Contactless Respiratory Rate Monitoring For ICU Patients Based On Unsupervised Learning

Zimeng Liu[1,2], Bin Huang[1*], Chun-Liang Lin[3*], Chieh-Liang Wu[4,5], Changchen Zhao[1], Wen-Cheng Chao[4,5,6], Yu-Cheng Wu[4], Yadan Zheng[7] and Zhiru Wang[8]
[1] Hangzhou Innovataion Institute, Beihang University.
[2] School of Automation Science and Electrical Engineering, Beihang University.
[3] College of Electrical Engineering and Computer Science, National Chung Hsing University.
[4] Department of Critical Care Medicine, Taichung Veterans General Hospital.
[5] Department of post-Baccalaureate Medicine, College of Medicine, National Chung Hsing University.
[6] Big Data Center, National Chung Hsing University.
[7] Computer Science School, Beijing University of Posts and Telecommunications.
[8] Graduate School of Translation and Interpretation, Beijing Foreign Studies University.
*Corresponding Authors: marshuangbin@buaa.edu.cn and chunlin@dragon.nchu.edu.tw.

Abstract

Recently, the task of contactless physiological signal monitoring based on deep learning technologies has attracted a large number of scholars. However, few studies focus on the application of real-world scenarios, especially in clinical medicine scenes. In this paper, a novel video-based contactless respiratory rate measurement algorithm is developed for the Intensive Care Unit (ICU) patients. Firstly, a large-scale clinical real-world database towards ICU patient is collected in this study. Then, based on the dataset, the unsupervised learning is first introduced to extract the respiration waveform from the chest area of patients. Lastly, a respiratory rate estimator based on neural networks is proposed and trained on a periodical physiological signal simulation dataset, and utilizes the transfer learning technique to extract the respiratory rate from only 10-second respiration waveform. We obtained estimated respiratory rate with an MAE of 2.8 breaths/min and an STD of 3.0 breaths/min against the reference value computed from the specialized medical device. Extensive experiments demonstrate that our proposed methods achieve competitive results over state-of-the-art (SOTA) method in the real-world scenario.

1. Introduction

Whether in the hospital or in daily life, the accurate measurement of physiological parameters is of vital importance, which can help monitor the health conditions of humans such as in the COVID-19 pandemic and its post pandemic. Among all physiological parameters, the respiratory rate (RR) is one of the most critical parameters, which is not only essential to the respiratory system but also a key indicator of a severe disorder in many body systems. In clinical scenarios, an abnormal value of RR can indicate some respiratory problems directly related to the lungs and unanticipated intensive care unit admission [6]. For example, patients with their lungs infected by the COVID-19 virus will breathe significantly faster, which means they are more likely to need hospital treatment [8, 9].

Broadly speaking, the theory of measuring respiratory rate is based on the movement of the chest area during contraction and relaxation. Physicians sometimes make a rough estimation of the respiratory rate of patients referring to the movement of the chest area.

To be more precise, the basic principle of RR measurement by the commonly used clinical vital sign monitors is the theory of thoracic impedance. When two electrodes are attached to the chest area of subjects, the contraction and relaxation of the abdominal and intercostal muscles cause changes in the body's electrical resistance, thus monitoring the RR waveform. However, this contact-based method might have a variety of disadvantages. Firstly, it is observed that the measurement procedure of RR in clinical settings is complicated, requiring specialized personnel to handle it. Secondly, considering public health factors, the sensor probes must be strictly sterilized before each use. Moreover, the contact of the sensor might cause irritation to some patients whose skin is fragile [10] and increase the chance

of cross-infection during the COVID-19 pandemic.

The contactless technique for measuring RR can provide an appropriate solution to the thorny problems mentioned above. The non-contact RR measurement system just needs a RGB camera and can realize long-term as well as user-friendly monitoring. The video-based method is also based on the movement of the chest area of subjects, but it is unnecessary to access the skin of chest, which can protect privacy. In most cases, RR is capable to be detected when the subject are wearing heavy clothes or even covered with a quilt. Jorge et al. [15], Chen et al. [5] and Massaroni et al. [21] [18] [20] estimated RR by analyzing the intensity of reflected light, while Alinovi et al. [1] , Antognoli et al. [2] and Brieva et al. [3] used the motion-based method to measure RR. They mostly focused on testing their method in the laboratory scenario, which is rather different from a real medical situation.

Janssen et al. [14] and Rossol et al. [22] tested their algorithms in the Neonatal Intensive Care Unit (NICU) scenario. However, the number of subjects recruited in the two papers is too small as only two newborns in NICU were tested. Recently, the RR monitoring algorithm based on deep learning have developed rapidly. Preterm infants in NICU [25] and post-operative patients in ICU [16] have been tested respectively. Both of the experiments require 30s video frames to compute RR. When it comes to the size of datasets, only 15 patients were involved in the validation studies in [16].

