SiVe: See Into the Vehicle's Hidden Persons via Laser Doppler Vibrometer

Zhu Xiao, *Senior Member, IEEE*, Shirong Guan, Jingyang Hu, Siyu Chen, Hongbo Jiang, *Senior Member, IEEE*, and Keqin Li, *Fellow, IEEE*

***Abstract*—Detecting stowaways hidden in various transport vehicles, including cars, trucks, containers, and trailers, is crucial to border security inspection systems. Existing solutions mainly rely on contact-based sensors and manual inspection, which significantly compromise the efficiency of border control operations. Therefore, there is an urgent need for an automated, efficient and fast non-contact vehicle hidden person detection system. In this paper, we propose SiVe, a novel border inspection system that utilizes laser Doppler vibrometer (LDV) to detect hidden people in the vehicle. We extract signals associated with human activities (such as breathing, heartbeat, low-frequency body movements, etc.) from complex laser reflection data to detect the presence of hidden people. Specifically, we first employs the Empirical Mode Decomposition (EMD) algorithm to extract and reconstruct signals associated with human activities in complex and noisy environments. Then based on the characteristics of EMD outputs, we design a Time-series Variation Feature extraction Identification network (TVFI-net) model that accurately captures complex time-varying patterns for efficient and reliable detection of hidden people presence. Extensive real-world experiments validate the effectiveness of SiVe in various environments. The system achieves an average presence detection accuracy of 99.93% for sedans and MPVs, 98.37% for light trucks, 98.07% for heavy trucks, and 95.23% for trailers in non-contact detection of hidden people across twelve different vehicle types, under various indoor and outdoor environments.**

***Index Terms*—Laser Doppler vibrometer, hidden person detection, deep learning.**

I. Introduction

In recent years, the robust growth of the cross-border transportation and logistics sectors has simultaneously created opportunities for illicit activities. Vehicle compartments (sedans, SUVs, vans, etc.) due to their high concealment capabilities and significant load capacity, are particularly susceptible to being used to hide illegal immigrants, victims of human trafficking, contraband, and even members of terrorist organizations. This illicit smuggling poses serious threats to social security, public health, and social stability. Since 2021, more than 3.3 million people have attempted illegal border crossings in the United States, leading to a total expenditure by the federal government of 72.8 billion dollars and state and local governments spending over 115.6 billion dollars on issues related to illegal immigration [1].

Existing solutions for detecting people in enclosed spaces include X-ray scanning [2], [3] and vibration sensors [4], [5]. While these methods have achieved good results in certain scenarios, they also have significant limitations. For example, X-ray scanning lacks penetration when encountering thick vehicle bodies. Vibration sensors require direct contact with the target object. This not only increases operational complexity but also poses safety risks to inspectors, especially when dangerous individuals may be hiding in the vehicle. Furthermore, manual inspection is time-consuming, labor-intensive, and inefficient, making them unsuitable for large-scale rapid security checks and automated customs clearance processes.

To tackle these challenges, in this paper, we propose SiVe, which centered around the use of Laser Doppler Vibrometers (LDV). SiVe aims to achieve precise and efficient detection of individuals within enclosed spaces through the high sensitivity and non-contact nature of LDV technology. The extremely short wavelength of lasers enables sub-micron to nanometer-level high spatial resolution [6], [7], ensuring highly accurate non-contact measurement of minute vibrations. The high directionality and focusing ability of lasers allow for precise targeting and measurement of vibrations in specific areas [8], [9]. Consequently, LDV can measure subtle vibrations on surfaces at a distance, capturing even the most minute biological signs, such as those induced by breathing or heartbeat [10], [11], [12], to ascertain the presence of living beings within confined spaces.

