 ± 0.21
		16 bits	65.74 ± 0.95	0.99 ± 0.01
	SC3000 SR	8 bits	<b>43.97 ± 0.22</b>	<b>0.96 ± 0.01</b>
		16 bits	66.50 ± 0.93	0.99 ± 0.01

**Table 2.** Root Mean Square Error calculated using a difference between reference RR values and RR values estimated from thermal sequences using two respiration rate estimators: *eRR<sub>sp</sub>* and *eRR<sub>as</sub>*. Raw pixel values inside manually selected regions of interests were aggregated using average or skewness operators. Best result for each dataset was marked in **bold** (separately for 8 and 16-bit data).

Scaling Factor	Dataset	Bit Resolution	RMSE of Respiratory Rate Estimation			
			RR Estimator		Aggregation Operation	
			eRR <sub>sp</sub> Average	eRR <sub>sp</sub> Skewness	eRR <sub>as</sub> Average	eRR <sub>as</sub> Skewness
0	Lepton	8 bits	4.97	6.28	15.61	12.80
		16 bits	5.15	6.35	5.68	7.39
	Lepton * EVM	8 bits	5.58	7.04	9.40	11.94
		16 bits	5.58	6.81	7.98	11.28
1/2 and 2	Lepton Bicubic	8 bits	5.66	7.21	9.14	7.96
		16 bits	4.93	7.20	8.08	7.77
	Lepton SR	8 bits	<b>4.89</b>	<b>5.64</b>	<b>4.95</b>	<b>6.21</b>
		16 bits	4.93	6.72	6.29	7.41
1/4 and 4	Lepton Bicubic	8 bits	5.61	7.64	8.57	7.90
		16 bits	6.40	7.32	8.37	7.34
	Lepton SR	8 bits	4.96	5.93	7.41	12.25
		16 bits	<b>4.89</b>	6.10	10.31	8.00
0	SC3000	8 bits	3.48	3.59	17.19	11.06
		16 bits	3.61	5.61	12.11	14.72
	SC3000 EVM	8 bits	5.00	6.15	5.98	7.82
		16 bits	4.56	6.09	5.84	7.65
1/2 and 2	SC3000 Bicubic	8 bits	6.35	6.04	17.05	11.26
		16 bits	5.91	8.46	34.43	14.52
	SC3000 SR	8 bits	<b>2.94</b>	<b>2.46</b>	<b>5.56</b>	<b>4.27</b>
		16 bits	4.09	3.59	8.39	8.90
1/4 and 4	SC3000 Bicubic	8 bits	5.73	8.23	17.05	11.65
		16 bits	5.73	6.32	14.31	11.35
	SC3000 SR	8 bits	3.48	5.11	13.36	12.12
		16 bits	3.48	5.38	14.48	14.61

\* Lepton dataset had to be upscaled due to requirements of the EVM tool [50], which does not accept input with spatial resolution smaller than  $100 \times 100$  pixels.

Qualitative results of applying each algorithm on sequences from both datasets are visualized in Figures 6 and 7. The restoration of facial features with the proposed model from images downsampled 2 and 4 times is presented in the last frame and the third to last frame, respectively in both figures. The image enhancement is especially visible for the Lepton dataset which consist of images of very small spatial resolution ( $80 \times 60$ ). One can notice that the LR image created for the scale 4 (second to the last image in Figure 6) is almost completely blurred, while its enhanced version (the last image) contains most of the detailed components and is very similar to the original HR input (the first image).



**Figure 6.** The corresponding frame extracted from the Lepton 8-bit sequences used in the study for respiratory rate evaluation. From the left: original, Eulerian Video Magnification, bicubic interpolation scale 2, super resolved using our model scale 2, bicubic interpolation scale 4, super resolved using our model scale 4.



**Figure 7.** The corresponding frame extracted from the SC3000 8-bit sequences used in the study for respiratory rate evaluation. From the left: original, Eulerian Video Magnification, bicubic interpolation scale 2, super resolved using our model scale 2, bicubic interpolation scale 4, super resolved using our model scale 4.

#### 4. Discussion

The evaluation of the proposed CNN-based SR model on both datasets proved that thermal images can be enhanced and lead to improvement of image quality. Both on 8 and 16-bit data enhanced sequences led to higher values of PSNR. In addition, it turned out that on the SC3000 set, the SR model trained on images with even lower resolution (scale 4) outperformed results of the model trained on images down-scaled 2 times. This proves the possibility of improving low-resolution thermal images using deep learning techniques.