In recent years, lots of studies have demonstrated that deep learning technologies can be devoted to the video-based physiological signal measurement [11–13, 26]. However, the majority of those researches focus on healthy subjects in the laboratory environment, and there are still many great challenges to accelerate the application of video-based vital sign monitoring technology in some real-world scenarios, especially in medical scenes. Besides, it is important to efficiently train supervised learning algorithms with synchronized information between the video sequence and physiological signals, both of which are hard to be synchronous, in that the golden standard of physiological information is even missing due to the movement of subject or other exceptional situation in the monitoring. For these reasons, a real-world clinical ICU patient database, devoted to the contactless monitoring of vital signs based on videos, is collected in Taichung Veterans General Hospital, Taiwan, and the unsupervised method is introduced to this clinical study of contactless RR monitoring.

In this paper, we have done extensive clinical experiments on ICU patient dataset using the contactless method. Since there has been no public ICU dataset which can be used to measure RR remotely, we collected data in the ICU wards of a hospital and built a database ourselves. Furthermore, the deep learning network was introduced to improve the robustness of our algorithm. Our method can be divided into two components: respiratory signal extractor and RR estimator, which are trained separately. We adopted the unsupervised learning method to train the first part of our network, and tested the deep learning model on our test dataset. Compared with the ground truth value of RR, our method can achieve the high accuracy and can be applied to the real clinical scenario.

In summary, the contributions of this paper are mainly listed as follows:

- In this study, a real-world large-scale dataset is collected for the task of contactless vital signs monitoring for ICU patients, in which 405 patients are involved and the physiological parameters such as heart rate (HR), respiration rate (RR), electrocardiogram (ECG), blood pressure (BP) and blood oxygen saturation (SpO2) are included.
- To the best of our knowledge, it is the first time that unsupervised learning method has been introduced for the task of video-based contactless respiration waveform extraction in ICU. Moreover, the respiratory rate estimator, a lightweight time series neural network, which only requires 10-second respiration waveform, is proposed to estimate RR value.
- The extensive experiments demonstrate that the unsupervised learning and RR estimator models show a high accuracy and robustness on the ICU patient dataset. It means a stride in the application of contactless methods in real-world clinical scenarios.

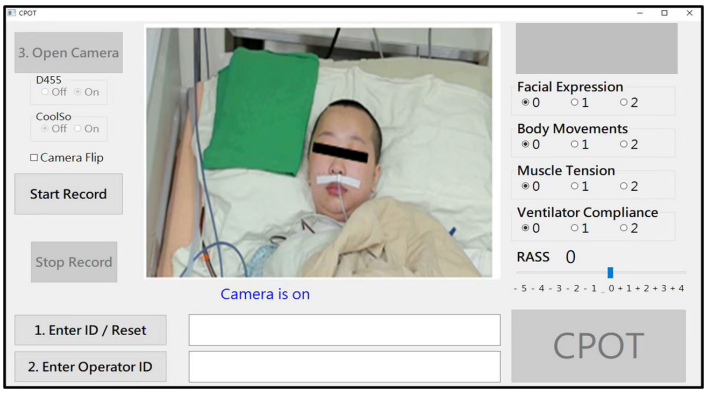
2. Subjects and Data

Most of the existing public datasets were collected in laboratory scenarios, excluding those where the chest wall area of patients is hidden. To do clinical experiments in the real world, we collected a large quantity of data including videos and the ground truth of physiological parameters in the Taichung Veterans General Hospital of Taiwan and built our own dataset. The overview of the dataset is presented in Table 1.

2.1. Subjects

The subjects of our experiments were all from ICU wards in the hospital of Taiwan. We have collected about 405 data in total, from July, 2022 to February, 2023. And the continuing process of data acquisition is scheduled to end until June, 2023. There were 152 female patients, accounting for 37.53%, and 253 male patients, accounting for 62.47%, the average age of whom was 67.82. We have obtained the permission of the ICU patients in our dataset. To protect their privacy, we will do some mosaic processing when presenting the data.

(a) The scene of an ICU ward where we collected data.



(b) The interface of the video recording.

Figure 1. The overview of data collection. In (a), an ICU patient was lying in the sickbed. The monitoring device BROADSIMS TeleTouch and the computer connected to a camera were collecting data. In (b), the captured video footage is shown. The level of facial expression, body movements, RASS and so on can be adjusted.

Table 1. The overview of our clinical ICU patient dataset.

Descriptions	Values
Total number of patients	405
Total length of videos (hour)	67.5
Gender(Females/Males)	152/253(37.53%/62.47%)
Average age (year)	67.82
Number of video clips of each patient	1-4
Number of text files	≈ 300
ECG sampling rate (Hz)	500
RR sampling rate (Hz)	500
Number of training data	103
Number of test data	44
RASS	≤ 0

2.2. Data Collection

The scene of an ICU ward where we collected data is shown in Figure 1a. In the process of data collection, a camera was held over the sickbed and maintained an overlooking angle. Since there was no special requirements, patients could just lie in bed and breathe spontaneously. The interface of the video recording is shown in Figure 1b. In this video, the patient was sleeping peacefully, so the parameter Richmond Agitation-Sedation Scale (RASS) [7] was set to 0. The chest area of patients should keep intact despite the connected monitoring instruments. The resolution of all videos is 1440×1080 or 1280×720 pixels. The frame rate of videos is set to 30 fps. For each patient, 1-4 video clips were collected and the total duration was around 10 minutes.