Received 12 August 2025; revised 9 January 2026; accepted 10 February 2026. Date of publication 13 February 2026; date of current version 5 June 2026. This work was supported in part by the National Key Research and Development Program of China under Grant 2023YFC3321601, in part by the National Natural Science Foundation of China under Grant U24A20247, in part by the Postdoctoral Fellowship Program of CPSF under Grant GZC20251058, in part by the China Postdoctoral Science Foundation through Anhui Joint Support Program under Grant 2025T022AH, in part by the National Science Foundation of China (NSFC) under Grant 62372161, in part by the Selected Best Candidates to Undertake Key Research Projects of Changsha City under Grant kq2503003, in part by the Yuelushan Center for Industrial Innovation under Grant 2025YCII0127, in part by the Shenzhen S&T Program under Grant JCYJ20250604185802004, in part by the Science and Technology Innovation Program of Hunan Province under Grant 2025PT3026, in part by by the Science and Technology Program of Hunan Provincial Department of Transportation under Grant 202507, and the Hunan Natural Science Foundation of China under Grant 2025JJ50351. Recommended for acceptance by X. Tian. *(Corresponding authors: Jingyang Hu; Hongbo Jiang.)*

Zhu Xiao, Shirong Guan, Siyu Chen, and Hongbo Jiang are with the College of Computer Science and Electronic Engineering, Hunan University, Changsha 410082, China, and also with Shenzhen Research Institute, Hunan University, Shenzhen 518055, China (e-mail: zhxiao@hnu.edu.cn; srguan@hnu.edu.cn; csy990406@hnu.edu.cn; hongbojiang2004@gmail.com).

Jingyang Hu is with the School of Computer Science and Technology, University of Science and Technology of China, Hefei 230000, China (e-mail: hujingyang@ustc.edu.cn).

Keqin Li is with the Department of Computer Science, State University of New York, New York, WA 98104 USA (e-mail: lik@newpaltz.edu).

Digital Object Identifier 10.1109/TMC.2026.3664371

SiVe operates based on the laser Doppler effect: when a laser beam strikes a vibrating surface, the reflected light experiences a frequency shift corresponding to the vibration frequency. By employing signal processing algorithms with deep learning, SiVe deciphers the information embedded in these vibration signals, providing strong evidence for the detection of illicitly concealed individuals within confined spaces. Compared to contact-based detection methods, SiVe not only eliminates the safety risks associated with direct contact but also enhances detection efficiency and accuracy.

Despite its promising potential, the SiVe solution utilizing LDV faces two core challenges in transitioning into a practical detection system:

- Signal Extraction and Noise Suppression: In real-world scenarios, the environments surrounding enclosed trucks or containers are typically rife with various vibration sources, ranging from ground traffic vibrations to wind effects, and even the operation sounds of nearby machinery. These factors introduce substantial background noise that can interfere with the signals captured by the LDV. Hence, precisely extracting weak vibration signals associated with human activity from a complex, noisy environment becomes a critical hurdle for the SiVe technology. Effective signal processing algorithms must be capable of identifying and filtering out irrelevant noise, ensuring that even under adverse conditions, subtle vibrations indicative of human vital signs can be accurately detected.
- Efficient and Accurate Human Detection: Even after successfully isolating vibration signals related to human activity, the SiVe must possess robust signal analysis capabilities to efficiently and accurately distinguish whether there are individuals concealed within the confined space. This entails a deep understanding and modeling of signal characteristics, including discerning human vital signs from other types of vibration patterns, such as cargo movement or vehicular vibrations. Given the diversity of vehicle interior structures and the uncertainty of hiding locations, developing a versatile and adaptable signal analysis model that performs consistently across different scenarios and conditions represents another significant challenge to achieving both the practicality and reliability of the SiVe.

To realize the detection of hidden people in the vehicle, we first employ Empirical Mode Decomposition (EMD) to preprocess and isolate faint signals associated with human motion. Subsequently, we design a neural network model, which, through training, learns to discriminate between human movement and noise in the processed signals, thereby enabling accurate determination of an individual's presence within an enclosed space. Extensive experiments demonstrate that our system achieves efficient and precise hidden-person detection across various vehicle types.

