Biomedical Signal Processing and Control 103 (2025) 107445

Contents lists available at ScienceDirect

Biomedical Signal Processing and Control

journal homepage: www.elsevier.com/locate/bspc

A comprehensive evaluation of multiple video compression algorithms for preserving BVP signal quality

Caiying Zhou [a], Xiaolang Ye [a,b], Yuanwang Wei [b,c,d,e,*], Vincenzo De Florio [f], Hong Sun [b,c,d,e,*], Xinlong Zhan [a], Yonggang Li [e], Chaochao Wang [e], Xianchao Zhang [b,c,d,e]

[a] College of Science, Jiangxi University of Science and Technology, Jiangxi, GanZhou, 341000, China
[b] Provincial Key Laboratory of Multimodal Perceiving and Intelligent Systems, Jiaxing University, ZheJiang, JiaXing, 314001, China
[c] Key Laboratory of Medical Electronics and Digital Health of Zhejiang Province, Jiaxing University, ZheJiang, JiaXing, 314001, China
[d] Engineering Research Center of Intelligent Human Health Situation Awareness of Zhejiang Province, Jiaxing University, ZheJiang, JiaXing, 314001, China
[e] Institute of Information Network & Artificial Intelligence, Jiaxing University, ZheJiang, JiaXing, 314001, China
[f] Evolution, Complexity and COgnition Research Group, Vrije Universiteit Brussel, Brussels, B-1160, Belgium

ARTICLE INFO

Keywords:
Remote Photoplethysmography (rPPG)
Video compression algorithms
Blood Volume Pulse (BVP)
PSNR

ABSTRACT

Remote Photoplethysmography (rPPG) detects heartbeat-induced skin color changes via camera to extract the Blood Volume Pulse (BVP) signal, which helps measure vital signs like heart and respiratory rates. Despite its popularity in health research for its user-friendly and noninvasive approach, rPPG's BVP signal quality can be compromised by factors like video compression. This research seeks to evaluate how various compression methods affect BVP signal quality, aiming to enhance rPPG's practical use. We created the ZJXU-MOTION dataset to assess video compression's impact on Blood Volume Pulse (BVP) signal quality during motion, considering activity and lighting. We compared common codecs like H.264, H.265, AV1/VP9, MJPEG, ProRes, and FFV1, analyzing how compression artifacts and bitrates influence BVP. Our results were confirmed with the UBFC-Phys dataset. Inter-frame compression can degrade BVP signal quality by adding noise. For clear signals, use intra-frame compression like H.265 with a GOP size of 1. If using inter-frame, keep GOP small and CQP stable. For static videos, H.265 CQP or VP9 VBR with GOP sizes of 1–3 work well. For low-motion videos, H.265 provides better performance. For complex motion, lossless compression methods like FFV1 outperform H.264 and H.265 in preserving BVP signal quality. Given that inter-frame compression introduces noise into BVP processing, we recommend intra-frame compression – particularly H.265 with a GOP of 1 – for optimal signal extraction. For complex motion, H.265 and FFV1 offer robust solutions, with FFV1 being especially effective in high-motion conditions. These findings are instrumental for advancing rPPG technology in health monitoring and diagnostics.

1. Introduction

In light of the swift advancement in the fields of computer vision and machine learning techniques [1–6], methods for measuring physiological parameters based on rPPG have attracted increasing attention from researchers due to their non-intrusive, contactless, low-cost, and user-friendly characteristics [7–9]. By analyzing BVP signals in both time and frequency domains, it is possible to calculate physiological parameters such as heart rate (HR) [10–15], respiratory rate (RR) [16,17], blood pressure (BP) [18–20], and heart rate variability (HRV) [21–23]. Therefore, extracting high-quality BVP signals is crucial for the accurate estimation of these physiological parameters, and there are numerous factors that can affect the quality of BVP signal extraction, including head movements, changes in ambient lighting, and the video compression coding used during video capture and transmission, among others. Despite the emergence of a large amount of literature on rPPG in recent years, most of the literature focuses on how to eliminate or reduce interference caused by head movements or changes in ambient lighting. There is a limited number of studies on how different video compression coding affects the quality of BVP signal extraction. The main reasons for this are as follows:

(1) **Lack of relevant research literature**: Existing literature mainly focuses on the basic principles, methods, and applications of heart rate estimation. Research on how video compression technology affects the extraction and analysis of BVP signals and how

* Corresponding authors.
E-mail addresses: yuanwang_wei@zjxu.edu.cn (Y. Wei), hongsun@zjxu.edu.cn (H. Sun).

https://doi.org/10.1016/j.bspc.2024.107445
Received 15 September 2024; Received in revised form 7 December 2024; Accepted 27 December 2024
Available online 7 January 2025

Table 1
Public-domain Databases for rPPG (RAW stands for uncompressed, RR stands for respiration rate).

Dataset	Video duration	Resolution	Fps	Provided signals	Codec	Bit rate (Mb/s)
HCI-Tagging	30 s	780 × 580	61	ECG	H.264	≈4.198
PURE	1 min	640 × 480	30	BVP	RAW	–
MMSE-HR	30 s	1040 × 1392	25	HR	–	–
COHFACE	1 min	640 × 480	20	BVP, RR	MPEG-4 Visual	0.251
OBF	5 min	1920 × 1080	60	ECG, BVP	–	–
ECG-Fitness	1 min	640 × 480	30	ECG, RR	RAW	211/218
VIPL-HR	30 s	1280 × 720	30	HR	MJPG	≈20
UBFC-rPPG	1 min	640 × 480	30	BVP, HR	RAW	211/218
UBFC-Phys	3 min	1024 × 1024	35	BVP, EDA	MJPG	220

to optimize heart rate estimation algorithms to adapt to different compression conditions is relatively scarce. This leads to a lack of sufficient theoretical support and practical guidance for researchers when designing and improving heart rate estimation algorithms. Compared to other factors, there is relatively little research on the impact of different video compression codecs on the quality of BVP signal extraction. In 2017, McDuff et al. [24] explored the effect of different bitrates (file sizes) and video quality settings on video pulse recovery rate by comparing two popular video compression algorithms, x264 and x265, adjusting the Constant Rate Factor (CRF) parameter. The study showed that in dynamic scenes, H.265-encoded videos performed better in recovering BVP signals than H.264-encoded videos. In 2018, Niu et al. [25] conducted a comparative analysis of five common video codecs – MJPG, FMP4, DIVX, PIM1, and X264 – during the construction of a dataset and ultimately chose the MJPG codec, which performed better on the dataset. However, this study used only the CHROM algorithm [26] to calculate the RMSE metric, and the experimental details were insufficient, especially in the analysis of motion scenarios. Since 2018, with the rise of deep learning methods in the rPPG field, the scale of datasets has expanded rapidly, and video compression codecs have become an effective means of optimizing storage space. Nevertheless, few studies have delved into the specific impact of video encoding on datasets. In 2019, C. Zhao et al. [27] proposed an emerging video compression strategy called POSSC, which significantly compresses non-facial regions to achieve high compression ratios. However, this approach may weaken the performance of rPPG methods that rely on background information. Therefore, more detailed research is urgently needed to explore the potential impact of different video compression codec parameters, motion intensities, and other variables on the quality of BVP signals.

(2) **Lack of relevant datasets**: Currently, most datasets do not provide raw uncompressed video data or standard BVP signals, making them unsuitable for conducting research on the impact of different video compression codecs on the quality of extracted BVP signals. We conducted a comparative analysis of existing publicly available datasets, as shown in Table 1. Datasets such as HCI-Tagging [28], MMSE-HR [29], ECG-Fitness [30], and VIPL-HR [31] do not provide standard reference BVP signals, they do not provide raw uncompressed video data. Furthermore, although PURE [32], COHFACE [33] and UBFC-rPPG [34] provide raw uncompressed video data and standard reference BVP signals, their resolutions are relatively low, at only 640 × 480. These datasets are not particularly suitable for conducting research on the impact of different video compression codecs on the quality of extracted BVP signals. The UBFC-Phys dataset [35], on the other hand, provides videos compressed with Motion JPEG, with a frame resolution of 1024 × 1024, and also includes standard reference BVP signals, making it very suitable for studying the effects of different video compression codecs on the quality of extracted BVP signals. The OBF dataset [36] is similar to the UBFC-Phys dataset, but it is not publicly available.

In summary, although research on rPPG-based physiological parameter measurement methods has increased in recent years, there is still relatively little research on the impact of different video compression codecs on the quality of BVP signal extraction during video capture and transmission. Furthermore, most publicly available datasets feature videos recorded under conditions where subjects remain stationary or only slightly move their heads, lacking videos of intense motion states, especially those dedicated to exercise scenarios. To address these shortcomings, this paper delves into the impact of different video compression codecs on the extraction of BVP signals, with the main contributions as follows:

(1) **We constructed a multi-scenario dataset named ZJXU-MOTION, specifically designed to explore the impact of different video compression codecs on the quality of BVP signal extraction.** This dataset encompasses diverse lighting conditions, motion amplitudes, shooting angles, and infrared videos, aiming to provide abundant resources for research on video compression codecs and training rPPG deep learning methods. The diversity of the ZJXU-MOTION dataset, especially in its motion scenarios, presents both challenges and opportunities for research on the adaptability of rPPG algorithms to motion.

(2) **Under the premise of maintaining video resolution, we conducted an in-depth exploration of the impact of mainstream video compression codecs (including H.264, H.265, AV1/VP9, MJPEG, ProRes, and FFV1), different bitrate controllers, and GOP values on the quality of BVP signal extraction.** By carefully selecting balanced codecs to compress the dataset, this study provides a novel perspective. Its findings are not only of significant value to research on BVP signal extraction in this field but also offer important references for video encoding methods in other video-related datasets.

2. Data acquisition

To better evaluate the impact of different video compression codecs on BVP signal extraction, we first constructed an uncompressed high-bitrate dataset. Subsequently, we processed this dataset using various video compression techniques to analyze the specific effects of different codecs on BVP signals.

2.1. Data acquisition environment setup

In motion scenarios, the sources of noise are diverse and complex, primarily including the following aspects: first, motion artifacts, typically manifesting as natural body movements such as vertical head movements; second, environmental light changes, involving fluctuations in light direction and intensity; third, the impact of video compression algorithms, as different compression techniques may introduce additional noise; fourth, the use of cosmetics, including different types and application thicknesses, which can interfere with image quality; and finally, the inherent noise of the camera itself. These factors collectively pose challenges to image clarity and signal accuracy [37].

Table 2
Experimental setup for rPPG data collection.

Group	Motion	Lighting	Camera	Number of video
01	P_s	L_1, L_2, L_3	C_1, C_2, C_3	$20 \times (6RGB + 1Inf)$
02	P_w	L_1, L_2	C_1, C_2, C_3	$20 \times (4RGB + 1Inf)$
03	P_r	L_1, L_2	C_1, C_2, C_3	$20 \times (4RGB + 1Inf)$

To effectively reduce the impact of noise [25,38], we adopted a multi-camera strategy for video recording to enhance the effectiveness and accuracy of rPPG signals. Additionally, following the framework proposed in [39], we incorporated an infrared camera as an input strategy. In the actual recording environment, we configured the following camera setup: RGB Camera 1 (C_1), RGB Camera 2 (C_2), and Infrared Camera (C_3). C_1 and C_3 were of the same model, with consistent placement height and vertical angle, and the recorded image resolution and frame rate were also identical. This configuration aimed to improve the quality and reliability of signal acquisition through synchronized multi-angle and multi-spectral recording.