Although the achieved results are promising, they are only preliminary and we discovered some limitations of the proposed method that should be further investigated. In the best case RMSE of RR estimation was equal 2.46 bpm. In general, the accuracy results are not ideal as in most cases they oscillate around values of 3.5–5 bpm, which is not satisfactory for professional medical applications. Yet, the aim of the study was to evaluate the effect of enhancing very low-resolution images on RR estimation accuracy, which was confirmed as the proposed SR DNN led to a significant decrease of the RMSE. To further reduce the estimation error, we should extend the measurement time and specify more details about data acquisition conditions, e.g., visibility of the nostril area, what has been previously explained in similar studies [44,46]. Another limitation of the proposed method is the requirement of the manual ROI selection. We believe that the estimation accuracy could be further improved by providing an automatic way of the precise ROI selection. In addition, it could also be interesting to algorithmically detect facial areas where the RR signal is the best to keep human interference to a minimum, as shown in [51]. We would like to include both of these ideas in the future studies conducted in the area of contactless RR estimation.

In this study, we analysed short data segments to imitate potential practical applications of the proposed algorithms for home monitoring systems. To increase patients' convenience and avoid influence of body motion on the estimation accuracy, the data acquisition process should be as short as possible. Thus, we experimentally selected lengths of analysed sequences for both datasets. The number of selected samples have a direct effect on frequency resolution and thus on the RR estimation accuracy. For the Lepton dataset 300 samples ( $N$ ) were used for RR estimation and the sampling frequency ( $f_s$ ) was 12 Hz. Thus, the frequency resolution was

$$\delta f = \frac{f_s}{N} * 60 = \frac{12}{300} * 60 = 2.4 \text{ bpm} \quad (3)$$

For the SC3000 with  $f_s = 30$  and  $N = 400$ ,  $\delta f = 4.5$  bpm. To further improve RR estimation on the SC3000 dataset, more samples could be used. We will investigate this setting in the future study. Other factor influencing the results of RR estimation is selection of the window of samples analysed for RR estimation. This parameter was manually selected by us in the study, yet we believe that the performance can be further improved by automating this process.

It's important to note that the problem of RR estimation from thermal sequences in static and dynamic poses have been widely studied in the literature. However, the objective of our study was not to outperform existing RR estimation algorithms, but to evaluate the influence of enhancing thermal data with the SR model on the RR estimation results. As proved by previous studies, the best results of breathing signals analysis are achieved if nostrils are clearly visible in a frame [52] to be able to capture temperature modulation during breathing events. Also, various methods are often used to remove possible motion artifacts, e.g., video stabilization [53]. However, in our work we wanted to minimize computation overhead and limit the requirements about data acquisition setup to provide fast responses about the health status of the monitored person. Specifically, we assumed basic data acquisition setting (no need of tilting head backward to improve visibility of nostril area) and minimal data pre-processing step (no additional motion/lightning compensation algorithms). In this way, we wanted to make sure that the system could be used by non-professionals, without the supervision of the third person and calculate results on the resource-constrained devices.

The analysis performed for various thermal sequence resolutions, showed that the gain of the RR estimation accuracy can be achieved by applying SR deep neural network on low-resolution thermal images. For all estimators and aggregation operators the lowest RMSE values were achieved for the proposed DL model with the scaling factor of 2. For the SR model, scale 2, the Lepton dataset, average aggregator, RMSE of RR accuracy was reduced by 0.77 (8-bit data) comparing to bicubic and by 0.08 (8-bit data) comparing to original sequences if  $eRR_{sp}$  was used. In case of the  $eRR_{as}$  estimator, these differences are 4.19 and 10.66, respectively. For the SR model, scale 2, the SC3000 dataset, average aggregator, RMSE of RR accuracy was reduced by 3.41 (8-bit data) comparing to bicubic and by 0.54 (8-bit data) comparing to original sequences if  $eRR_{sp}$  was used. In case of the  $eRR_{as}$  estimator, these differences are 11.49 and 11.63, respectively.