Overall, we summarize the contributions of SiVe as follows:

- We propose a LDV-based system for non-contact, high-precision detection of persons in enclosed spaces, addressing the limitations of conventional detection methods.
- We design signal processing algorithms that enable SiVe to extract faint vibration signals associated with human activity from substantial background noise in complex real-world environments, ensuring the accuracy and reliability of the detection process.
- Through the construction of a deep neural network model, SiVe is capable of efficiently and accurately distinguishing whether individuals are concealed within enclosed spaces, maintaining high detection accuracy even amidst varied vehicular interior structures and uncertain hiding positions.
- SiVe exhibits versatility across diverse scenarios and demonstrates high-precision detection capabilities in intricate environments. This substantiates its practicality and dependability as a solution for detecting individuals in enclosed spaces, offering robust technological support for enhancing societal security measures.

II. Background and Motivation

A. Laser Doppler Basics

The LDV operates based on the Doppler effect,which describes the frequency shift of light reflected from a moving surface. When a laser beam strikes a vibrating surface, the reflected light experiences a frequency shift proportional to the velocity of the surface motion. This principle is mathematically expressed as:

$$f_D = \frac{-2v(t)}{\lambda_n}, \tag{1}$$

where f_D is the Doppler frequency shift, $v(t)$ is the surface velocity, and λ_n is the wavelength of the laser adjusted for the refractive index of the medium (e.g., air).

B. Interferometric Measurement of the Light Phase

1) Detection of Light: For a given detector bandwidth B, the photodiode is capable of converting the light energy (measured in terms of optical power P) irradiated on it into a corresponding electrical signal, i.e., the photogenerated current $i_P = K_P \cdot Z_0 \cdot < E^2(t) >$.

In photodiodes, sensitivity K_P, reflecting optical-to-electrical power conversion, is 0.5 A/mW in the visible, rising to 1 A/mW at 1550 nm, showing wavelength dependence. Free space impedance Z_0 is approximately 376,73 Ω is based on constants μ_0 (magnetic permeability) and ϵ_0 (electric permittivity).

2) Interferometric Detection: For the accurate detection of swift or infinitesimal changes in electric fields, the application of 'beat frequency' techniques, which involve the superposition of measurement beams(E_m) and reference beams(E_r) to generate a lower frequency beat signal, becomes indispensable. Following this, the signal is processed by a photodiode and a low-pass filter to ensure its frequency is suitable for measurement by a high-speed photodetector. The interferometer physically superimposes these two light beams, facilitating their spatial interference.

$$\widetilde{E}(t) = \sqrt{r} \cdot \widetilde{E}_{m,0} \cos(\omega t + \varphi(t)) + \widetilde{E}_{r,0} \cdot \cos(\omega t). \tag{2}$$

In the experiment, the beam separator superimposes two beams and the reference light intensity is halved. The interference is converted to an electrical signal, low-pass filtered to

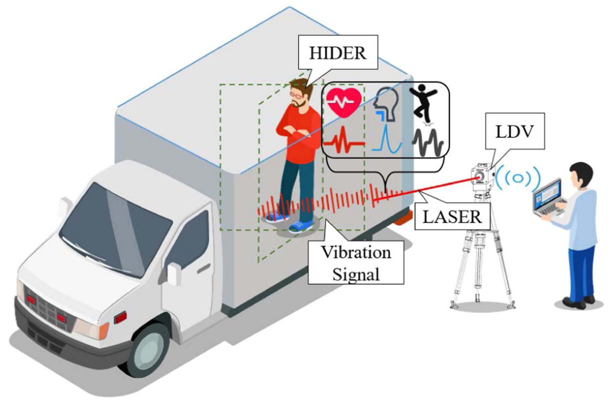


Fig. 1. Usage scenarios of SiVe, non-contact detection of hidden persons in vehicles using laser Doppler vibrometer.