To deeply study how motion artifacts interfere with the accuracy of experimental data, we systematically recorded video footage under three distinct lighting conditions to capture the subtle effects of light changes on artifacts. The specific operations were as follows: (1) Natural lighting conditions (L_1): videos were recorded during the daytime, using abundant natural sunlight as the light source. (2) Experimental lighting conditions (L_2): at night, we carefully designed and implemented a multi-directional LED lighting layout to ensure balanced and sufficient illumination in the study area. (3) Low lighting conditions (L_3): we deliberately selected a dimly lit room or corners with uneven LED lighting distribution to simulate insufficient lighting environments. Through these controlled experiments, we were able to observe the behavior characteristics of motion artifacts under different lighting levels in detail, and thus accurately quantify their potential impact on experimental results, laying a solid foundation for subsequent analysis.

To explore the impact of different exercise intensities and video compression codecs on BVP signal extraction quality, we carefully designed the following three scenarios on a treadmill: (1) Static state (P_s): the participant remains completely still, serving as the baseline condition. (2) Slow walking state (P_w): the treadmill speed is set to a slow walking pace, simulating a casual stroll. (3) Jogging state (P_r): the treadmill is adjusted to a suitable jogging speed to test the effects of light aerobic exercise.

As shown in Table 2, the experimental design included different combinations and total sample sizes. During the experiments, all participants refrained from using cosmetics and did not wear masks to ensure complete exposure of their forehead and cheek areas. Given the importance of privacy protection, we processed the recorded experimental videos by applying Gaussian blur to the participants' eye and mouth areas and overlaying the contours of these features. This measure aimed to protect personal privacy while not affecting the accurate detection of 68 facial landmarks by image recognition algorithms [2,6,40].

2.2. Device setting

Fig. 1 shows an overview of the experimental setup. The experiment took place in an indoor environment where participants stood on a treadmill, with a wrist-worn PPG sensor. Lighting equipment was positioned slightly above and directly in front of the participants' faces to ensure adequate illumination. Three different types of cameras were installed 1.5 m away from the participants, positioned parallel to the front of the treadmill. These cameras were adjusted to the height of the participants' faces for optimal viewing angles. Specifically, an infrared camera was centrally positioned, with RGB cameras placed on each side. This layout was designed to comprehensively capture experimental data.

(1) Under static testing conditions, participants are requested to perform a unified posture: standing naturally on the treadmill, maintaining a forward-facing posture to the best of their ability. Auxiliary bulbs were placed around the cameras to improve lighting conditions, so as to ensure adequate and unified illumination.
(2) In the walking scenario, to mitigate variations in the strides of different participants, the treadmill was set to synchronize the participants' pace with a metronome. The target pace was approximately 78 steps per minute for walking, and the speed is adjusted to ensure comfort for each participant so as to minimize the variance between different participants.
(3) In the running scenario, participants jogged on the treadmill at a pace synchronized with a metronome, set to approximately 120 steps per minute, with the flexibility to adjust the speed for individual comfort, so as to minimize the variance between different participants.

In the experimental recordings under the P_r condition, the PPG signals collected by the finger-type PPG sensor fluctuate due to motion interference. To mitigate the effects of motion artifacts and signal delay on PPG signal acquisition, we specifically used a fixed device to stabilize the participant's arm. This device reduces minor displacements between the participant's finger and the sensor during movement, thereby minimizing signal instability.

2.3. Data processing

In the P_s video, the subject's hand remains mostly stationary, resulting in PPG signals with minimal noise during the recording experiment. In contrast, in the P_w and P_r motion scenarios, the swinging of the hand causes significant PPG signal noise, necessitating signal restoration processing. The noise in the PPG signal mainly stems from the periodic swinging of the hand during running, with a non-fixed frequency. Low-pass filtering in the experiment did not yield significant effects.

To effectively remove artifacts caused by hand and sensor movement during exercise, the PPG signal is first decomposed. Common methods for signal decomposition include Singular Spectrum Analysis (SSA) [41,42], Single Channel Independent Component Analysis (SCICA) [43,44], and Variational Mode Decomposition (VMD) [45,46]. These methods decompose the signal based on different criteria and procedures. In our experiments, the VMD method was found effective in decomposing the PPG signal into two fixed components, IMF1 and IMF2, which effectively separate noise. Adjusting VMD parameters for different exercise modes helps mitigate noise peaks, although this decomposition process reduces other information contained in the original signal. Nevertheless, the frequency peaks align more closely with ECG frequencies, effectively removing many noise artifacts. Fig. 2 illustrates this decomposition, where IMF2 closely matches the frequency of hand movements during running.

2.4. Participants

The constructed ZJXU-MOTION dataset includes 280 uncompressed RGB videos and 60 near-infrared videos from 20 participants (10 males and 10 females), aged between 20 and 45 years. Each video has a recording length of approximately 1 min.

- **Video Summary:** $Size \approx 1500G$, $Fps = 30$, $Resolution = 1024 \times 1024$, $Bitrate \approx 755M/s$
- **Pace:** The cycles of walking and running swing up and down, influenced by factors such as treadmill pace and experimenter's leg length. During the experiment, the pace was adjusted to maintain approximately consistent cycles based on leg length.
- **Gender:** $50\%\ M, 50\%\ F$.

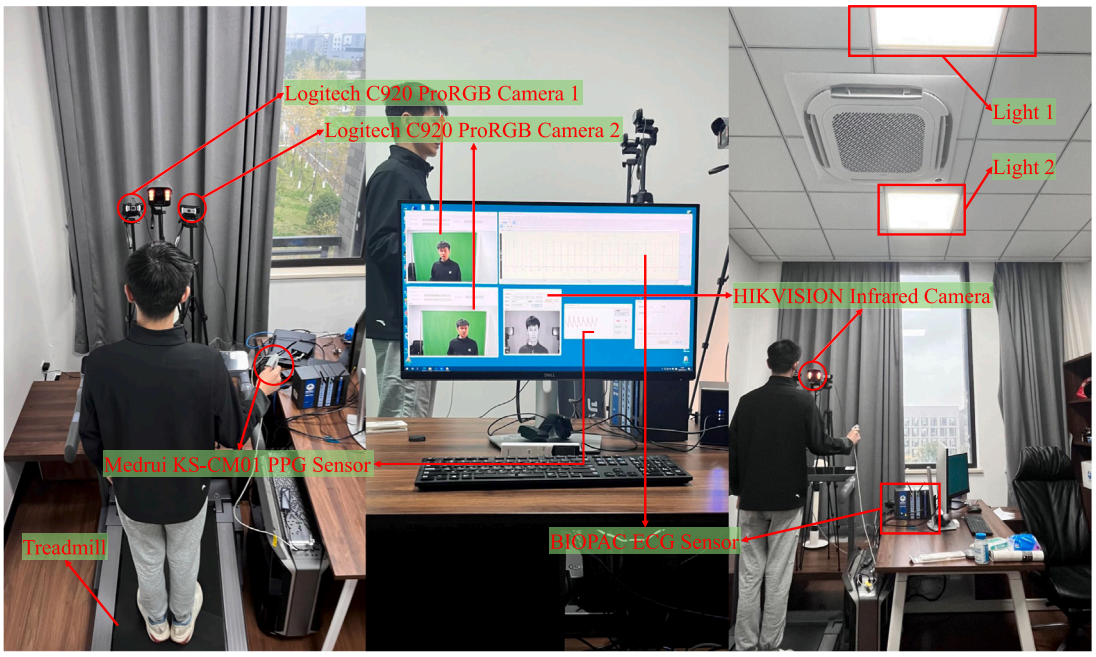

Fig. 1. Schema of the device Setting.

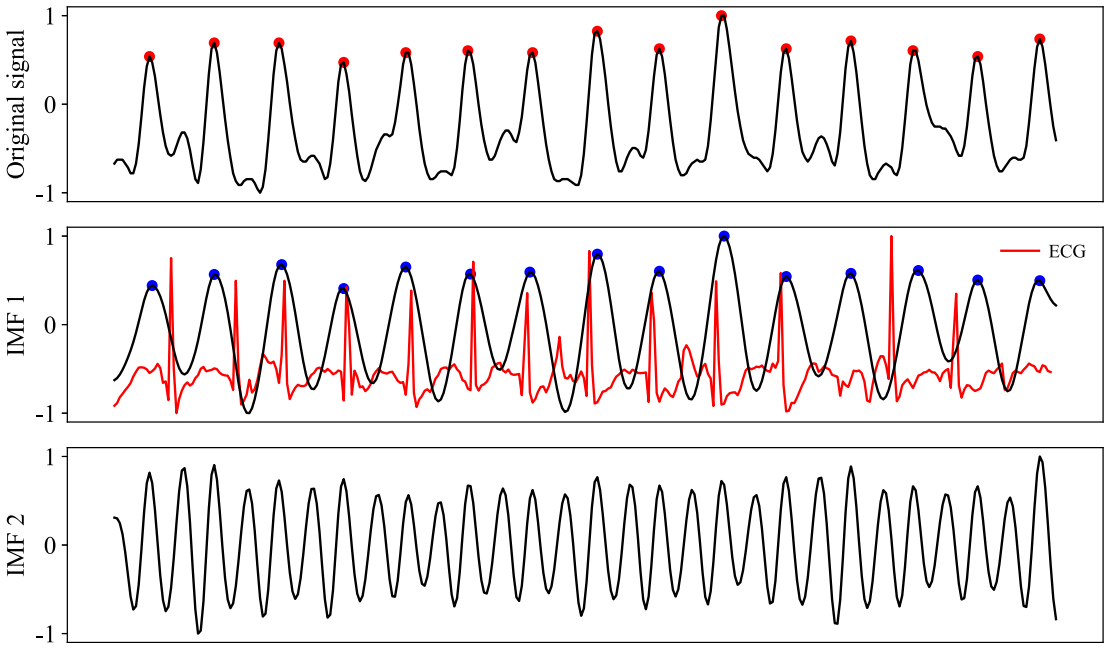

Fig. 2. The original bvp signal, VMD component and ECG signal.

- **Skin Color:** The majority of participants are from mainland China, predominantly exhibiting yellowish skin tones with a skewed distribution. The results demonstrate skin tone universality, with a histogram of facial skin color distribution shown in Fig. 3.
- **Exercise Heart Rate:** Intense exercise induces variations in overall heart rate. Box plots of heart rates during different exercise states are depicted in Fig. 4.

2.5. Data format

To enhance the storage efficiency of the dataset, we adopted a standardized data format, significantly reducing storage space. Each video is associated with a unique identifier pointing to a binary file with the "$.info$" extension. This file structure consists of a fixed-size 8-byte file header followed by a variable-length file body. Details of the information contained in the file are outlined in Table 3, ensuring synchronization of information duration with the respective videos and consistent sampling intervals. Additionally, we provide a dedicated Python script for parsing and processing the information within these "$.info$" files to facilitate data interpretation and utilization.

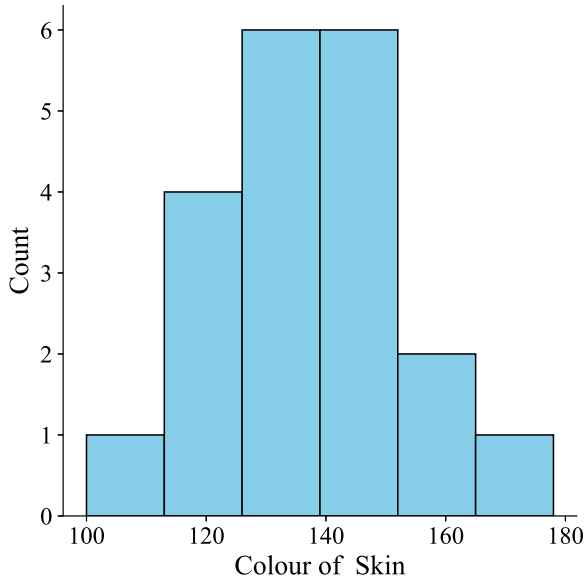

Fig. 3. Skin color distribution histogram.