Simultaneously, we have collected the ground truth of physiological parameters using the professional monitoring equipment BROADSIMS TeleTouch. The data recording device can measure ECG, RR, non-invasive blood pressure (NIBP), SpO2 and pulse rate (PR) in real time, and output the data to the computer. Therefore, our ICU patient dataset can be used not only to measure RR, but also to complete the estimation task of HR, SpO2 and other physiological parameters. Specific to the RR estimation in this paper, the sampling rate of RR is set to 500 Hz and the measurement accuracy can achieve $\pm 2\%$.

3. Method

Our proposed method consists of respiratory signal extractor and RR estimator, which forms a multitask pipeline to estimate respiratory signals and RR simultaneously. The two components are trained in different ways, the first mentioned of which is trained on our ICU patient dataset in an unsupervised way while the second is trained on a simulation dataset based on supervised learning.

3.1. Data Preprocessing

For respiratory signal extractor, the videos should be processed to regions of interest (ROIs) of the same size and time length. For RR estimator, a large simulation dataset including three different signals should be generated for training.

3.1.1 ROI Selecting

Except for the videos of which the recorded thoracic area is incomplete, the level of RASS is greater than zero or the light in the ward is unstable, the proper data for RR estimation are selected from the ICU patient dataset. The number of training and test data is 103 and 44, respectively. For every video, ROI selecting is an essential procedure. Since the periodic change in the thoracic area is most obvious during respiration, we select the chest wall area of the patients as the ROI. Due to the occlusion of the medical devices, it is troublesome to apply the algorithm that automatically iden-

Figure 2. A selected ROI sample of an ICU patient. The blue box denotes the ROI we label manually, and the red spots denote the feature points of KLT tracking algorithm.

tifies the thorax region. Since the problem is not what we really focus on, we label the bounding box of ROI manually. A selected ROI sample of the patient in ICU is shown in Figure 2. The ROI is labeled only in the first frame of each video, and the KLT tracking algorithm [23] is used to track and extract the ROIs in the subsequent frames.

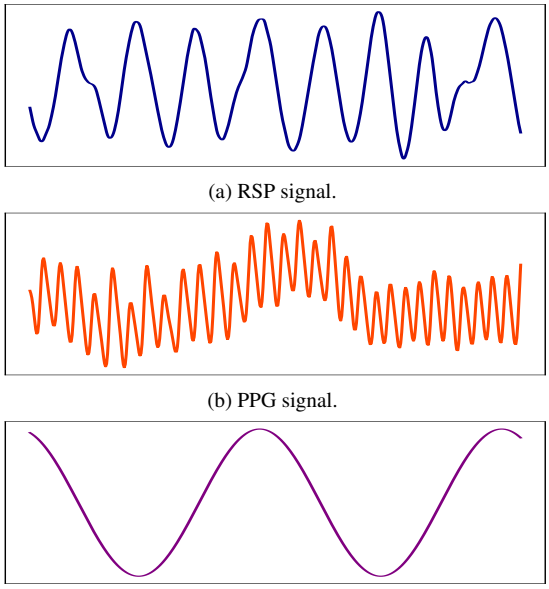
For each patient video, 1800 consecutive frames are selected, adding to one minute. The ROIs are resized to 128×128 uniformly in order to be input into the neural network. Furthermore, the real values of RR are read from the text files. The two kinds of information are saved to one h5 file, which facilitates the repeated reading of data.

3.1.2 Simulation Dataset for RR Estimator

To train the RR estimator more efficiently, we use built-in functions in *neurokit2* [17] to generate three kinds of periodic signals of morphology. To ensure the diversity of the training data, we generate not only Respiratory (RSP) signals, but also Photoplethysmography (PPG) signals as well as the standard sinusoidal signals. The three kinds of simulation signals are shown in Figure 3.

For RSP and PPG, the sampling rate is set to 120 Hz. Then, the signals are downsampled, reducing the sampling rate to 30 Hz. The duration of every sample is 60s, which means 1800 frames in total. The data are sliced in order to diversify the phase. For each sample, the slicing length is fixed to 500 frames. The beginning location of the slicing operation is random. In addition, the motion amplitude for PPG is set as a nonzero value to simulate the inhomogeneous amplitude in the real breathing process. For the standard sinusoidal signals, different frequencies and phases are set. The detailed parameters of the three kinds of simulation signals are illustrated in Table 2.

Each kind of signal is generated in equal amounts and concatenated to a larger matrix. Each sample corresponds



(a) RSP signal.

(b) PPG signal.

(c) Sinusoidal signal.

Figure 3. A random sample of the three kinds of morphological signals. 20k samples of the signals are generated respectively for training RR estimator and 4k are generated respectively for validation.

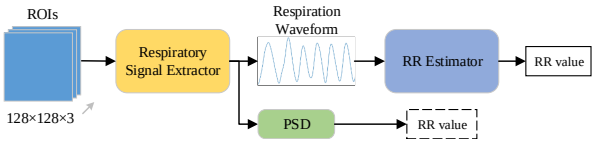


Figure 4. The overall architecture of our deep neural network. It consists of respiratory signal extractor and RR estimator.

to a label, used for the supervised learning of the RR estimator. For training, 20k samples of RSP, PPG and sinusoidal signals are generated respectively. For validation, 4k samples of each kind of signal are generated respectively.