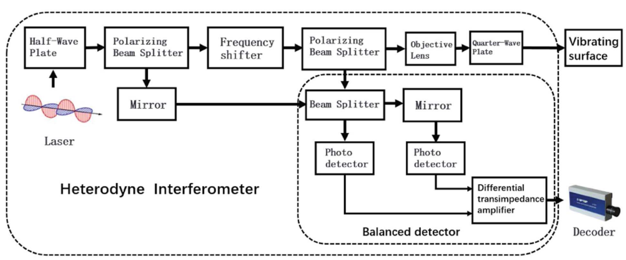


Fig. 2. The basics of LDV.

remove noise, and the electric field strength squared is calculated to extract the interference key data.

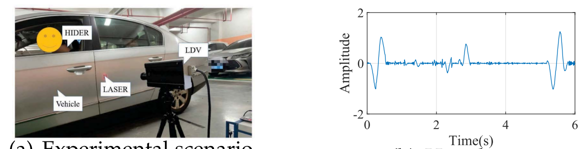
$$< \widetilde{E}^2(t) >= r\cdot \widetilde{E}^2_{m,0} + 2\cdot\sqrt{r}\cdot\eta\cdot\widetilde{E}_{m,0}\cdot\widetilde{E}_{r,0} \cdot\cos(\varphi(t)) + \widetilde{E}^2_{r,0}. \tag{3}$$

η is a correction factor used to characterize and quantify the degradation of signal quality caused by wavefront aberrations.

3) Heterodyne Interferometers: The heterodyne method minimizes the technical noise in the detector and suppresses harmonic distortions in the detector by means of suitable filters, as well as enhancing the signal-to-noise ratio of the signal. Fig. 2 show an example of such an interferometer setup based on the asynchronous frequency mixing principle.

In this setup, frequency mixing is achieved by changing the frequency of the measured light relative to the reference light using a carrier frequency ω_c. The resulting photocurrent is converted into a voltage signal u_{TIA}. Using a transimpedance amplifier (TIA). The detector bandwidth B is typically set to twice the carrier frequency ω_c. To eliminate DC bias and ensure response to only dynamic signals, the system often employs AC coupling or bias removal techniques. Balanced detection is one of the most effective approaches for this purpose, using two nearly identical photodetectors receiving optical signals that are 90 degrees out of phase.

Therefore, with the balanced detection technique, the final output signal of the detector is essentially twice the product of the electric field amplitude of the measured light (E_m) and the



(a) Experimental scenario. (b) Heartbeat.

Fig. 3. Preliminary experiment.

reference light (E_r), i.e., $2\cdot\tilde{E}_m\cdot\tilde{E}_r = E_m\cdot E_r E_m$.

$$u_{TIA,B}(t) = 2\cdot R_{TIA}\cdot K_P\cdot\sqrt{r}\cdot\eta\cdot\sqrt{P_m}\cdot\sqrt{P_r} \cdot\cos\left(\omega_c t + M\cdot\sin(\Omega t)\right), \tag{4}$$

P_r is the reference beam power and P_m is the measured beam power.

4) Proof-of-Concept: A rudimentary experiment was devised and executed to substantiate the practicability of employing LDV for the detection of obscured occupants within an automobile. As depicted in Fig. 3(a), the experimental protocol required the participant to be placed inside a motionless vehicle. The LDV was utilized to irradiate the car's exterior, enabling continuous surveillance to discern infinitesimal oscillations, which were subsequently attributed to heartbeat activity.

Our experiments validate our hypothesis: regardless of an individual's position within the vehicle cabin, the data collected by LDV can be processed using our extraction algorithm to reveal signals like the one shown in Fig. 3(b), which aligns with human Heartbeat patterns (as confirmed by the baseline device worn by the subject). In contrast, such signals are typically absent in tests conducted with an empty vehicle. This finding highlights the potential of LDV as a reliable non-contact tool for health monitoring and detection, especially in scenarios where individuals may be concealed in hidden locations.

III. System Design

A. System Overview

Our SiVe system is a cutting-edge personnel presence detection system that employs LDV technology, dedicated to achieving high-precision detection of individuals concealed within enclosed spaces under non-contact conditions.