For the scale 4, similar conclusion is true for the  $eRR_{sp}$  estimator. For the SR model, scale 4, the Lepton dataset, average aggregator, RMSE of RR accuracy was reduced by 0.65 (8-bit data) comparing to bicubic and by 0.01 (8-bit data) comparing to original sequences if  $eRR_{sp}$  was used. In case of the  $eRR_{as}$  estimator, these differences are 1.16 and 8.20, respectively. For the SR model, scale 4, the SC3000 dataset, average aggregator, RMSE of RR accuracy was reduced by 2.25 (8-bit data) comparing to bicubic and it was as good as original sequences if  $eRR_{sp}$  was used. In case of the  $eRR_{as}$  estimator, these differences are 3.69 and 3.83, respectively.

During the experiments, we found out that the  $eRR_{as}$  estimator is very sensitive to the selected ROI, resulting in unreliable RR values that led to a significant increase of RMSE. Thus, results for this estimator are inconclusive and should be further investigated in future studies.

In addition, for bicubic and EVM, 16-bit data turned out to be beneficial in improving accuracy of RR estimation. This indicates the need for providing lossless conversion of raw radiation values to temperature data. In case of SR, although higher PSNR and SSIM values were achieved for 16-bit sequences, this improvement hasn't led to better accuracy of RR estimation. This result may be caused by the difficulty of DL network training. Both 8-bit and 16-bit models were trained using the same hyperparameters, while the 16-bit network had to deal with more information. As a result, it may have overfitted to the training set or suffered from the vanishing gradient problem, what may have caused worse RR estimation accuracy. This issue will be further investigated by us in the future work.

The important finding of our study is that the proposed SR model outperformed the EVM algorithm on 8-bit and 16-bit data with scale 2 for all estimators and aggregation operators. For scale 4 the SR network was also better than EVM in most of the cases. It is worth noticing that for scale 4 images from the Lepton dataset were as small as  $20 \times 15$  pixels and still a similar RR accuracy was

achieved as for the original resolution ( $80 \times 60$ ) with color changes enhanced using EVM. Taking it into account, many applications that utilize very low resolution thermal images can potentially benefit from applying SR algorithms, leading to results similar to values estimated from data with much bigger spatial resolution. Examples of such applications include detecting of unusual behaviours at security checkpoint by analysis of vital signs patterns, monitoring of passengers' state of health in self-driving vehicles or contact-less estimation of respiratory rate in newborns. Considering the cost and availability of high-resolution thermal cameras, this finding can be crucial for enabling solutions that previously were very difficult to achieve, e.g., vital signs estimation at longer distances [54].

## 5. Materials and Methods

The proposed Convolutional-based Super Resolution Network and checkpoints trained on both datasets will be available at <https://github.com/akwasnie/Super-Resolved-Thermal-Imagery> for further exploration upon acceptance.

## 6. Conclusions

The aim of this study was to evaluate the influence of applying the Deep Learning-based Super-Resolution model on low resolution thermal video sequences for the purpose of improving remote measurement of a respiratory rate. While telemedicine was originally intended as a tool to treat patient in remote or not accessible locations, it is becoming increasingly popular as a tool to receive health care in a convenient way. Respiration rate is one of the useful metrics that can help in determining remote patient's health condition, and one of the means to obtain this signal is with the use of thermal cameras. Thermal cameras, while still expensive in comparison with standard web cameras, are becoming more affordable due to technological advances. Their modules are also becoming smaller, which allows for embedding them in more portable form factors, and some day we might even see them implemented in our personal computers just next to the web camera. The preliminary results indicate that applying such technique is beneficial and can lead to the decrease of the respiratory rate estimation error when compared with other enhancement methods, i.e., EVM by 41.2% and bicubic by 45.3% in the best case, as proven by tests on two independent databases. The performed experiments proved that DL can help to improve accuracy of vital signs estimation, however, we believe that further SR improvements could lead to achieving even higher accuracy and thus we will focus on this area in the future work. Another element of our work that could be improved is the selection of ROI and the window of samples used for RR estimation. Both parameters could be selected automatically, as described earlier. In our study we also identified other limitations of the proposed method, such as short acquisition time or specific conditions for data collection, i.e., visibility of the nostril area and provided ideas for addressing these issues in future studies.

**Author Contributions:** The authors specify following individual contribution to this study: conceptualization, A.K. and J.R.; methodology, A.K. and J.R.; software, A.K., J.R. and M.S.; validation, A.K.; formal analysis, A.K.; data acquisition, A.K. and J.R.; data curation, M.S.; writing—original draft preparation, A.K., J.R. and M.S.; writing—review and editing, A.K., J.R. and M.S.; visualization, A.K. and M.S.; supervision, J.R.; funding acquisition, J.R., A.K. and M.S.