3.2. Architecture of Deep Neural Network

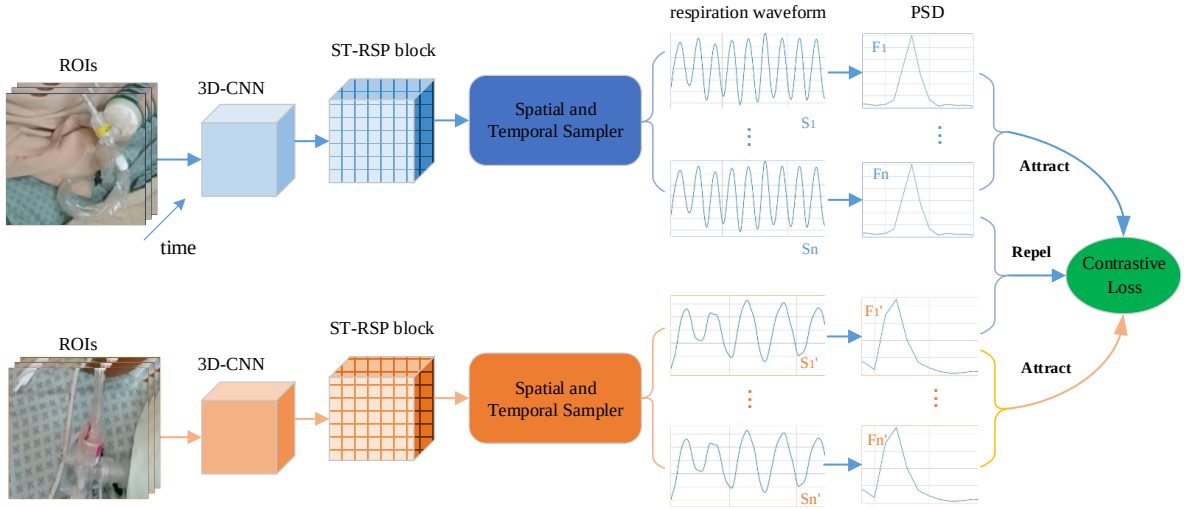
Based on the ICU patient dataset, we propose an efficient deep neural network. It consists of two components: respiratory signal extractor and RR estimator. The overall architecture of the network is depicted as Figure 4. It can be regarded as a versatile pipeline, which can estimate respiratory waveforms and RR values.

3.2.1 Respiratory Signal Extractor

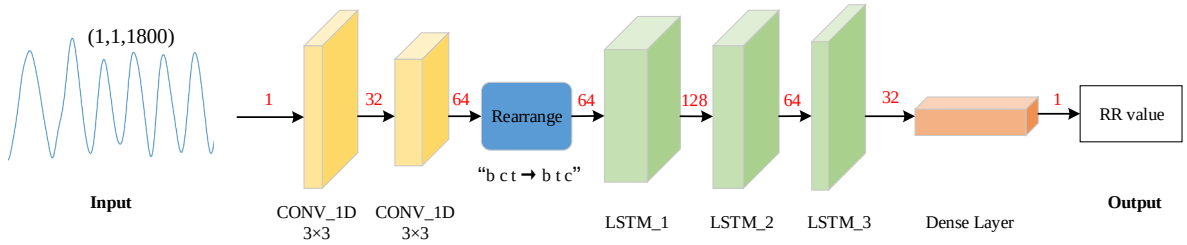
The architecture of the respiratory signal extractor is inspired from Contrast-Phys [24]. Due to the similarity of the principle of RR and HR estimation, minor changes are made to the original model. As shown in Figure 5a, chest ROI frames of a pair of videos are fed into the same 3D-

Table 2. The detailed parameters of the three kinds of simulation signals.

Signal	Duration (s)	Sampling rate (Hz)	Range (bpm)	Noise	Slicing length	Motion amplitude
RSP	60	120	10-30	0.01	500	-
PPG	60	120	40-150	0.01	500	0.1-1
SIN	60	120	10-150	0	500	-



(a) The architecture of the respiratory signal extractor.



(b) The architecture of the RR estimator.

Figure 5. The two components of our method. In (a), a pair of ROIs are input to the 3D-CNN. Different respiration waveforms and PSDs are generated to calculate loss for unsupervised learning. In (b), the input of RR estimator is respiration waveform, and the output is RR value.

CNN block. Next, ST-RSP blocks which follows [24] are generated, which is a collection of respiratory signals in spatio-temporal dimensions. Then, respiratory signals are sampled both in spatial and temporal positions to generate multiple samples. Denote the width and height of the ROI frames as w and h. For spatial sampling, we express the respiratory signal clip as $S(\cdot, h, w)$. For temporal sampling, we express the signal clip as $S(t \rightarrow t + \Delta t, h, w)$, where t and Δt denote the start time and the duration of the sample, respectively.

The respiratory signals are converted to Power Spectral Density (PSD) subsequently in order to compute loss in the frequency domain. The PSDs from the same ROI video are attracted, while those from different videos are repelled. To gain the dynamic performance of the respiration and higher accuracy, the extracted respiratory signals are input to the following RR estimator.