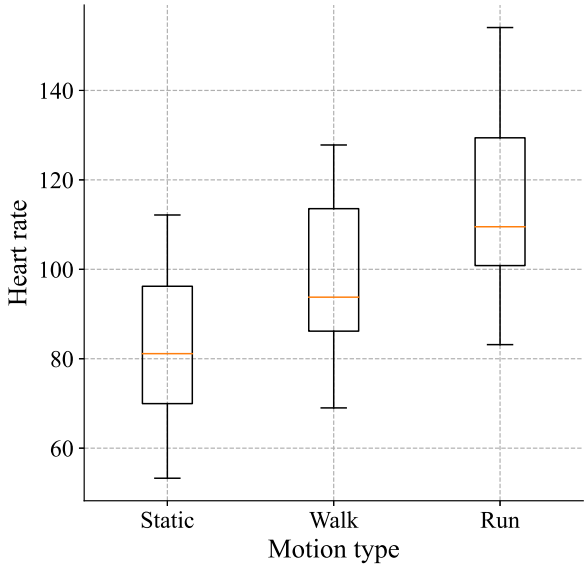

Fig. 4. Heart rate of different motion type.

Table 3
Data format information.

Size/bit	Describe
16	Number of BVP data
16	Number of SO_2 data
16	Number of Pr data
16	Number of PI data
24	Number of ECG data
16	Unset, reservation
*	BVP data,read by 8 bits
*	SO_2 data,read by 8 bits
*	Pr data,read by 8 bits
*	PI data,read by 8 bits
*	ECG data,read by 32 bits (float32)

The final data set is encoded according to the experimental results. For the P_s dataset, the VP9 CRF is suitable with a compression ratio of 10% to 15%. For the P_w dataset, the H.265 CQP is preferable with a compression ratio of 15% to 20%. However, for the P_r dataset, dynamic encoding is not suitable, and encoding using FFV1 is recommended with a compression ratio of 30%.

3. Evaluation of video compression algorithms

To quantify and analyze the specific impact of various video encoding schemes on the quality of BVP signal extraction from videos, we designed the experimental workflow illustrated in Fig. 5. The core process involves constructing a database composed of raw videos, which is then used to train a BVP signal extraction model. Subsequently, we input video samples compressed by different encoding techniques into the aforementioned model to obtain predicted BVP signals. These predicted signals are compared with the actual BVP signals, and the superior performance between the two sets of results is used as the evaluation metric to determine the impact of the encoding method on the signal extraction quality. To further confirm the robustness and general applicability of our experimental conclusions, we also repeated the validation on subsets T1 and T3 of the UBFC-Phys dataset, ensuring the trustworthiness and consistency of our research findings.

3.1. Evaluation methods

We first designed a set of evaluation metrics to measure the specific impact of different video encoding strategies on the quality of BVP signals. Through comparative analysis, we can select the most suitable video compression codec for rPPG applications. Suppose there are two $m \times n$ RGB color images I and K, where the maximum pixel value of each RGB channel is L. Based on this, we will use image processing and signal analysis techniques to assess the potential effects of the encoding process on BVP signal recovery capability.

In the field of objective image quality assessment, performance measurement mainly relies on two major indicators: the Structural Similarity Index (SSIM) and the Peak Signal-to-Noise Ratio (PSNR). SSIM focuses on evaluating the structural similarity between images, while PSNR is a standard method for quantitatively analyzing the difference between the original image and the compressed or processed image. SSIM captures the consistency of local structural information of the image, whereas PSNR reflects the degree of image distortion by calculating the logarithmic form of the maximum possible pixel intensity error (usually represented by the mean squared error between the original image and the compressed or distorted image).

The combination of SSIM and PSNR allows for a comprehensive assessment of image quality, considering both the structural similarity in terms of visual perception and the signal purity in quantitative terms. The mean squared error between the original and the compressed or distorted image (in dB) is defined as follows:

$$\widehat{MSE} = \frac{1}{nm}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I_{diff}(i,j)]^2 \tag{1}$$

where

$$I_{diff} = I - K \tag{2}$$

In this context, I_{diff} represents the difference between images I and K, and MSE is the mean squared error between images I and K. Therefore, $PSNR$ can be calculated from MSE using the following formula:

$$PSNR = 10 \times log_{10}\left(\frac{L^2}{Avg(\widehat{MSE})}\right) \tag{3}$$

Generally speaking, images with $PSNR < 30dB$ have poor quality and noticeable distortion to the naked eye; images with PSNR between

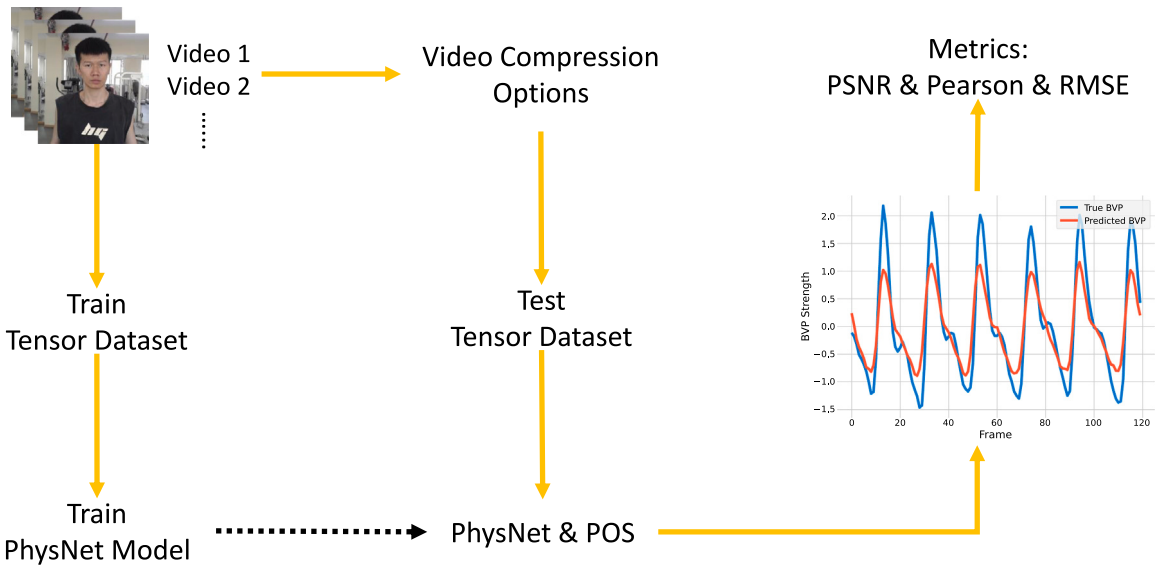

Fig. 5. The experimental workflow of the method in this paper.

$30dB$ and $40dB$ have moderate quality with some distortion, but are acceptable; images with PSNR between $40dB$ and $50dB$ have good quality with minor distortion that is hard to detect; and images with $PSNR > 50dB$ have excellent quality with almost no distortion.

Inspired by literature [47], we explored and implemented various BVP signal restoration techniques. Through comparative analysis, we identified two strategies that performed exceptionally well: the classic traditional method (POS) and an innovative approach based on deep learning (PhysNet). Notably, the PhysNet model was comprehensively trained on uncompressed datasets under different motion scenarios. To objectively evaluate the effectiveness of these methods in processing BVP signals, we used the Pearson correlation coefficient as key performance metrics. The specific calculation formulas are shown in Eq. (4) and Eq. (5).

$$Pearson = \max(pearson(bvp_{pos}), pearson(bvp_{physnet})) \tag{4}$$

$$pearson(bvp) = \frac{N\sum bvp_i bvp'_i - \sum bvp_i \sum bvp'_i}{\sqrt{N\sum bvp_i^2 - (\sum bvp_i)^2}\sqrt{N\sum (bvp'_i)^2 - (\sum bvp'_i)^2}} \tag{5}$$

where bvp_i represents ground truth signal, while bvp'_i represents the predicted signal.

3.2. Compression codec

MPEG-4 encompasses a series of compression technologies, with the XviD and H.264/AVC [48] codecs representing the second part (MPEG-4 Part 2) and the tenth part (MPEG-4 Part 10) of the MPEG-4 standard, respectively. XviD is suitable for general video compression needs but is not suitable as a compression codec for datasets. On the other hand, H.264/AVC is more commonly used in scenarios requiring higher compression efficiency and better video quality.

H.265/HEVC [49] is a video encoding standard that succeeded H.264/AVC, offering higher efficiency in video compression, better video quality, and greater efficiency than H.264/AVC at the same bitrate. With the same compression ratio, H.265/HEVC achieves higher compression rates. These three codecs are more universally applicable and popular among researchers for decompression, offering good compression rates.

VP9/AV1 [50,51] boasts high compression rates and is an emerging open-source, royalty-free video compression format.

The key characteristics of the main encoding technologies used in this study are shown in Table 4. Unlike MJPG, which encodes video as a series of independent still images, the video encoding technologies we selected serialize consecutive images into a Group of Pictures (GOP). Each GOP forms a sequence unit, starting with an I-frame (intra-coded frame or keyframe) followed by several B-frames (bi-predictive frames) or P-frames (predictive frames). This encoding strategy allows for high video compression, but it comes at the cost of potentially significant effects on encoding performance during rapid motion changes, especially when accurately constructing the keyframes.

Table 4
Introduction of compression coding.