Fundamental Principle: SiVe is based on the principle of structural vibration transmission rather than laser penetration. When occupants inside a vehicle breathe, move, or have cardiac activity, their micro-movements exert dynamic forces on the seat and cabin structure. These forces propagate through the vehicle's mechanical system (contact surface → chassis → body shell), inducing minute vibrations on the external surface (e.g., door, hood). The LDV detects these surface vibrations in a non-contact manner by measuring the Doppler shift of the reflected laser beam. Although the displacement amplitude is extremely small (sub-micron to micron level), modern LDVs offer sufficient sensitivity to capture such signals, especially when combined with advanced signal processing techniques. This physical mechanism enables non-invasive detection of

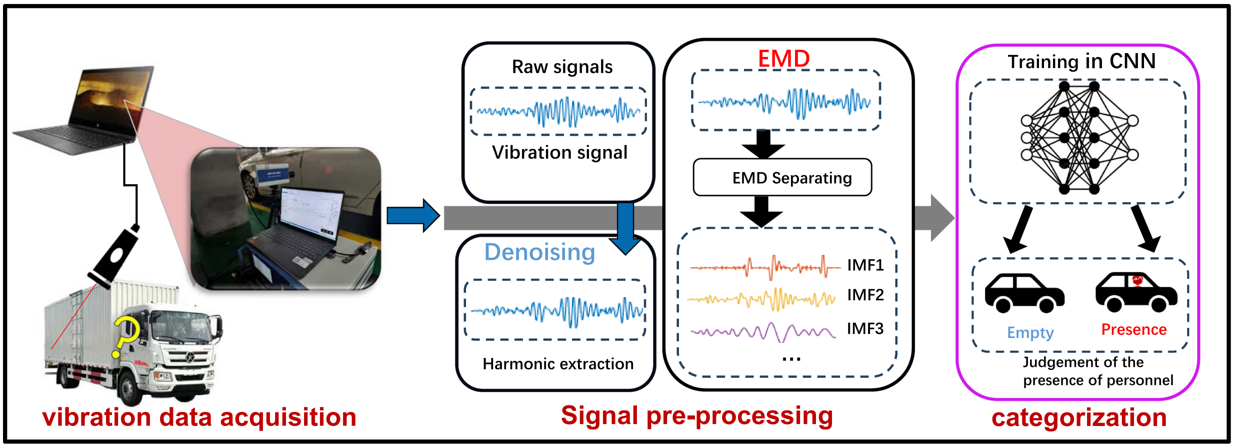


Fig. 4. System overview (1) Vibration Data Acquisition using LDV, (2) Signal Pre-processing via low-pass filtering and EMD decomposition, and (3) Categorization through a TVFI-net model. The system leverages LDV's high-resolution vibration signals and EMD-based feature extraction to enable non-contact detection of hidden individuals in vehicles.

hidden individuals without requiring any modification to the vehicle.

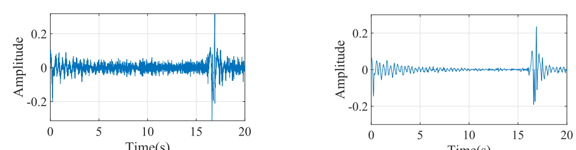
Similar structural-vibration-based human sensing approaches have been demonstrated in infrastructure monitoring and smart building applications [4], [5].

As depicted in Fig. 4, the system comprises three main phases: (1) Vibration Data Acquisition; (2) Signal Pre-processing; (3) Categorization. During the Vibration Data Acquisition phase, the SiVe system utilizes LDV technology to emit laser beams onto the vehicle under inspection, capturing raw vibration signals as the system input. These raw signals are contaminated with substantial noise, hence in the Signal Pre-processing phase, the SiVe system initially applies a low-pass filter to eliminate background noises that have significantly different frequencies from human biometric signals. Subsequently, the EMD algorithm is used to decompose the preliminary denoised signals into multiple IMF components, extracting vibration information closely associated with human biometric features.