3.2.2 RR Estimator

The overview of the RR estimator of our network is presented in Figure 5b. It is composed of two 1D convolutional layers, three Long Short Term Memory (LSTM) layers and one linear layer. The respiration waveform is input into the estimator, and the model predicts the numerical value of RR as the output. The respiration array should be rearranged before it is input into LSTM layer, or mistakes will take place. The convolution operation is employed not only to extract the features of the input sequence, but also to reduce

the number of parameters. The LSTM layer is used to handle the problems of long distance dependence and gradient vanishing or explosion. The linear layer at the end of the neural network is used to map the feature channels to the proper one.

The respiration waveform is an 1D array in essence, which can be expressed as $S_{RSP} = array(T,)$, where T denotes the frames of ROIs. In our experiments, the length of all respiratory signal sequences is 1800. The estimator is trained on the simulation dataset and obtains the weight parameters. It predicts the value of RR with sliding windows [12], where the window length and stride can be adjusted. In this way, each waveform is sliced into several pieces and the dynamic variation trend of respiration can be displayed. The estimation process can be illustrated as

$$\hat{V}_i = \mathcal{F}_{estimator}\left(S_i, \omega, b\right) \tag{1}$$

where $\hat{V}_i$ and S_i denote the predicted value of RR and the respiratory signal piece of every window, respectively. ω and b denote the parameters of the 1D convolution layers or LSTM layers, which can be learned through the back propagation process.

All values are averaged as the final prediction result of RR:

$$\hat{V}_{RR} = \frac{1}{n}\sum_{i=1}^{n}\hat{V}_i \tag{2}$$

where $\hat{V}_{RR}$ and n denote the final predicted RR value and the total number of predicted RR value with sliding windows, respectively.

3.3. Training Strategy

To improve the efficiency of training process, two parts of our network were trained separately. The respiratory signal extractor was trained on our ICU patient dataset based on unsupervised learning. The RR estimator was trained on the *neurokit2* simulation dataset based on supervised learning.

3.3.1 Loss Function

For the respiratory signal extractor, the loss function consists of a positive term and a negative term. Suppose f_i and f_j are PSDs from the same chest ROI video, and f_i' and f_j' are PSDs from another ROI video. Referring to the spatio-temporal similarity of physiological signals in the same video and the dissimilarity across videos [24], we can get $f_i \approx f_j$, $f_i' \approx f_j'$, $i \neq j$. Then, the positive term can be defined as

$$L_p = \sum_{i=1}^{N}\sum_{\substack{j=1 \\ j\neq i}}^{N}\left(\left\|f_i - f_j\right\|^2 + \left\|f_i' - f_j'\right\|^2\right)/(2N(N-1)) \tag{3}$$

where N denotes the total number of signal clips of one video.

The negative term can be defined as

$$L_n = -\sum_{i=1}^{N}\sum_{j=1}^{N}\left\|f_i - f_j'\right\|^2/N^2 \tag{4}$$

Thus, the loss function can be expressed as $Loss = L_p + L_n$.

For the RR estimator, Mean Squared Error (MSE) loss is used to evaluate the error between the predicted RR and the ground truth since the RR estimation can be considered as a regression task. MSE loss can be expressed as

$$L_{MSE} = \frac{1}{n}\sum_{i=1}^{n}\left(\hat{f}_{RR}(i) - f_{RR}(i)\right)^2 \tag{5}$$

where $\hat{f}_{RR}(i)$ and $f_{RR}(i)$ denote the predicted and real value of RR, respectively.

3.3.2 Training Details

The training dataset of the respiratory signal extractor includes 103 video clips from our ICU patient dataset. And 60k simulation signals are generated to train the RR estimator. Both of the models are trained on one TITAN RTX (24G RAM) GPU.

For the respiratory signal extractor, the iteration epoch is set as 400. The learning rate is set as 1×10^{-5}, and the batch size is set as 2. Besides, the temporal and spatial dimensions of ST-RSP block are set as 600 and 2, respectively. The time length of each respiratory signal clip is set as 300. In the spatial dimension, the number of signal clips is 4. The high-pass and low-pass cut-off frequency of filters are set to 0.13 Hz and 0.5 Hz according to the normal range of RR. The AdamW optimizer is adopted to train the model.

For the RR estimator, the batch size is set as 50. The learning rate is set as 1×10^{-4}, and the iteration epoch is set as 390. The Adam optimization algorithm is used to optimize the model.

4. Experiments and Results

4.1. Evaluation Metrics

In our experiments, we use Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE) and Standard Deviation (STD) to evaluate the accuracy of our non-contact method of measuring RR. The effectiveness of the method depends on the values of the metrics. The evaluation metrics can be defined as the following equations respectively:

$$MAE = \frac{1}{N}\sum_{i=1}^{N}\parallel \hat{V}_{RR}(i) - V_{RR}(i) \parallel \tag{6}$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\hat{V}_{RR}(i) - V_{RR}(i))^2} \quad (7)$$

$$MAPE = \frac{1}{N}\sum_{i=1}^{N} \parallel (\hat{V}_{RR}(i) - V_{RR}(i))/V_{RR}(i) \parallel \quad (8)$$

$$STD = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(error(i) - error\_mean)^2} \quad (9)$$

where $\hat{V}_{RR}$ and V_{RR} denote the predicted value and ground truth of RR, respectively. N denotes the total number of samples in the test set. The error can be expressed as $error(i) = \hat{V}_{RR}(i) - V_{RR}(i)$ in Equation 9.