Coder /Decoder	Description	Video type
H.264(AVC)	This video codec is MPEG-4 Part 10, commonly known as H.264/AVC. It is used for online video streaming, Blu-ray discs, and other applications. It serves as the foundation for many popular video encoding and decoding standards.	.mp4/.mkv /.avi
H.265(HEVC)	It provides better support for high-resolution videos such as 4K and 8K but requires more computational resources.	.mp4/.mkv
AV1	Developed by AOMedia, it provides more efficient compression than VP9, with over 30% bit rate reduction without sacrificing video quality or hardware feasibility.	.webm/.mkv
VP9	It was released earlier than AV1, so it is more widely supported on some platforms.	.webm/.mkv

In the experiments conducted on the P_s and P_w datasets, we encoded the videos by adjusting the CRF parameter to obtain video versions with different compression ratios. As revealed in the experiments reported in [24], even setting the CRF to the minimum value of 1 has a significant impact on video quality evaluation metrics. As shown in Fig. 6, during the video compression process, the number of I-frames, B-frames, and P-frames in the video dynamically changes with different motion scenes. The I-frames, which do not rely on other frames for decoding, exhibit the best signal quality with peaks at their positions. In contrast, P-frames and B-frames depend on I-frames or other P-frames for reference, and there is inherent loss during compression, resulting in relatively poorer signal quality. Furthermore, the frame quality of compressed videos fluctuates with the number of frames; as the video

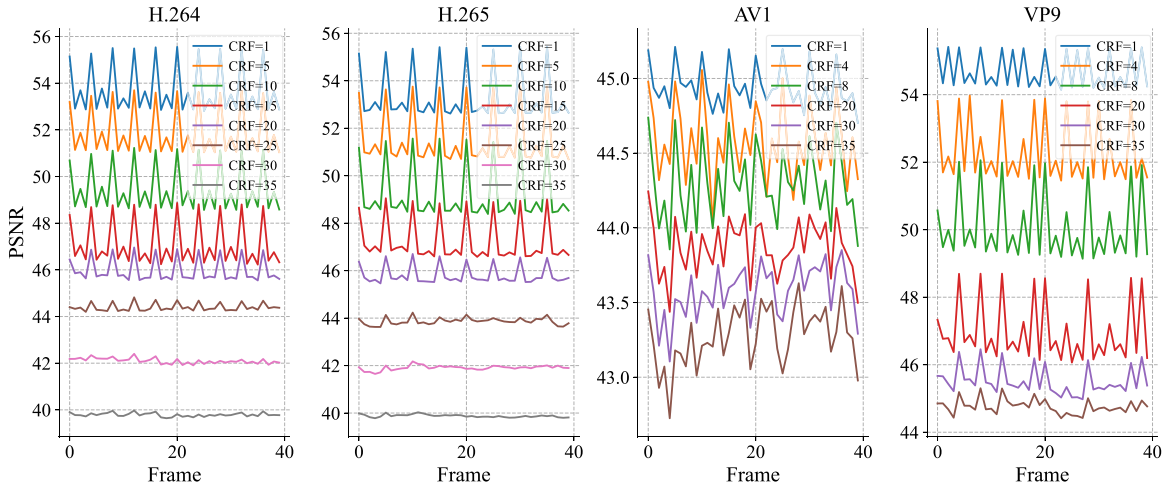

Fig. 6. The PSNR values for the first 40 frames of P_s videos encoded with H.264, H.265, AV1, and VP9 codecs.

quality decreases, its PSNR value correspondingly decreases, and the decreasing trend gradually stabilizes. It is worth noting that this quality fluctuation is closely related to the magnitude of motion, and different encoding schemes have varying adaptability in handling motion, among which the fluctuation of AV1 encoding is more pronounced in comparison.

3.3. Rate control mode

Video bitrate control, as a crucial component of encoders, adjusts the inter-frame compression strategy and is responsible for regulating the size of the bitstream output using specific algorithms, independently of encoding standards. In video compression practice, although many publicly available datasets prefer using the CRF mode for bitrate control, it may not always be the optimal choice. In fact, common bitrate control strategies include:

(1) **Variable Bit Rate (VBR)**: Adjusts bitrate dynamically based on scene complexity or motion. Higher bitrate is used for complex or highly dynamic scenes, and lower bitrate for simpler scenes.
(2) **Constant Bit Rate (CBR)**: Maintains a constant bitrate output regardless of scene complexity or motion, aiming to meet a specified target bitrate.
(3) **Constant Quantization Parameter (CQP)**: Uses the same quantization parameter for encoding each frame, maintaining consistent quality.
(4) **Constant Rate Factor (CRF)**: Adjusts the quality parameter based on scene complexity. It increases the quality parameter for scenes with more motion and decreases it for scenes with less motion.

Video scenes typically involve dynamic elements, and while CBR is suitable for network transmission, it performs poorly in scenarios where image quality fluctuations due to motion are significant. The present study examines and applies bitrate control technologies such as CRF, CQP, and VBR. The compression effects of the videos were evaluated using the I_{diff} metric, and specific results are depicted in Fig. 7. Given the varying effectiveness of different bitrate control schemes, our research not only delves into exploring compression encoding technologies but also emphasizes analyzing the specific impacts of these bitrate control strategies on encoder performance.

We used FFmpeg version 6.1.2 to compress the video. This tool provides a comprehensive command-line interface for video transcoding and supports multiple formats and containers. Table 5 lists the bitrate control methods and detailed parameters used in the experiments. It is worth noting that setting a too large GOP value can lead to decreased video quality. Multiple experiments have shown that setting GOP values between 1 and 3 achieves better compression quality with lower variance.

4. Results

4.1. Evaluation results on internal ZJXU-MOTION dataset

In this study, the datasets used were recorded in a lossless format to ensure the integrity and quality of the facial video data. Considering that different levels of motion intensity may have a significant impact on compression results, this article particularly took into account the limitations of existing models in handling lighting variations and motion recognition. Therefore, we selected the L1 dataset, which has good lighting conditions, and chose sample videos from stationary (P_s) and walking (P_w) states as the test set. Notably, during the experiments, it was found that the Pearson correlation coefficients of BVP signals obtained under running conditions (P_r) were lower, and the effect of noise on the Pearson values was random. Consequently, we did not include the running state (P_r) in the scope of further analysis.

In the selected test set videos, this study evaluated a series of commonly used lossy video compression codecs. By applying different video compression encoding strategies and adjusting bitrate control parameters, the aim was to explore optimal encoding configurations, striving to achieve effective data compression while maintaining video quality.

Fundamentally, the video compression process can be viewed as a mechanism introducing noise. By deeply analyzing the characteristics of noise, it is possible to effectively mitigate its negative effects during compression. In statistical terms, mean and variance are commonly used measures to describe trends and variability within datasets, and they are similarly applicable for assessing noise properties. In this study, different compression ratios corresponded to different levels of video compression quality. Under the condition of controlling GOP to 1, i.e., without inter-frame compression, Fig. 8 illustrates the trends in average Peak Signal-to-Noise Ratio (PSNR) and variance across four intra-frame compression encoding technologies as compression rates vary. From the figure, it can be observed that the average PSNR values of these four encoding technologies are comparable numerically. Particularly, until the average PSNR value reaches 55 dB (typically considered high-quality video above 50 dB), the performance of H.265 All I encoding technology consistently outperforms the other three technologies. Furthermore, the variance of the ProRes encoding technology is significantly higher than the other two encoding technologies, indicating that it introduces noise with greater variability during the compression process.

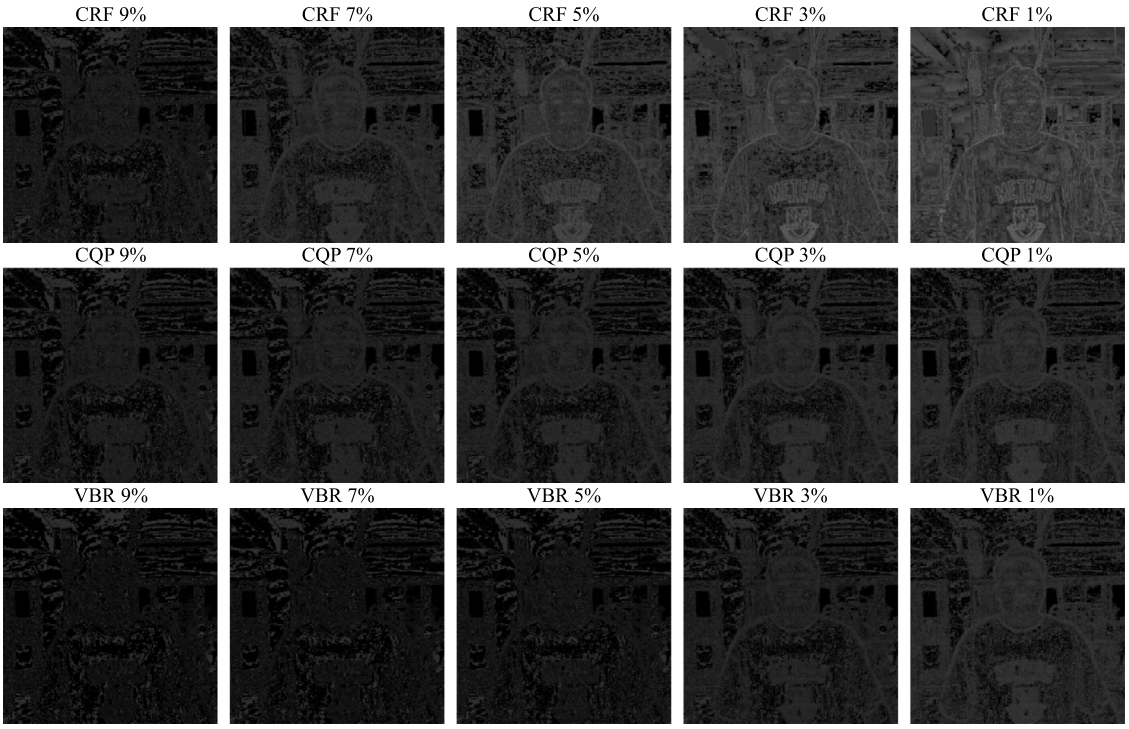

Fig. 7. H.264 encoding, CRF, CQP, VBR rate control method, comparison of I_{diff} after Gamma transformation (the same frame).

Table 5
FFmpeg parameter for four coding three rate control.

Coder/Decoder	Rate control	Parameter	-preset	-pix_fmt	-g	-tune
libx264	CRF/CQP	-[crf/qp] [value]	slow			psnr
	VBR	-b:v [value] -maxrate [value]				
libx265	CRF/CQP	-[crf/qp] [value]				
	VBR	-b:v [value] -maxrate [value]				
prores_ks	–	-profile:v [value]	———	yuv444p	1–3	———
mjpeg	–	-q:v value				
ffv1	–	–				
libsvtav1	CRF/CQP	-[crf/qp][value] -svtav1-params	1			1
	VBR	-b:v [value] -svtav1-params rc=1	1			1
libvpx-vp9	CRF	-crf [value] -b:v 0	1			
	VBR	-b:v [value] -maxrate [value]	1			

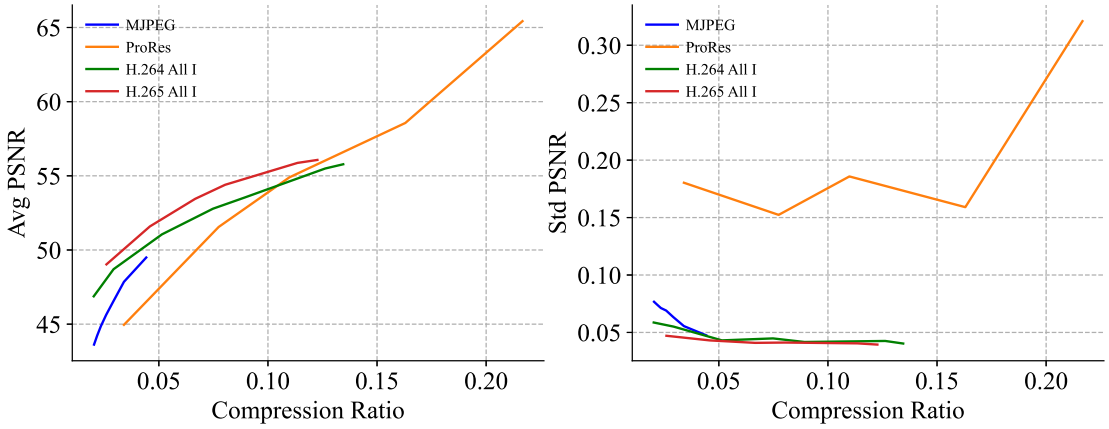

Fig. 8. Variations of Avg PSNR and Std PSNR with Compression Rate for P_s Videos, H.264 All I and H.265 All I ($GOP = 1$) Encoding (Results are Comparable to P_w).