In the Categorization module, the SiVe system takes the IMF components derived from the previous step as input, utilizing their statistical features to train a neural network model. Ultimately, this module is capable of making precise judgments on the initial input signals, categorizing them as "No anomaly", "Single person present", "Two people present", and "Multiple people present", thereby realizing accurate detection of individuals hidden within confined spaces. This series of processes ensures that the SiVe system maintains high-performance detection capabilities even in complex environments.

B. LDV Signal Preprocessing

The raw vibration signals (collected from a Peugeot 2002, one measurement point, laser distance of 80 cm, no special surface treatment (natural reflection), outdoor environment) captured are contaminated with a significant amount of noise, as illustrated in Fig. 5(a). The presence of this noise greatly complicates the subsequent extraction of vital signs signals. To improve signal quality and enhance detection accuracy, considering that



(a) Raw Signal. (b) Filtered Signal.

Fig. 5. Enhanced signal quality through digital filtering techniques.

human-related signals are primarily low-frequency signals, we applied a 10 Hz Butterworth low-pass filter to process the signals. The filter order was set at 4th order, providing sufficient attenuation capability while avoiding excessive computational complexity, ensuring processing efficiency. Fig. 5(b) presents the signals after filtering, and we can clearly observe that, compared to the original signals, the noise components have been significantly suppressed.

After noise reduction, we apply the EMD to decompose the preliminarily denoised signal. EMD proposed by Dr. Norden E. Huang [13] from the NASA Goddard Space Flight Center, is an advanced time-frequency signal analysis method. This method is particularly well-suited for the analysis of nonlinear and non-stationary signals, while also being effective for linear and stationary signals. Compared to other time-frequency analysis techniques, EMD-based methods not only capture the intrinsic variation characteristics of signals more accurately, but also demonstrate superior performance in analyzing linear and stationary signals, providing a more precise representation of the physical meaning underlying the data. Therefore, EMD offers a unique and powerful tool for investigating the time-frequency characteristics of complex signals, regardless of whether they exhibit linearity or stationarity. The distinct advantage of this approach lies in its adaptability, making it an ideal choice for the analysis of a wide variety of signal types.

Specifically, EMD has been widely applied in biomedical signal processing, where it has proven effective in extracting meaningful physiological components from noisy or mixed signals. Representative applications include: Electrocardiogram

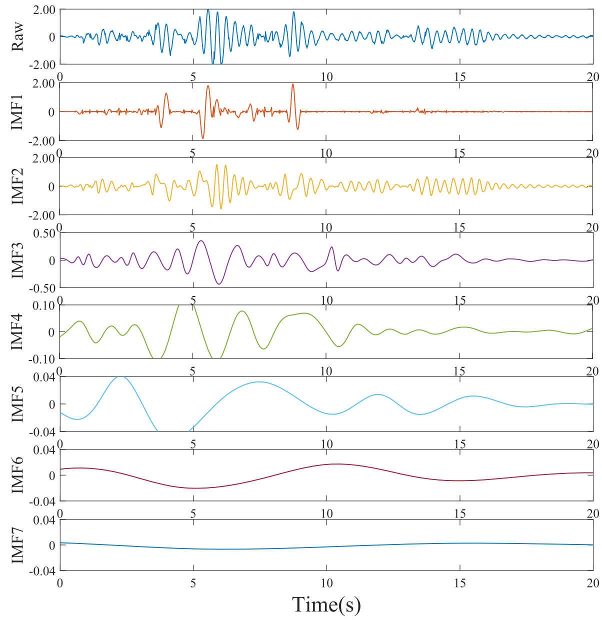


Fig. 6. EMD decomposition yields seven modes from the original signal (Collected from a Peugeot 2002 under outdoor conditions, using a single measurement point at a laser distance of 80 cm, with natural surface reflection and no special surface treatment).