4.2. Clinical Experiments

To the best of our knowledge, previous studies have not explored more in the real clinical environment on a large scale [16]. Therefore, we do extensive clinical experiments on our ICU patient dataset, applying our method to the real-world clinical scenarios.

By the respiratory signal extractor, we can simultaneously get the respiration waveform and PSD map. Several samples of respiratory signals and the corresponding PSD maps are shown in Figure 6. It is intuitive to observe the respiratory pattern from the waveform, where the peaks indicate the expiratory process while the troughs indicate the inspiratory process. In the frequency domain, the peak of the PSD map shows the value of RR computed by fast Fourier transform (FFT) [4].

However, in our experiments, at least 800-frame of respiration waveform must be fed into the estimation model in the PSD-based RR estimation method, otherwise the accuracy of this method will drop dramatically. This is because the sampling rate is 30 fps and the minimum frequency of respiration is 0.15 Hz, which means that the 800-frame respiration waveform only contains $800/30 * 0.15 = 4.0$ cycles. Therefore, if the input respiration waveform less than four periods (800 frames), the PSD-based method can not compute the RR efficiently.

To improve the real-time performance and accuracy of non-contact RR estimation, we predict the final values by RR estimator based on deep learning method, which only needs 5-second (150 frames) respiration waveform. We test our method on 44 ICU patient samples from our clinical test set and compare the predicted RR value with the ground truth. The dynamic contrast between the predicted RR and real RR is presented in Figure 7, indicating that the estimated signal of our method is similar to the ground truth of respiratory signal. We can also see that the first 5s waveform suggests the relatively slow breathing while the next 10s reveals that the patient accelerates breathing, and slows down breathing afterwards.

Our experimental results are shown in Table 3. We have also done the sensitive analysis of the window length and stride of RR estimator. Due to the limitation of FFT, PSD [24] can calculate RR values with at least 800-frame signal. It can prove that our method gets the MAE of 2.651 bpm, RMSE of 3.525 bpm, MAPE of 13.8% and STD of 3.383 bpm when the window length and stride (*win* and *s* in Table 3 respectively) are set to 500 and 60 respectively, achieving a competitive performance in real clinical scenarios. Moreover, our method can almost achieve the same high accuracy as that of PSD, even though only 5s waveform is needed for our method. Referring to the results in Figure 7, our proposed method achieves a prominent dynamic performance in tracking the fluctuations of RR. In other words, it can monitor abnormal RR fluctuations, which is vitally important for the real-world applications in the clinical medicine scenarios.

4.3. Comparative Experiments

In contrast to the performance of traditional methods [19] of extracting respiratory signals, we have also done comparative experiments. Referring to the motion-based method, we extract the signals by analyzing the intensity change of the ROI pixels and select the 5% of the curves with higher standard deviation [19]. Then, three curves with the closest trend are selected as the final figure of the respiration pattern. The same RR estimator is applied to predict the numerical value of RR. Table 4 presents the comparative results of two algorithms. It shows that the our method based on unsupervised learning outperforms the traditional motion-based method on respiratory signal extraction and RR estimation.

5. Conclusions

To expand the contactless vital signs monitoring technologies to clinical medicine scenarios, a real-world large-scale dataset is collected for the task of video-based contactless monitoring for ICU patient. In addition, the unsupervised learning is first introduced to extract the respiration waveform from the chest area of patients, and a RR estimator is proposed to measure the respiratory rate from only 10-second respiration waveform. Based on these experimental results in clinical scenarios, the main conclusion of this study can be summarized as follows:

In the clinical validation studies, the experimental results demonstrate that the proposed contactless RR estimation method can almost achieve the same high accuracy as that of contact-based medical devices. More significantly, the results show the robustness and effectiveness of our pro-

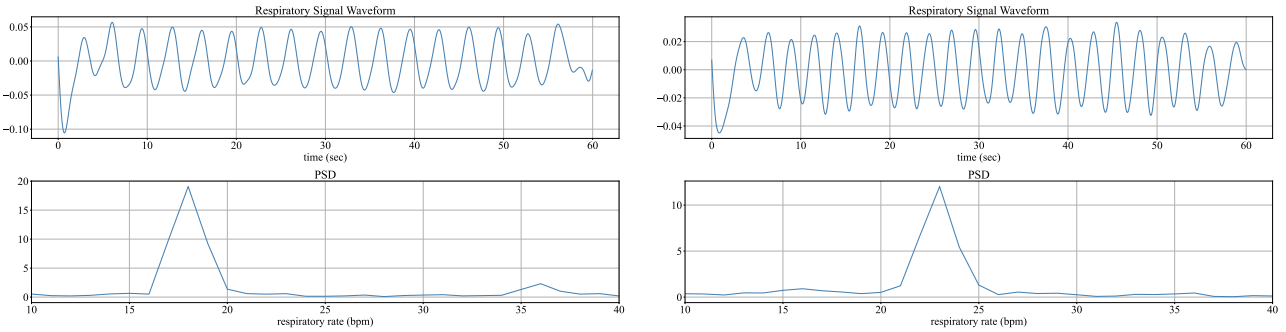


Figure 6. Testing samples of respiratory signals and the corresponding PSDs. It is intuitive to observe the respiratory pattern. And the PSD peak denotes the RR value computed by FFT.