**Funding:** This work has been partially supported by Statutory Funds of Electronics, Telecommunications and Informatics Faculty, Gdansk University of Technology and by Intel Corporation, USA. We thank all our colleagues, who provided insight and expertise that greatly assisted the research.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.



## Abbreviations

The following abbreviations are used in this manuscript:

CNN	Convolutional Neural Network
DL	Deep Learning
EVM	Eulerian Video Magnification
FPS	Frames Per Second
HR	High Resolution
LR	Low Resolution
NN	Neural Network
PSNR	Peak Signal to Noise Ratio
RR	Respiratory Rate
SISR	Single Image Super Resolution
SR	Super Resolution
SSIM	Structural Similarity Index

## References

1. Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. Sphreface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 212–220.
2. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823.
3. Liu, W.; Anguelov, D.; Erhan, D.; Szegedy, C.; Reed, S.; Fu, C.Y.; Berg, A.C. Ssd: Single shot multibox detector. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2016; pp. 21–37.
4. Girshick, R. Fast r-cnn. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1440–1448.
5. Qi, C.R.; Su, H.; Mo, K.; Guibas, L.J. Pointnet: Deep learning on point sets for 3d classification and segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 652–660.
6. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
7. Chaichulee, S.; Villarroel, M.; Jorge, J.; Arteta, C.; Green, G.; McCormick, K.; Zisserman, A.; Tarassenko, L. Multi-task convolutional neural network for patient detection and skin segmentation in continuous non-contact vital sign monitoring. In Proceedings of the 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), Washington, DC, USA, 30 May–3 June 2017; pp. 266–272.
8. Wang, Z.; Yang, Z.; Dong, T. A review of wearable technologies for elderly care that can accurately track indoor position, recognize physical activities and monitor vital signs in real time. *Sensors* **2017**, *17*, 341. [[CrossRef](#)] [[PubMed](#)]
9. Abbas, A.K.; Heimann, K.; Jergus, K.; Orlikowsky, T.; Leonhardt, S. Neonatal non-contact respiratory monitoring based on real-time infrared thermography. *Biomed. Eng. Online* **2011**, *10*, 93. [[CrossRef](#)] [[PubMed](#)]
10. Ng, E.Y.K. Is thermal scanner losing its bite in mass screening of fever due to SARS? *Med. Phys.* **2005**, *32*, 93–97. [[CrossRef](#)] [[PubMed](#)]
11. Kwasniewska, A.; Ruminski, J.; Szankin, M.; Czuszyński, K. Remote Estimation of Video-Based Vital Signs in Emotion Invocation Studies. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 17–21 July 2018; pp. 4872–4876.
12. Fei, J.; Zhu, Z.; Pavlidis, I. Imaging breathing rate in the co 2 absorption band. In Proceedings of the 2005 IEEE Engineering in Medicine and Biology 27th Annual Conference, Shanghai, China, 17–18 January 2006; pp. 700–705.
13. Murthy, R.; Pavlidis, I. Non-contact monitoring of respiratory function using infrared imaging. *IEEE Eng. Med. Biol. Mag.* **2006**, *25*, 57. [[CrossRef](#)]