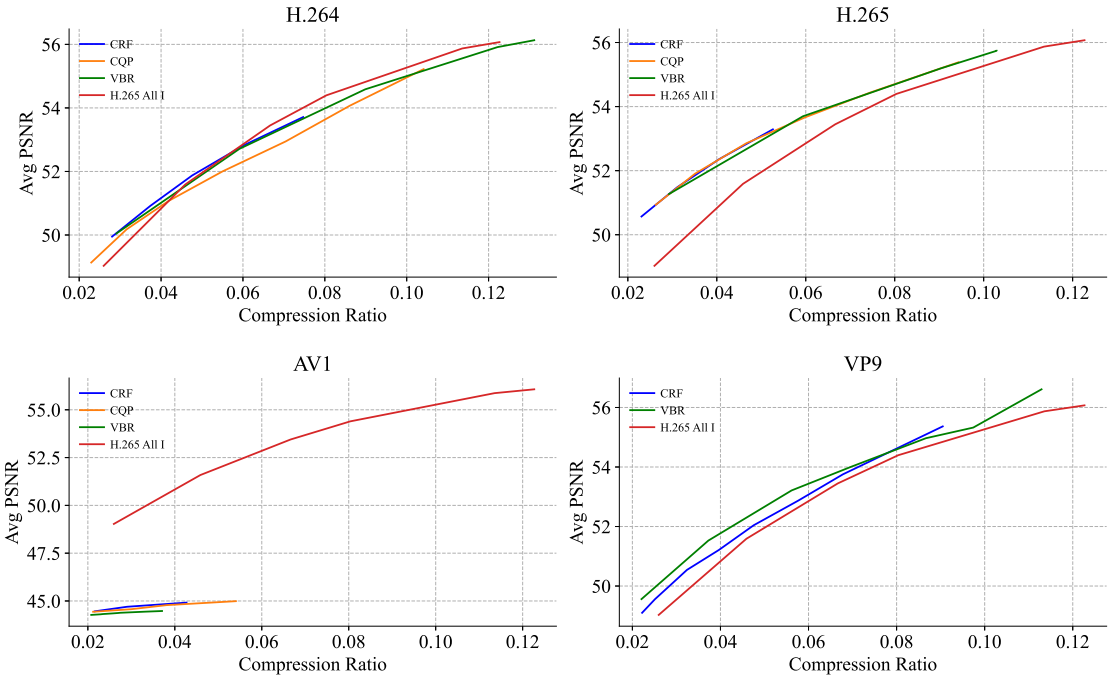

Fig. 9. The Avg PSNR for the P_s videos encoded with H.264, H.265, AV1, and VP9 codecs.

To explore the impact of three bitrate control methods on video compression quality, this study used H.265 All I encoding technology as the baseline. Fig. 9 show the trends in the average PSNR values with varying compression rates under different video compression settings in the P_s dataset. The results indicate that the PSNR values obtained using the VP9 encoding technology are generally superior to those obtained with other encoding methods, especially when the compression rate exceeds 25%, where its advantage becomes more pronounced.

Furthermore, in both the P_s and P_w datasets, the constant quantization parameter (CQP) mode always exhibits smaller variance compared to the variable bitrate (VBR) mode. This indicates that in H.264 and H.265 encoding technologies, the noise introduced by the constant quantization parameter (CQP) rate control method exhibits a higher level of stability. It is worth noting that inter-frame compression, due to the introduction of temporal redundancy, increases noise volatility. The variance data in Fig. 10 further confirm this, where intra-frame compression shows significantly lower variance than inter-frame compression until the compression rate reaches a high level, at which point the variance gradually decreases. Additionally, the AV1 encoding quality is relatively poor, resulting in less variance fluctuation.

In the experimental results, both for the P_s and P_w datasets, encoding using the Constant Quantization Parameter (CQP) mode consistently shows lower variance, while Variable Bit Rate (VBR) mode exhibits higher variance. This finding suggests that the CQP mode introduces less fluctuation in noise during the compression process. Therefore, under similar video quality (Avg PSNR), employing CQP mode can effectively mitigate the impact of noise and enhance signal stability.

Finally, we calculated the Pearson correlation values to evaluate the impact of different encoding methods on the quality of BVP signal extraction. The results are summarized in Table 6, and Fig. 11 illustrates the most effective bitrate control methods for each of the four encoding techniques within the P_s dataset. The study revealed that when the compression rate falls below 4% (approximately corresponding to a PSNR value of 50), the Pearson values of the videos processed by different encoding methods exhibit a sharp decline. This phenomenon indicates that when the PSNR value exceeds 50, its impact on the Pearson value is relatively minimal.

Analysis of the P_s dataset shows that the VP9 encoding technology has a relatively small impact on the Pearson value, which can be attributed to its excellent encoding efficiency and noise control capabilities. For the P_w dataset, when the compression rate exceeds 4%, the H.264 CRF encoding method can ensure good compression quality, thereby maintaining a high Pearson value.

4.2. External validation on the UBFC-Phys dataset

To further confirm the robustness and generalizability of the experimental conclusions, we also conducted experiments on the UBFC-Phys dataset. It is noteworthy that the video material in this dataset uses MJPG encoding, ensuring high compression quality and providing a high-quality data source for analysis. The dataset is specifically designed with experimental conditions in which the subjects' bodies remain static, but controlled head rotations are introduced as variables. Specifically, the experimental subjects in the T1 task scenario are relatively stationary, while the T2 task scenario simulates dynamic scenes with slight head movements and accompanying verbal communication, with the motion amplitudes strictly limited below P_w to ensure consistency and repeatability in the experiments.

During the experimental preparation phase, to eliminate potential interference from encoding differences in subsequent analysis, we converted the video encoding in the dataset to uncompressed Raw format and mapped the color space to the standard RGB model, allowing for more precise visual information capture and analysis. The experimental results are shown in Figs. 12 and 13. As a key metric for image quality evaluation, the trend of Avg PSNR aligns with the curves drawn in the testing on our dataset. Further analysis revealed that in both the T1 and T3 scenarios, the variance of PSNR values using the CQP encoding mode was significantly lower than that using the VBR encoding mode, demonstrating the advantage of CQP mode in maintaining video quality stability. Therefore, in subsequent data processing and analysis, the CQP encoding mode is prioritized to ensure the reliability and accuracy of the experimental results.

In the UBFC-Phys dataset, a detailed analysis of the experimental results is presented in Table 6. To deeply explore the performance of different encoding strategies, we selected the most notable bitrate control methods for each encoding scheme and illustrated them in Fig. 14. Notably, when the compression rate falls below the 10% threshold (corresponding approximately to a PSNR of 50 dB), the Pearson

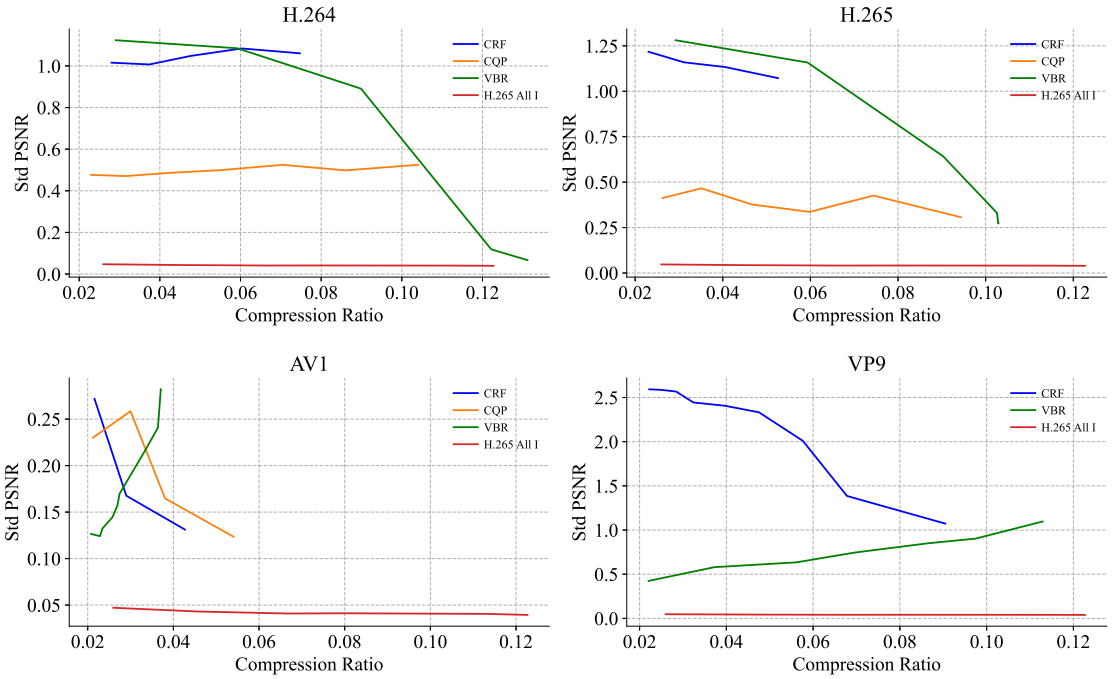

Fig. 10. The Std PSNR for the P_s videos encoded with H.264, H.265, AV1, and VP9 codecs.

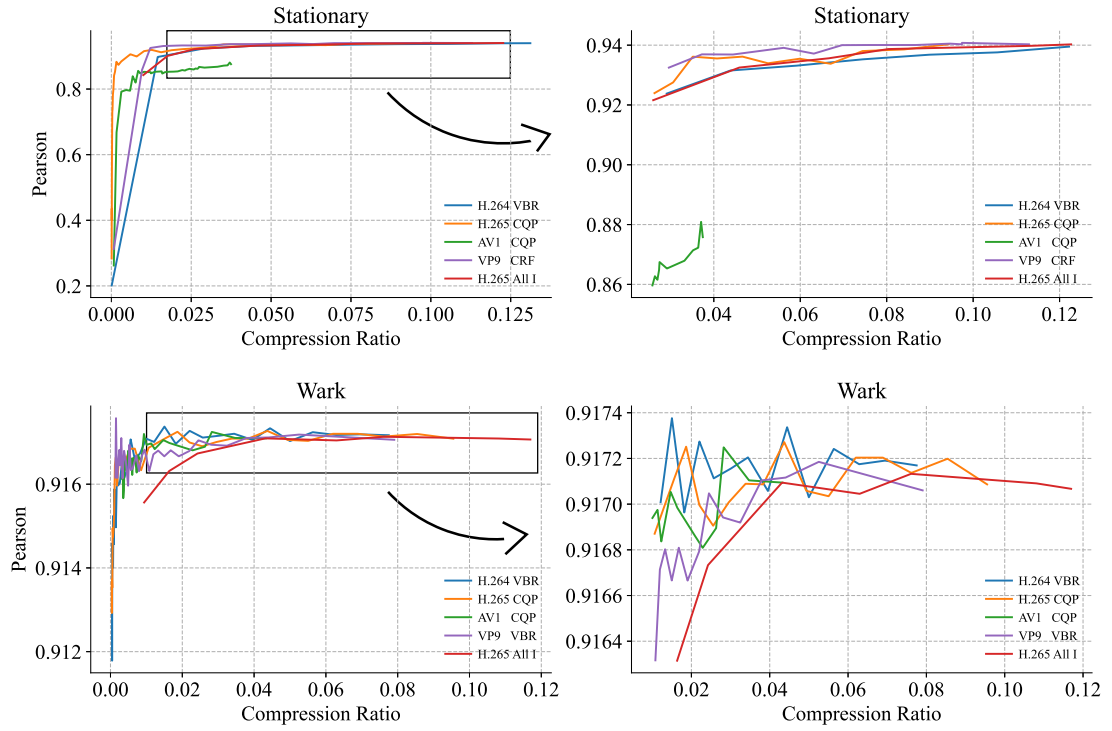

Fig. 11. Pearson correlation coefficients obtained from PhysNet and POS calculations on the compressed P_s and P_w dataset for four encoding methods, varying with compression rate.

values of the videos processed by each encoding scheme show a sharp decline. This phenomenon suggests that, under conditions where the PSNR exceeds 50 dB, the influence of PSNR on Pearson correlation tends to weaken. However, the robustness of this conclusion requires further consideration of the variance factor.