Table 3. The results of our clinical experiments.

Method	Parameters		MAE (bpm)	RMSE (bpm)	MAPE (%)	STD (bpm)
PSD [24]	win=800	s=30	3.664	4.632	18.3	3.531
	win=900	s=30	3.466	4.367	17.4	3.270
	win=1800	s=30	3.990	5.215	20.5	4.965
Ours	win=150	s=30	3.955	4.656	20.2	3.072
		s=60	4.048	4.745	20.6	3.017
	win=300	s=30	2.807	3.626	14.2	3.025
		s=60	2.788	3.628	14.1	**3.014**
	win=500	s=30	2.663	3.543	13.8	3.414
		s=60	**2.651**	**3.525**	**13.8**	3.383

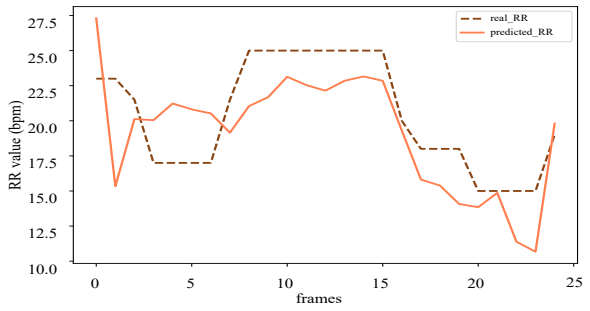


Figure 7. The dynamic contrast between predicted RR and real RR. The estimated trend is very close to the actual trend.

posed method. Even when part of the chest area of patients are hidden by medical equipment, our method can also extract the respiration waveform and estimate RR accurately.

We expand the scope of our study by migrating the experimental environment from laboratories to the real clinical scenarios of hospitals. Moreover, we provide insights that unsupervised learning may benefit and facilitate the development of contactless vital signs monitoring algorithms in the clinical scenario. Particularly, when the golden standard of physiological information is incomplete or missing, the unsupervised learning can also make these data sense, which offers great potential to employ the large-scale clinical data and increases the practical significance of the contactless RR estimation algorithm.

Table 4. The comparative results of our method and a traditional motion-based method (window length=300, stride=30).

Method	MAE (bpm)	RMSE (bpm)	MAPE (%)	STD (bpm)
Ours	2.807	3.626	14.2	3.025
Massaroni et al. [19]	3.364	4.300	18.3	3.847

Acknowledgment

This work is supported by the National Natural Science Foundation of China under grant No. 62203037. It is also jointly supported by National Chung Hsing University and Taichung Veterans General Hospital under grant No. TCVGH-NCHU 1110104.

References

[1] Davide Alinovi, Gianluigi Ferrari, Francesco Pisani, and Riccardo Raheli. Respiratory rate monitoring by video processing using local motion magnification. In *2018 26th European Signal Processing Conference (EUSIPCO)*, pages 1780–1784. IEEE, 2018. 2

[2] Luca Antognoli, Paolo Marchionni, Susanna Spinsante, Stefano Nobile, Virgilio Paolo Carnielli, and Lorenzo Scalise. Enanced video heart rate and respiratory rate evaluation: standard multiparameter monitor vs clinical confrontation in newborn patients. In *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–5. IEEE, 2019. 2

[3] Jorge Brieva, Ernesto Moya-Albor, Orlando Rivas-Scott, and Hiram Ponce. Non-contact breathing rate monitoring system based on a hermite video magnification technique. In *14th International Symposium on Medical Information Processing and Analysis*, volume 10975, pages 19–25. SPIE, 2018. 2

[4] E Oran Brigham. *The fast Fourier transform and its applications*. Prentice-Hall, Inc., 1988. 7

[5] Mingliang Chen, Qiang Zhu, Harrison Zhang, Min Wu, and Quanzeng Wang. Respiratory rate estimation from face videos. In *2019 IEEE EMBS International Conference on Biomedical & Health Informatics (BHI)*, pages 1–4. IEEE, 2019. 2

[6] Michelle A Cretikos, Rinaldo Bellomo, Ken Hillman, Jack Chen, Simon Finfer, and Arthas Flabouris. Respiratory rate: the neglected vital sign. *Medical Journal of Australia*, 188(11):657–659, 2008. 1

[7] E Wesley Ely, Brenda Truman, Ayumi Shintani, Jason WW Thomason, Arthur P Wheeler, Sharon Gordon, Joseph Francis, Theodore Speroff, Shiva Gautam, Richard Margolin, et al. Monitoring sedation status over time in icu patients: reliability and validity of the richmond agitation-sedation scale (rass). *Jama*, 289(22):2983–2991, 2003. 3

[8] Trisha Greenhalgh, Matthew Knight, Matt Inada-Kim, Naomi J Fulop, Jonathan Leach, and Cecilia Vindrola-Padros. Remote management of covid-19 using home pulse oximetry and virtual ward support. *bmj*, 372, 2021. 1