14. Murthy, J.N.; van Jaarsveld, J.; Fei, J.; Pavlidis, I.; Harrykissoo, R.I.; Lucke, J.F.; Faiz, S.; Castriotta, R.J. Thermal infrared imaging: A novel method to monitor airflow during polysomnography. *Sleep* **2009**, *32*, 1521–1527. [\[CrossRef\]](#)
15. Fei, J.; Pavlidis, I. Thermistor at a distance: Unobtrusive measurement of breathing. *IEEE Trans. Biomed. Eng.* **2009**, *57*, 988–998.
16. Rumiński, J. Evaluation of the respiration rate and pattern using a portable thermal camera. In Proceedings of the 13th Quantitative Infrared Thermography Conference, Gdansk, Poland, 4–8 July 2016.
17. Rumiński, J. Reliability of pulse measurements in videoplethysmography. *Metrol. Meas. Syst.* **2016**, *23*, 359–371. [\[CrossRef\]](#)
18. Zhou, Y.; Tsiamyrtzis, P.; Lindner, P.; Timofeyev, I.; Pavlidis, I. Spatiotemporal smoothing as a basis for facial tissue tracking in thermal imaging. *IEEE Trans. Biomed. Eng.* **2012**, *60*, 1280–1289. [\[CrossRef\]](#)
19. Al-Khalidi, F.Q.; Saatchi, R.; Burke, D.; Elphick, H. Tracking human face features in thermal images for respiration monitoring. In Proceedings of the ACS/IEEE International Conference on Computer Systems and Applications-AICCSA 2010, Hammamet, Tunisia, 16–19 May 2010; pp. 1–6.
20. Chauvin, R.; Hamel, M.; Brière, S.; Ferland, F.; Grondin, F.; Létourneau, D.; Tousignant, M.; Michaud, F. Contact-free respiration rate monitoring using a pan-tilt thermal camera for stationary bike telerehabilitation sessions. *IEEE Syst. J.* **2014**, *10*, 1046–1055. [\[CrossRef\]](#)
21. Flir Lepton Camera Modules. Available online: <https://www.flir.com/products/lepton/> (accessed on 24 August 2019).
22. Wu, H.Y.; Rubinstein, M.; Shih, E.; Guttag, J.; Durand, F.; Freeman, W.T. Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Trans. Graph.* **2012**, *31*, 1–8. [\[CrossRef\]](#)
23. Wadhwa, N.; Rubinstein, M.; Durand, F.; Freeman, W.T. Phase-based video motion processing. *ACM Trans. Graph. (TOG)* **2013**, *32*, 1–9. [\[CrossRef\]](#)
24. Oh, T.H.; Jaroensri, R.; Kim, C.; Elgharib, M.; Durand, F.; Freeman, W.T.; Matusik, W. Learning-based video motion magnification. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 633–648.
25. Estrada, M.; Stowers, A. Amplification of Heart Rate in Multi-Subject Videos. Available online: [https://web.stanford.edu/class/ee368/Project\\_Spring\\_1415/Reports/Stowers\\_Estrada.pdf](https://web.stanford.edu/class/ee368/Project_Spring_1415/Reports/Stowers_Estrada.pdf) (accessed on 30 September 2019).
26. Balakrishnan, G.; Durand, F.; Guttag, J. Detecting Pulse from Head Motions in Video. In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013; pp. 3430–3437.
27. Aubakir, B.; Nurimbetov, B.; Tursynbek, I.; Varol, H.A. Vital sign monitoring utilizing Eulerian video magnification and thermography. In Proceedings of the 2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, FL, USA, 16–20 August 2016; pp. 3527–3530.
28. Bennett, S.L.; Goubran, R.; Knoefel, F. Comparison of motion-based analysis to thermal-based analysis of thermal video in the extraction of respiration patterns. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Seogwipo, Korea, 11–15 July 2017; pp. 3835–3839.
29. Dong, C.; Loy, C.C.; He, K.; Tang, X. Images super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Kim, J.; Kwon Lee, J.; Mu Lee, K. Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1637–1645.
31. Tai, Y.; Yang, J.; Liu, X. Image super-resolution via deep recursive residual network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; Volume 1, p. 5.
32. Zhang, X.; Li, C.; Meng, Q.; Liu, S.; Zhang, Y.; Wang, J. Infrared Image Super Resolution by Combining Compressive Sensing and Deep Learning. *Multidiscip. Digit. Publ. Inst. Sens. J.* **2018**, *8*, 2587. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Liu, C.; Sun, X.; Chen, C.; Rosin, P.L.; Yan, Y.; Jin, L.; Peng, X. Multi-Scale Residual Hierarchical Dense Networks for Single Image Super-Resolution. *IEEE Access* **2019**, *7*, 60572–60583. [\[CrossRef\]](#)