Upon further analysis of the P_s curve, we found that VP9 encoding, compared to other encoding methods, has a lesser negative impact on the Pearson correlation coefficient, demonstrating its advantage in maintaining data correlation. Shifting focus to the P_w curve, a significant feature is observed: when the compression rate exceeds the 15% threshold, H.264 using the CRF encoding mode can more effectively maintain the compression quality of the video content at a high level. This indicates that within this compression rate range, the H.264 CRF encoding strategy is the preferred choice for achieving

Table 6
Summary of experimental results.

Compression ratio (%)	H.264		H.265		AV1		VP9		H.265 All in I	
	R	RMSE	R	RMSE	R	RMSE	R	RMSE	R	RMSE
					P_s					
0.020	0.9076	0.4133	0.9195	0.3897	0.8533	0.5184	0.9323	0.3603	0.9069	0.4149
0.030	0.9243	0.3805	0.9271	0.3715	0.8659	0.4980	0.9327	0.3580	0.9238	0.3782
0.040	0.9295	0.3665	0.9356	0.3520	–	–	0.9369	0.3470	0.9293	0.3653
0.050	0.9322	0.3590	0.9349	0.3515	–	–	0.9379	0.3443	0.9331	0.3563
0.060	0.9333	0.3555	0.9354	0.3519	–	–	0.9380	0.3446	0.9346	0.3526
0.070	0.9346	0.3521	0.9355	0.3512	–	–	0.9400	0.3395	0.9364	0.3481
0.080	0.9358	0.3488	0.9383	0.3424	–	–	0.9400	0.3394	0.9386	0.3417
0.090	0.9368	0.3454	0.9396	0.3389	–	–	0.9402	0.3384	0.9390	0.3407
0.100	0.9373	0.3447	–	–	–	–	0.9407	0.3367	0.9393	0.3398
0.110	0.9381	0.3430	–	–	–	–	0.9404	0.3375	0.9397	0.3390
0.120	0.9393	0.3402	–	–	–	–	–	–	0.9401	0.3378
					P_w					
0.020	0.9171	0.3968	0.9169	0.3974	0.9169	0.3970	0.9167	0.3976	0.9165	0.3982
0.025	0.9171	0.3966	0.9169	0.3973	0.9169	0.3972	0.9170	0.3968	0.9167	0.3976
0.030	0.9172	0.3966	0.9168	0.3975	0.9172	0.3964	0.9169	0.3972	0.9168	0.3974
0.035	0.9172	0.3965	0.9169	0.3973	0.9171	0.3966	0.9170	0.3969	0.9169	0.3972
0.040	0.9171	0.3968	0.9170	0.3971	0.9171	0.3967	0.9171	0.3967	0.9170	0.3969
0.045	0.9173	0.3963	0.9171	0.3967	–	–	0.9171	0.3966	0.9171	0.3968
0.050	0.9170	0.3969	0.9171	0.3967	–	–	0.9172	0.3966	0.9171	0.3968
0.055	0.9172	0.3965	0.9172	0.3964	–	–	0.9172	0.3965	0.9171	0.3968
0.060	0.9172	0.3965	–	–	–	–	0.9171	0.3966	0.9171	0.3969
0.065	0.9172	0.3965	–	–	–	–	0.9171	0.3966	0.9171	0.3968
0.070	0.9172	0.3965	–	–	–	–	0.9171	0.3967	0.9171	0.3968
0.075	0.9172	0.3966	–	–	–	–	0.9171	0.3968	0.9171	0.3967
					UBFC-Phys T1					
0.020	0.9266	0.3655	0.9212	0.3764	0.8145	0.5712	0.9184	0.3855	0.8878	0.4375
0.032	0.9300	0.3557	0.9263	0.3646	0.8253	0.5555	0.9298	0.3582	0.9127	0.3942
0.044	0.9298	0.3575	0.9238	0.3681	0.8350	0.5402	0.9318	0.3519	0.9179	0.3826
0.056	0.9341	0.3466	0.9308	0.3537	–	–	0.9327	0.3499	0.9231	0.3711
0.068	0.9348	0.3442	0.9330	0.3481	–	–	0.9337	0.3477	0.9283	0.3595
0.080	0.9352	0.3431	0.9353	0.3435	–	–	0.9340	0.3455	0.9305	0.3543
0.092	0.9367	0.3401	0.9348	0.3448	–	–	0.9333	0.3474	0.9321	0.3505
0.104	0.9370	0.3382	0.9323	0.3499	–	–	0.9358	0.3412	0.9337	0.3467
0.116	0.9358	0.3404	0.9338	0.3469	–	–	0.9350	0.3433	0.9353	0.3429
0.128	0.9351	0.3421	0.9348	0.3446	–	–	0.9355	0.3420	0.9369	0.3391
0.140	0.9350	0.3428	0.9351	0.3432	–	–	0.9345	0.3446	0.9373	0.3379
0.152	0.9342	0.3455	0.9359	0.3412	–	–	0.9356	0.3414	0.9370	0.3384
0.164	0.9350	0.3427	0.9367	0.3392	–	–	0.9385	0.3360	0.9367	0.3388
0.176	0.9362	0.3392	0.9360	0.3408	–	–	0.9384	0.3356	0.9369	0.3382
0.188	0.9364	0.3390	0.9355	0.3419	–	–	0.9372	0.3377	0.9373	0.3372
0.200	0.9374	0.3365	0.9356	0.3411	–	–	0.9372	0.3383	0.9377	0.3362
0.212	0.9381	0.3347	0.9363	0.3394	–	–	0.9384	0.3350	0.9382	0.3352
0.224	0.9380	0.3353	0.9374	0.3367	–	–	0.9383	0.3352	0.9386	0.3342
0.236	0.9385	0.3335	0.9380	0.3352	–	–	0.9375	0.3370	0.9390	0.3332
0.248	0.9388	0.3326	0.9377	0.3356	–	–	0.9359	0.3405	0.9391	0.3329
					UBFC-Phys T3					
0.020	0.8862	0.4568	0.8890	0.4529	0.8712	0.4853	0.8879	0.4561	0.8867	0.4572
0.032	0.8903	0.4490	0.8917	0.4476	0.8679	0.4906	0.8803	0.4700	0.8874	0.4554
0.044	0.8898	0.4502	0.8920	0.4466	0.8693	0.4883	0.8896	0.4510	0.8888	0.4529
0.056	0.8938	0.4431	0.8922	0.4461	0.8698	0.4872	0.8906	0.4496	0.8901	0.4503
0.068	0.8936	0.4431	0.8926	0.4446	–	–	0.8891	0.4517	0.8914	0.4478
0.080	0.8937	0.4427	0.8928	0.4443	–	–	0.8904	0.4495	0.8919	0.4470
0.092	0.8931	0.4435	0.8942	0.4421	–	–	0.8922	0.4462	0.8922	0.4464
0.104	0.8933	0.4426	0.8955	0.4398	–	–	0.8931	0.4444	0.8925	0.4458
0.116	0.8936	0.4421	0.8952	0.4400	–	–	0.8944	0.4418	0.8928	0.4452
0.128	0.8940	0.4414	–	–	–	–	0.8933	0.4434	0.8931	0.4446
0.140	0.8952	0.4390	–	–	–	–	0.8942	0.4414	0.8933	0.4442
0.152	0.8951	0.4391	–	–	–	–	0.8943	0.4417	0.8935	0.4438

high-quality compression.

In summary, this study further reveals the specific impacts of different encoding and bitrate control strategies on video quality assessment metrics, such as the Pearson correlation coefficient. It also emphasizes the importance of selecting an appropriate encoding scheme based on specific requirements in practical applications.

5. Discussion

In light of the previous experimental results, we observed that when the compression ratio increases and the PSNR values of various encoding methods exceed a certain threshold (approximately 50 dB), regardless of the variance in the video, the Pearson correlation coefficient of the BVP signal reaches its optimal state and stabilizes. Therefore, regardless of which compression method is chosen, it is

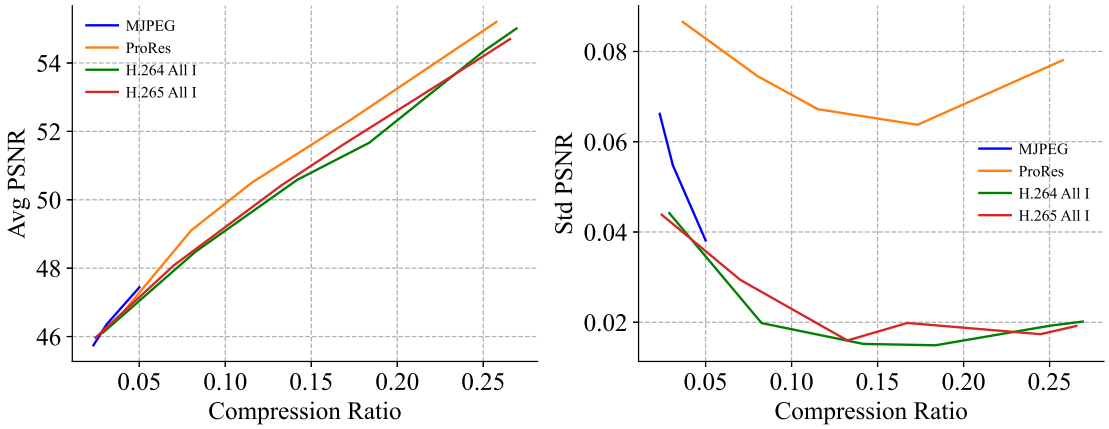

Fig. 12. Variations of Avg PSNR and Std PSNR with Compression Rate for P_s Videos, H.264 All I and H.265 All-I ($GOP = 1$) encoding (Results are Comparable to T1).