[9] Kara Hanson, Nouria Brikci, Darius Erlangga, Abebe Alebachew, Manuela De Allegri, Dina Balabanova, Mark Blecher, Cheryl Cashin, Alexo Esperato, David Hipgrave, et al. The lancet global health commission on financing primary health care: putting people at the centre. *The Lancet Global Health*, 10(5):e715–e772, 2022. 1

[10] Mirae Harford, Jacqueline Catherall, Stephen Gerry, John Duncan Young, and P Watkinson. Availability and performance of image-based, non-contact methods of monitoring heart rate, blood pressure, respiratory rate, and oxygen saturation: a systematic review. *Physiological measurement*, 40(6):06TR01, 2019. 1

[11] Bin Huang, Weihai Chen, Chun-Liang Lin, Chia-Feng Juang, and Jianhua Wang. Mlp-bp: A novel framework for cuffless blood pressure measurement with ppg and ecg signals based on mlp-mixer neural networks. *Biomedical Signal Processing and Control*, 73:103404, 2022. 2

[12] Bin Huang, Weihai Chen, Chun-Liang Lin, Chia-Feng Juang, Yuanping Xing, Yanting Wang, and Jianhua Wang. A neonatal dataset and benchmark for non-contact neonatal heart rate monitoring based on spatio-temporal neural networks. *Engineering Applications of Artificial Intelligence*, 106:104447, 2021. 2, 6

[13] Bin Huang, Chun-Liang Lin, Weihai Chen, Chia-Feng Juang, and Xingming Wu. A novel one-stage framework for visual pulse rate estimation using deep neural networks. *Biomedical Signal Processing and Control*, 66:102387, 2021. 2

[14] Rik Janssen, Wenjin Wang, Andreia Moço, and Gerard De Haan. Video-based respiration monitoring with automatic region of interest detection. *Physiological measurement*, 37(1):100, 2015. 2

[15] Joao Jorge, Mauricio Villarroel, Sitthichok Chaichulee, Alessandro Guazzi, Sara Davis, Gabrielle Green, Kenny McCormick, and Lionel Tarassenko. Non-contact monitoring of respiration in the neonatal intensive care unit. In *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*, pages 286–293. IEEE, 2017. 2

[16] João Jorge, Mauricio Villarroel, Hamish Tomlinson, Oliver Gibson, Julie L Darbyshire, Jody Ede, Mirae Harford, John Duncan Young, Lionel Tarassenko, and Peter Watkinson. Non-contact physiological monitoring of post-operative patients in the intensive care unit. *NPJ digital medicine*, 5(1):4, 2022. 2, 7

[17] Dominique Makowski, Tam Pham, Zen J Lau, Jan C Brammer, François Lespinasse, Hung Pham, Christopher Schölzel, and SH Annabel Chen. Neurokit2: A python toolbox for neurophysiological signal processing. *Behavior research methods*, pages 1–8, 2021. 4

[18] Carlo Massaroni, Daniela Lo Presti, Domenico Formica, Sergio Silvestri, and Emiliano Schena. Non-contact monitoring of breathing pattern and respiratory rate via rgb signal measurement. *Sensors*, 19(12):2758, 2019. 2

[19] Carlo Massaroni, Daniel Simões Lopes, Daniela Lo Presti, Emiliano Schena, and Sergio Silvestri. Contactless monitoring of breathing patterns and respiratory rate at the pit of the neck: A single camera approach. *Journal of Sensors*, 2018, 2018. 7, 8

[20] Carlo Massaroni, Emiliano Schena, Sergio Silvestri, and Soumyajyoti Maji. Comparison of two methods for estimating respiratory waveforms from videos without contact. In *2019 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6. IEEE, 2019. 2

[21] Carlo Massaroni, Emiliano Schena, Sergio Silvestri, Fabrizio Taffoni, and Mario Merone. Measurement system based on rbg camera signal for contactless breathing pattern and respiratory rate monitoring. In *2018 IEEE International Symposium on Medical Measurements and Applications (MeMeA)*, pages 1–6. IEEE, 2018. 2

[22] Scott L Rossol, Jeffrey K Yang, Caroline Toney-Noland, Janine Bergin, Chandan Basavaraju, Pavan Kumar, and Henry C Lee. Non-contact video-based neonatal respiratory monitoring. *Children*, 7(10):171, 2020. 2

[23] Jae Kyu Suhr. Kanade-lucas-tomasi (klt) feature tracker. *Computer Vision (EEE6503)*, pages 9–18, 2009. 4

[24] Zhaodong Sun and Xiaobai Li. Contrast-phys: Unsupervised video-based remote physiological measurement via spatiotemporal contrast. In *European Conference on Computer Vision*, pages 492–510. Springer, 2022. 4, 5, 6, 7, 8

[25] Mauricio Villarroel, Sitthichok Chaichulee, João Jorge, Sara Davis, Gabrielle Green, Carlos Arteta, Andrew Zisserman, Kenny McCormick, Peter Watkinson, and Lionel Tarassenko. Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit. *NPJ digital medicine*, 2(1):128, 2019. 2

[26] Changchen Zhao, Menghao Zhou, Zheng Zhao, Bin Huang, and Bing Rao. Learning spatio-temporal pulse representation with global-local interaction and supervision for remote prediction of heart rate. *IEEE Journal of Biomedical and Health Informatics*, 2023. 2