34. Wang, X.; Yu, K.; Dong, C.; Change Loy, C. Recovering realistic texture in image super-resolution by deep spatial feature transform. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 606–615.
35. Bulat, A.; Tzimiropoulos, G. Super-FAN: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with GANs. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 109–117.
36. Blau, Y.; Mechrez, R.; Timofte, R.; Michaeli, T.; Zelnik-Manor, L. The 2018 PIRM challenge on perceptual image super-resolution. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.
37. Cho, Y.; Bianchi-Berthouze, N.; Marquardt, N.; Julier, S.J. Deep Thermal Imaging: Proximate Material Type Recognition in the Wild through Deep Learning of Spatial Surface Temperature Patterns. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; p. 2.
38. Kniaz, V.; Gorbatshevich, V.; Mizginov, V. THERMALNET: A Deep Convolutional Network for Synthetic Thermal Image Generation. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2017**, *42*, 41. [[CrossRef](#)]
39. Bhattacharya, P.; Riechen, J.; Zölzer, U. Infrared Image Enhancement in Maritime Environment with Convolutional Neural Networks. *VISIGRAPP* **2018**, *4*, 37–46.
40. Kuang, X.; Sui, X.; Liu, Y.; Chen, Q.; Gu, G. Single infrared image enhancement using a deep convolutional neural network. *Neurocomputing* **2019**, *332*, 119–128. [[CrossRef](#)]
41. He, Z.; Tang, S.; Yang, J.; Cao, Y.; Yang, M.Y.; Cao, Y. Cascaded Deep Networks with Multiple Receptive Fields for Infrared Image Super-Resolution. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *29*, 2310–2322. [[CrossRef](#)]
42. Zhang, X.; Li, C.; Meng, Q.; Liu, S.; Zhang, Y.; Wang, J. Infrared Image Super Resolution by Combining Compressive Sensing and Deep Learning. *Sensors* **2018**, *18*, 2587. [[CrossRef](#)]
43. Kwasniewska, A.; Szankin, M.; Ruminski, J.; Kaczmarek, M. Evaluating Accuracy of Respiratory Rate Estimation from SuperResolved Thermal Imagery. In Proceedings of the IEEE EMBC Conference, Berlin, Germany, 23–27 July 2019; pp. 1–4.
44. Ruminski, J.; Kwasniewska, A. Evaluation of respiration rate using thermal imaging in mobile conditions. In *Application of Infrared to Biomedical Sciences*; Springer: Singapore, 2017; pp. 311–346.
45. Kaczmarek, M.; Bujnowski, A.; Wtorek, J.; Polinski, A. Multimodal platform for continuous monitoring of the elderly and disabled. *J. Med. Imaging Health Inform.* **2012**, *2*, 56–63. [[CrossRef](#)]
46. Ruminski, J.; Bujnowski, A.; Czuszyński, K.; Kocejko, T. Estimation of respiration rate using an accelerometer and thermal camera in eGlasses. In Proceedings of the 2016 Federated Conference on Computer Science and Information Systems (FedCSIS), Gdańsk, Poland, 11–14 September 2016; pp. 1431–1434.
47. Kwaśniewska, A.; Giczewska, A.; Rumiński, J. Big data significance in remote medical diagnostics based on deep learning techniques. *Task Q.* **2017**, *21*, 309–319.
48. Kim, J.; Kwon Lee, J.; Mu Lee, K. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
49. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
50. Eulerian Video Magnification Online Tool. Available online: <https://lambda.qrilab.com/site/> (accessed on 10 August 2019).
51. Ruminski, J.; Kwasniewska, A.; Szankin, M.; Kocejko, T.; Mazur-Milecka, M. Evaluation of Facial Pulse Signals Using Deep Neural Net Models. In Proceedings of the IEEE EMBC Conference, Berlin, Germany, 23–27 July 2019; pp. 1–4.
52. Hochhausen, N.; Barbosa Pereira, C.; Leonhardt, S.; Rossaint, R.; Czaplík, M. Estimating Respiratory Rate in Post-Anesthesia Care Unit Patients Using Infrared Thermography: An Observational Study. *Sensors* **2018**, *18*, 1618. [[CrossRef](#)] [[PubMed](#)]

53. Alam, S.; Singh, S.P.; Abeyratne, U. Considerations of handheld respiratory rate estimation via a stabilized Video Magnification approach. In Proceedings of the 2017 39th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Jeju Island, Korea, 11–15 July 2017; pp. 4293–4296.
54. Szankin, M.; Kwasniewska, A.; Sirlapu, T.; Wang, M.; Ruminski, J.; Nicolas, R.; Bartscherer, M. Long Distance Vital Signs Monitoring with Person Identification for Smart Home Solutions. In Proceedings of the 2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Honolulu, HI, USA, 17–21 July 2018; pp. 1558–1561.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).