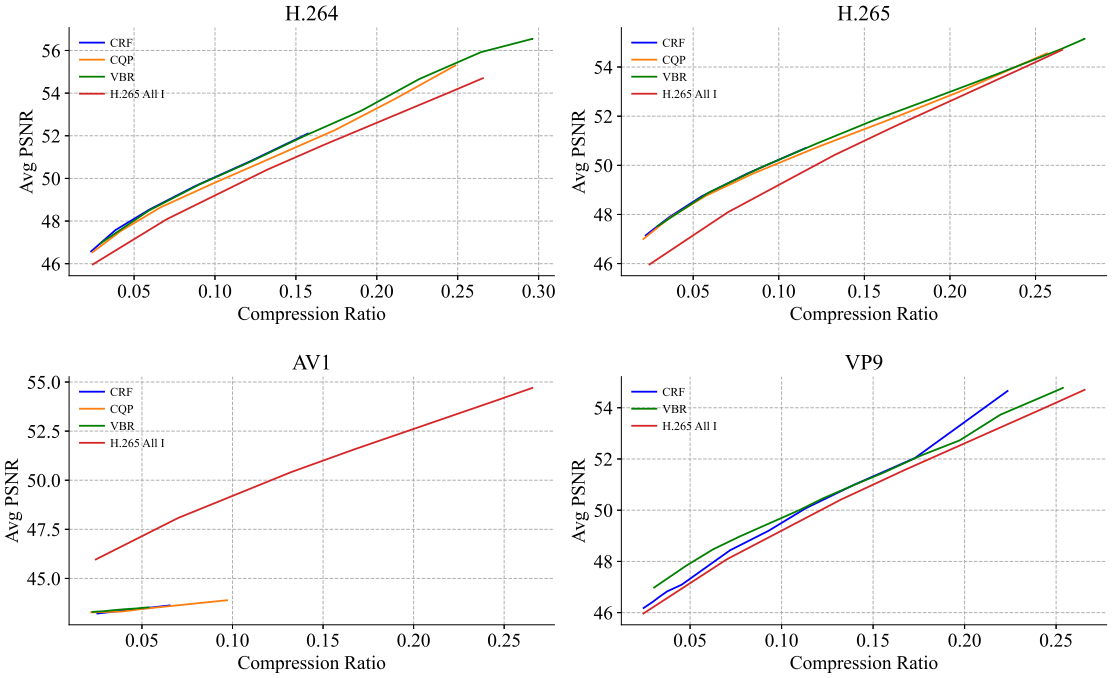

Fig. 13. The Avg PSNR for the T1 videos encoded with H.264, H.265, AV1, and VP9 codecs.

recommended to use whether the PSNR value exceeds this threshold as a criterion for evaluating the compression effect. When the compression ratio decreases and the PSNR value drops below 50 dB, different encoding methods, bitrate control strategies, motion amplitude, and GOP values exhibit varying impacts on the Pearson correlation coefficient of the BVP signal.

When comparing bitrate control schemes, the experiment selected the CQP mode of H.264 and H.265 (only the curves of the optimal bitrate control method are shown in Figs. 11 and 14), indicating that in complex inter-frame encoding scenarios, CQP can effectively control the generation and propagation of noise through consistent quantization parameters. In contrast, the VP9 encoding in FFmpeg does not offer corresponding CQP parameters, and using VBR yields better results than CRF.

For datasets with smaller motion amplitudes such as P_s and UPBC-Phys T1, the H.265 All I encoding ($GOP = 1$) shows the worst Pearson coefficient curve (red), ranking below other $GOP > 1$ bitrate control methods, while H.265 CQP (orange) performs best, with VP9 VBR (purple) performing similarly well. Therefore, static videos are more suitable for H.265 CQP or VP9 VBR encoding, with the GOP value set between 1 and 3, balancing compression efficiency and signal quality. Meanwhile, the PSNR value should be kept as high as possible to minimize the impact on the BVP signal.

In experiments with the dataset of higher motion amplitude P_w, although there is no significant improvement in the Pearson coefficient of H.265 All I encoding ($GOP = 1$), ts performance remains relatively stable with minimal fluctuations. This indicates that under scenes with intense motion, H.265 All I encoding can maintain a certain level of stability. Since the motion amplitude of UPBC-Phys T3 data is less than that of P_w, the difference between the T3 and T1 experimental results is not significant, but it can be seen that the performance of H.265 All I encoding ($GOP = 1$) improves on T3. Simultaneously, the VP9 encoding curve performs worse under $GOP > 1$ conditions, ranking below other encoding methods, suggesting that VP9's adaptability to motion is inferior to that of H.264 and H.265. Therefore, in scenarios with larger motion amplitudes and PSNR values below 50, H.265 encoding is a better choice.

Research involving scenes with intense motion is still challenging due to significant motion noise, leading to very low Pearson coefficients in the derived BVP signals, making them unsuitable for useful analysis.

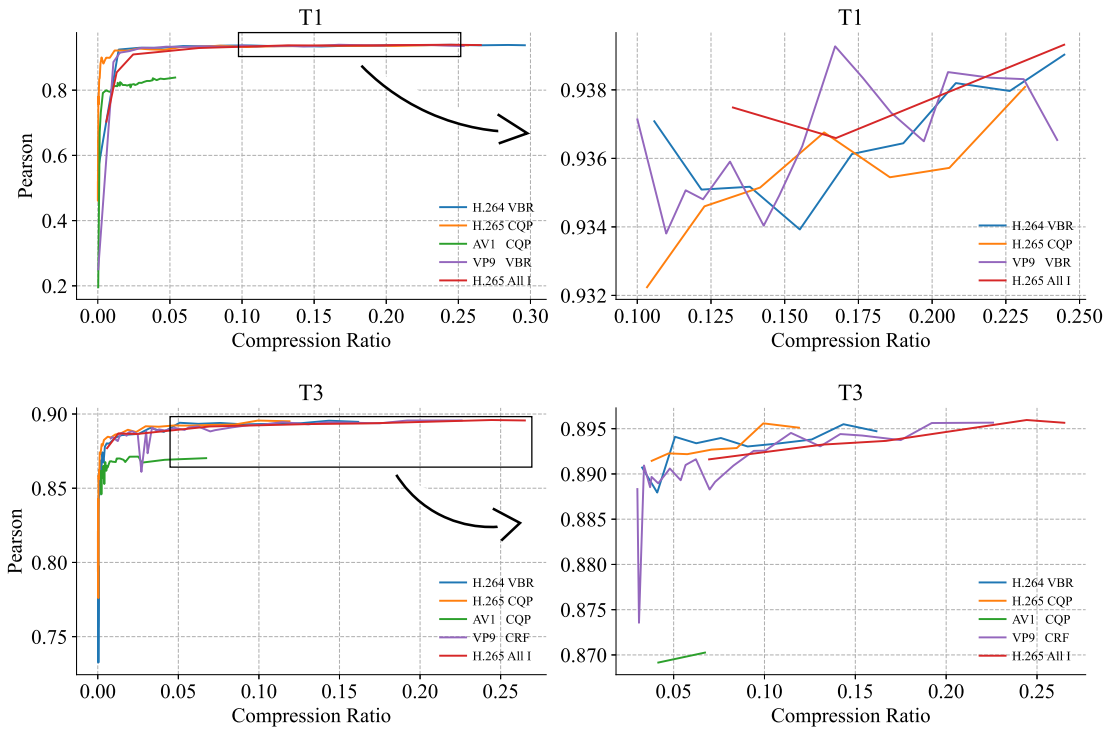

Fig. 14. Pearson correlation coefficients obtained from PhysNet and POS calculations on the compressed UBFC-Phys dataset (T1 and T3) for four encoding methods, varying with compression rate.

Lossless compression strategies are more appropriate. This experiment compared the performance of lossless compression encodings such as H.264, H.265, and FFV1, where FFV1 encoding maintains video quality while providing lower compression ratios, demonstrating superior performance in BVP signal extraction under intense motion scenarios and becoming the preferred choice for applications requiring the highest signal integrity.

6. Conclusion

A variety of factors influence the quality of BVP signal extraction, with video compression encoding technology being one of them. To explore the specific impact of different video compression encoding algorithms on the quality of BVP signal extraction, this paper constructs a dataset that includes raw video data. The dataset not only covers variations in different levels of motion intensity and lighting conditions but is also equipped with a standardized evaluation method. Subsequently, we conducted a systematic experimental analysis on a range of mainstream video compression encoding technologies (such as H.264, H.265, AV1/VP9, MJPEG, ProRes, and FFV1, etc.) using this dataset, focusing on the impact of factors such as noise introduced by inter-frame compression, different bit rate control strategies, and the dynamic characteristics of video content on the quality of BVP signal extraction. To further verify the validity of the research, we also conducted supplementary experiments on the publicly available UBFC-Phys dataset.

The experimental results reveal the specific mechanisms by which various video compression encoding technologies affect the quality of BVP signal extraction, providing valuable reference for researchers in related fields.

CRediT authorship contribution statement

Caiying Zhou: Writing – original draft, Methodology, Investigation, Conceptualization. **Xiaolang Ye:** Writing – original draft, Visualization, Validation, Software, Formal analysis. **Yuanwang Wei:** Writing – review & editing, Supervision, Project administration, Investigation, Funding acquisition, Data curation, Conceptualization. **Vincenzo De Florio:** Writing – review & editing. **Hong Sun:** Investigation. **Xinlong Zhan:** Writing – review & editing, Software, Methodology, Investigation. **Yonggang Li:** Writing – review & editing. **Chaochao Wang:** Writing – review & editing. **Xianchao Zhang:** Writing – review & editing.

Statements of ethical approval

All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards. Informed consent was obtained from all individual participants included in the study.

Code availability

The source codes for the investigations performed in this paper are available on Github: https://github.com/remote-PPG/rppg_video_compression.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work is partially supported by the grants from the Zhejiang Provincial Natural Science Foundation of China (Grant Nos. LTGG24F020001, LQ23F010006 and Z25F020008); and the Jiaxing Science and Technology Project (Grant No. 2023AY11030). We thank all the anonymous reviewers who generously contributed their time and efforts.

Data availability

Data will be made available on request.

References

[1] Hongsheng Yin, Honggang Qi, Jingwen Xu, William NN Hung, Xiaoyu Song, Generalized framework for similarity measure of time series, Math. Probl. Eng. 2014 (1) (2014) 572124.

[2] Weiyi Lv, Ning Zhang, Junjie Zhang, Dan Zeng, One-shot multiple object tracking with robust id preservation, IEEE Trans. Circuits Syst. Video Technol. (2023).

[3] Yuteng Xiao, Kaijian Xia, Hongsheng Yin, Yu-Dong Zhang, Zhenjiang Qian, Zhaoyang Liu, Yuehan Liang, Xiaodan Li, AFSTGCN: Prediction for multivariate time series using an adaptive fused spatial-temporal graph convolutional network, Digit. Commun. Netw. (2022).

[4] Weiyi Lv, Yuhang Huang, Ning Zhang, Ruei-Sung Lin, Mei Han, Dan Zeng, DiffMOT: A real-time diffusion-based multiple object tracker with non-linear prediction, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 19321–19330.

[5] Yuteng Xiao, Zhaoyang Liu, Hongsheng Yin, Xingang Wang, Yudong Zhang, STFormer: A dual-stage transformer model utilizing spatio-temporal graph embedding for multivariate time series forecasting, J. Intell. Fuzzy Systems (Preprint) (2024) 1–17.

[6] Hang Du, Hailin Shi, Dan Zeng, Xiao-Ping Zhang, Tao Mei, The elements of end-to-end deep face recognition: A survey of recent advances, ACM Comput. Surv. 54 (10s) (2022) 1–42.

[7] Ashish Kumar, Aryaman Ashdhir, Rama Komaragiri, Manjeet Kumar, et al., Analysis of photoplethysmogram signal to estimate heart rate during physical activity using fractional fourier transform–A sampling frequency independent and reference signal-less method, Comput. Methods Programs Biomed. 229 (2023) 107294.

[8] Jieying Wang, Caifeng Shan, Lin Liu, Zongshen Hou, Camera-based physiological measurement: Recent advances and future prospects, Neurocomputing (2024) 127282.

[9] Pireh Pirzada, Adriana Wilde, David Harris-Birtill, Remote photoplethysmography for heart rate and blood oxygenation measurement: a review, IEEE Sens. J. (2024).

[10] Ru Jing Lee, Saaveethya Sivakumar, King Hann Lim, Review on remote heart rate measurements using photoplethysmography, Multimedia Tools Appl. 83 (15) (2023) 44699–44728, http://dx.doi.org/10.1007/s11042-023-16794-9.

[11] Xiaowen Chen, Guanci Yang, Yang Li, Qingsheng Xie, Xiang Liu, Heart rate measurement based on spatiotemporal features of facial key points, Biomed. Signal Process. Control 96 (2024) 106650.

[12] Hemant Sharma, Amita Kumari Sharma, Var-HR: Non-contact heart rate measurement using an RGB camera based on adaptive region selection with singular value decomposition, IEEE Sens. Lett. (2024).

[13] Fangfang Zhu, Qichao Niu, Xiang Li, Qi Zhao, Honghong Su, Jianwei Shuai, FM-FCN: a neural network with filtering modules for accurate vital signs extraction, Res. 7 (2024) 0361.

[14] Xiaowen Chen, Guanci Yang, Yang Li, Qingsheng Xie, Xiang Liu, Heart rate measurement based on spatiotemporal features of facial key points, Biomed. Signal Process. Control. 96 (2024) 106650, http://dx.doi.org/10.1016/J.BSPC.2024.106650.

[15] Moajjem Hossain Chowdhury, Muhammad E.H. Chowdhury, Mamun Bin Ibne Reaz, Sawal Hamid Md. Ali, Seyed Mehdi Rakhtala, M. Murugappan, Sakib Mahmud, Md Nazmul Islam Shuzan, Ahmad Ashrif A. Bakar, Mohd Ibrahim Bin Shapiai, Muhammad Salman Khan, Amith Khandakar, LGI-rPPG-Net: A shallow encoder-decoder model for rPPG signal estimation from facial video streams, Biomed. Signal Process. Control. 89 (2024) 105687, http://dx.doi.org/10.1016/J.BSPC.2023.105687.

[16] Giuseppe Boccignone, Alessandro D'Amelio, Omar Ghezzi, Giuliano Grossi, Raffaella Lanzarotti, An evaluation of non-contact photoplethysmography-based methods for remote respiratory rate estimation, Sensors 23 (7) (2023) 3387.

[17] Huahua Chen, Xiang Zhang, Zongheng Guo, Na Ying, Meng Yang, Chunsheng Guo, ACTNet: Attention based CNN and transformer network for respiratory rate estimation, Biomed. Signal Process. Control 96 (2024) 106497, http://dx.doi.org/10.1016/j.bspc.2024.106497.

[18] Hanguang Xiao, Tianqi Liu, Yisha Sun, Yulin Li, Shiyi Zhao, Alberto Avolio, Remote photoplethysmography for heart rate measurement: A review, Biomed. Signal Process. Control 88 (2024) 105608, http://dx.doi.org/10.1016/j.bspc.2023.105608.

[19] Weiying Xing, Yinni Shi, Chaoyong Wu, Yiqiao Wang, Xu Wang, Predicting blood pressure from face videos using face diagnosis theory and deep neural networks technique, Comput. Biol. Med. 164 (2023) 107112.

[20] Osama A Omer, Mostafa Salah, Loay Hassan, Ahmed Abdelreheem, Ammar M Hassan, Video-based beat-by-beat blood pressure monitoring via transfer deep-learning, Appl. Intell. 54 (6) (2024) 4564–4584.

[21] Fabian Schrumpf, Patrick Frenzel, Christoph Aust, Georg Osterhoff, Mirco Fuchs, Assessment of non-invasive blood pressure prediction from PPG and rPPG Signals Using Deep Learning, Sensors 21 (18) (2021) 6022, http://dx.doi.org/10.3390/s21186022.

[22] Yongshen Zeng, Dongfang Yu, Xiaoyan Song, Qiqiong Wang, Liping Pan, Hongzhou Lu, Wenjin Wang, Camera-based cardiorespiratory monitoring of preterm infants in NICU, IEEE Trans. Instrum. Meas. (2024).

[23] Ismoil Odinaev, Kwan Long Wong, Jing Wei Chin, Raghav Goyal, Tsz Tai Chan, Richard HY So, Robust heart rate variability measurement from facial videos, Bioeng. 10 (7) (2023) 851.

[24] Daniel J. McDuff, Ethan B. Blackford, Justin R. Estepp, The impact of video compression on remote cardiac pulse measurement using imaging photoplethysmography, in: 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017), IEEE, Washington, DC, DC, USA, 2017, pp. 63–70, http://dx.doi.org/10.1109/FG.2017.17.

[25] Xuesong Niu, Hu Han, Shiguang Shan, Xilin Chen, VIPL-HR: A multi-modal database for pulse estimation from less-constrained face video, 2018, arXiv:1810.04927.

[26] Gerard De Haan, Vincent Jeanne, Robust pulse rate from chrominance-based rPPG, IEEE Trans. Biomed. Eng. 60 (10) (2013) 2878–2886, http://dx.doi.org/10.1109/TBME.2013.2266196.

[27] Changchen Zhao, Weihai Chen, Chun-Liang Lin, Xingming Wu, Physiological signal preserving video compression for remote photoplethysmography, IEEE Sens. J. 19 (12) (2019) 4537–4548, http://dx.doi.org/10.1109/JSEN.2019.2899102.

[28] Mohammad Soleymani, Jeroen Lichtenauer, Thierry Pun, Maja Pantic, A multimodal database for affect recognition and implicit tagging, IEEE Trans. Affect. Comput. 3 (1) (2011) 42–55.

[29] Sergey Tulyakov, Xavier Alameda-Pineda, Elisa Ricci, Lijun Yin, Jeffrey F. Cohn, Nicu Sebe, Self-adaptive matrix completion for heart rate estimation from face videos under realistic conditions, in: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, Las Vegas, NV, USA, 2016, pp. 2396–2404, http://dx.doi.org/10.1109/CVPR.2016.263.

[30] Radim Špetlík, Vojtech Franc, Jirí Matas, Visual heart rate estimation with convolutional neural network, in: Proceedings of the British Machine Vision Conference, Newcastle, UK, 2018, pp. 3–6.

[31] Xuesong Niu, Hu Han, Shiguang Shan, Xilin Chen, VIPL-HR: A multi-modal database for pulse estimation from less-constrained face video, in: Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part V 14, Springer, 2019, pp. 562–576.

[32] Ronny Stricker, Steffen Muller, Horst-Michael Gross, Non-contact video-based pulse rate measurement on a mobile service robot, in: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, IEEE, Edinburgh, UK, 2014, pp. 1056–1062, http://dx.doi.org/10.1109/ROMAN.2014.6926392.

[33] Guillaume Heusch, André Anjos, Sébastien Marcel, A Reproducible Study on Remote Heart Rate Measurement, 2017, arXiv:1709.00962.

[34] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, Julien Dubois, Unsupervised skin tissue segmentation for remote photoplethysmography, Pattern Recognit. Lett. 124 (2019) 82–90, http://dx.doi.org/10.1016/j.patrec.2017.10.017.

[35] Rita Meziati Sabour, Yannick Benezeth, Pierre De Oliveira, Julien Chappé, Fan Yang, UBFC-Phys: A multimodal database for psychophysiological studies of social stress, IEEE Trans. Affect. Comput. 14 (1) (2023) 622–636, http://dx.doi.org/10.1109/TAFFC.2021.3056960.

[36] Xiaobai Li, Iman Alikhani, Jingang Shi, Tapio Seppanen, Juhani Junttila, Kirsi Majamaa-Voltti, Mikko Tulppo, Guoying Zhao, The OBF Database: A large face video database for remote physiological signal measurement and atrial fibrillation detection, in: 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), IEEE, Xi'an, 2018, pp. 242–249, http://dx.doi.org/10.1109/FG.2018.00043.

[37] Pieter-Jan Toye, Vital videos: A dataset of face videos with PPG and blood pressure ground truths, 2023, http://dx.doi.org/10.48550/ARXIV.2306.11891, arXiv preprint arXiv:2306.11891.

[38] Daniel McDuff, Sarah Gontarek, Rosalind W. Picard, Improvements in remote cardiopulmonary measurement using a five band digital camera, IEEE Trans. Biomed. Eng. 61 (10) (2014) 2593–2601, http://dx.doi.org/10.1109/TBME.2014.2323695.

[39] Sang Bae Park, Gyehyun Kim, Hyun Jae Baek, Jong Hee Han, Joon Ho Kim, Remote pulse rate measurement from near-infrared videos, IEEE Signal Process. Lett. 25 (8) (2018) 1271–1275, http://dx.doi.org/10.1109/LSP.2018.2842639.

[40] Yue Wu, Tal Hassner, KangGeon Kim, Gerard Medioni, Prem Natarajan, Facial landmark detection with tweaked convolutional neural networks, IEEE Trans. Pattern Anal. Mach. Intell. 40 (12) (2018) 3067–3074, http://dx.doi.org/10.1109/TPAMI.2017.2787130.

[41] T.J. Harris, Hui Yuan, Filtering and frequency interpretations of singular spectrum analysis, Physica D 239 (20–22) (2010) 1958–1967, http://dx.doi.org/10.1016/j.physd.2010.07.005.

[42] Jialiang Gu, Kevin Hung, Bingo Wing-Kuen Ling, Daniel Hung-Kay Chow, Yang Zhou, Yaru Fu, Sio Hang Pun, Generalized singular spectrum analysis for the decomposition and analysis of non-stationary signals, J. Franklin Inst. 361 (6) (2024) 106696, http://dx.doi.org/10.1016/j.jfranklin.2024.106696, URL https://www.sciencedirect.com/science/article/pii/S0016003224001170.

[43] Christopher J. James, David Lowe, Extracting multisource brain activity from a single electromagnetic channel, Artif. Intell. Med. 28 (1) (2003) 89–104, http://dx.doi.org/10.1016/S0933-3657(03)00037-X.

[44] Quanbo Lu, Mei Li, A method combining fractal analysis and single channel ICA for vibration noise reduction, Shock Vib. 2021 (1) (2021) 5583587.

[45] Norden E. Huang, Zheng Shen, Steven R. Long, Manli C. Wu, Hsing H. Shih, Quanan Zheng, Nai-Chyuan Yen, Chi Chao Tung, Henry H. Liu, The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis, Proc. R. Soc. Lond. Ser. A Math. Phys. Eng. Sci. 454 (1971) (1998) 903–995, http://dx.doi.org/10.1098/rspa.1998.0193.

[46] Haroon Yousuf Mir, Omkar Singh, VMD based powerline interference cancellation in ECG signals, in: 2023 Twelfth International Conference on Image Processing Theory, Tools and Applications, IPTA, IEEE, 2023, pp. 1–4.

[47] Zitong Yu, Wei Peng, Xiaobai Li, Xiaopeng Hong, Guoying Zhao, Remote heart rate measurement from highly compressed facial videos: An end-to-end deep learning solution with video enhancement, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, ICCV, 2019.

[48] Bolin Chen, Jie Chen, Shiqi Wang, Yan Ye, Generative face video coding techniques and standardization efforts: A review, in: 2024 Data Compression Conference, DCC, IEEE, 2024, pp. 103–112.

[49] Milan Jayaratne, L.K. Gunawardhana, Uthpala Samarathunga, Comparison of H. 264 and H. 265, Eng. Technol. Q. Rev. 5 (2) (2022).

[50] Pinar Akyazi, Touradj Ebrahimi, Comparison of compression efficiency between HEVC/H.265, VP9 and AV1 based on subjective quality assessments, in: 2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX), IEEE, Cagliari, 2018, pp. 1–6, http://dx.doi.org/10.1109/QoMEX.2018.8463294.

[51] Miroslav Uhrina, Lukas Sevcik, Juraj Bienik, Lenka Smatanova, Performance comparison of VVC, AV1, HEVC, and AVC for high resolutions, Electronics 13 (5) (2024) 953.