(ECG) denoising, using EMD to separate cardiac components from muscle artifact noise [14]; EEG rhythm extraction, decomposing electroencephalographic signals to isolate α, β, and other neural oscillations [15]; Respiration extraction from non-respiratory signals, leveraging EMD to recover respiratory information embedded in ECG, blood pressure (BP), and photoplethysmographic (PPG) signals, as demonstrated on the MIMIC database [16]. These studies highlight EMD's capability to disentangle physiologically meaningful oscillations from complex, multi-component signals—particularly in low signal-to-noise ratio (SNR) environments. Inspired by these successes, we employ the EMD as a foundational step to decompose vehicle's micro-vibration signal into intrinsic mode functions. Each function can be used to capture distinct physical dynamics, including those induced by human respiration and heartbeat.

Note that EMD is a data-driven method that adaptively decomposes a signal $x(t)$ into a set of Intrinsic Mode Functions (IMFs), each representing different time-scale characteristics inherent in the original signal. This decomposition is performed iteratively through a process known as sifting.

Two key criteria must be satisfied for a component to qualify as an IMF:

- *R1:* The number of local maxima and local minima in the IMF should differ by at most one, and the number of zero crossings should be close to the number of extrema.
- *R2:* At any given time instant, the mean of the upper envelope (defined by local maxima) and the lower envelope (defined by local minima) should be approximately zero.

These rules ensure that each IMF reflects meaningful oscillatory modes within the signal. Specifically, R1 prevents excessively long monotonic intervals, preserving local fluctuations,

Algorithm 1: Empirical Mode Decomposition (EMD).

Require: Signal $x(t)$, Threshold ε
Ensure: Intrinsic Mode Functions (IMFs) IMF_j

```
Initialize residual signal r_0(t) ← x(t)
Initialize IMF counter j ← 1
while number of extrema in r_j(t) ≥ 2 do
  Initialize h_{j,0}(t) ← r_{j-1}(t)
  Initialize iterations i ← 1
  while SD(i) ≥ ε do
    Identify all local maxima and minima
    of h_{j,i-1}(t)
    step for R1: Ensure extrema count
    difference ≤ 1
    Compute the upper and lower envelopes
    U_{j,i-1}(t), L_{j,i-1}(t) by interpolating
    step for R2: Symmetry via envelope mean
    Calculate the mean of the envelope
    μ_{j,i-1}(t) = (U_{j,i-1}(t) + L_{j,i-1}(t)) / 2
    R2: Zero-mean envelope
    Update h_{j,i}(t) ← h_{j,i-1}(t) − μ_{j,i-1}(t), i ← i + 1
    Sifting process to enforce R2
    Calculate standard deviation SD(i) as stopping
    criterion.
  end while
  Set IMF_j(t) ← h_{j,i}(t)
  Update residual signal r_j(t) ← r_{j-1}(t) − IMF_j(t)
  Increment IMF counter j ← j + 1
end while
```

while R2 ensures symmetry and eliminates any global trend in the IMF.

Unlike traditional wavelet or filter-based decomposition methods, EMD does not rely on predefined basis functions or filters. Instead, it dynamically constructs IMFs based on the intrinsic characteristics of the signal itself, making it highly adaptive. It is important to note that for EMD to function properly, the original signal $x(t)$ must contain at least one local maximum and one local minimum. This ensures sufficient information for initiating the sifting process.

After EMD the initial signal x(t) is decomposed into $\mathrm{IMF}_j(t)$, $j = 1, \ldots$, C, and the residual $r_C(t) = x(t) - \sum_{j=1}^{C} \mathrm{IMF}_j(t)$, C represents the number of modes, which is determined based on the stopping criterion SD, so that C varies according to the characteristics of the signal itself. The screening process helps to eliminate excess fluctuations and smooth amplitudes. By limiting the size of the standard deviation SD of two consecutive screening results (typically the SD threshold is set between 0.2 and 0.3) it is ensured that the extracted IMF components retain the full physical significance of the signal in terms of amplitude and frequency modulation. Fig. 6 shows that our EMD method decomposes the raw signal into seven modes, including heartbeat (IMF1), ground vibration (IMF2) and respiration (IMF3), as well as other residual noise and trend component (IMF4-IMF7).

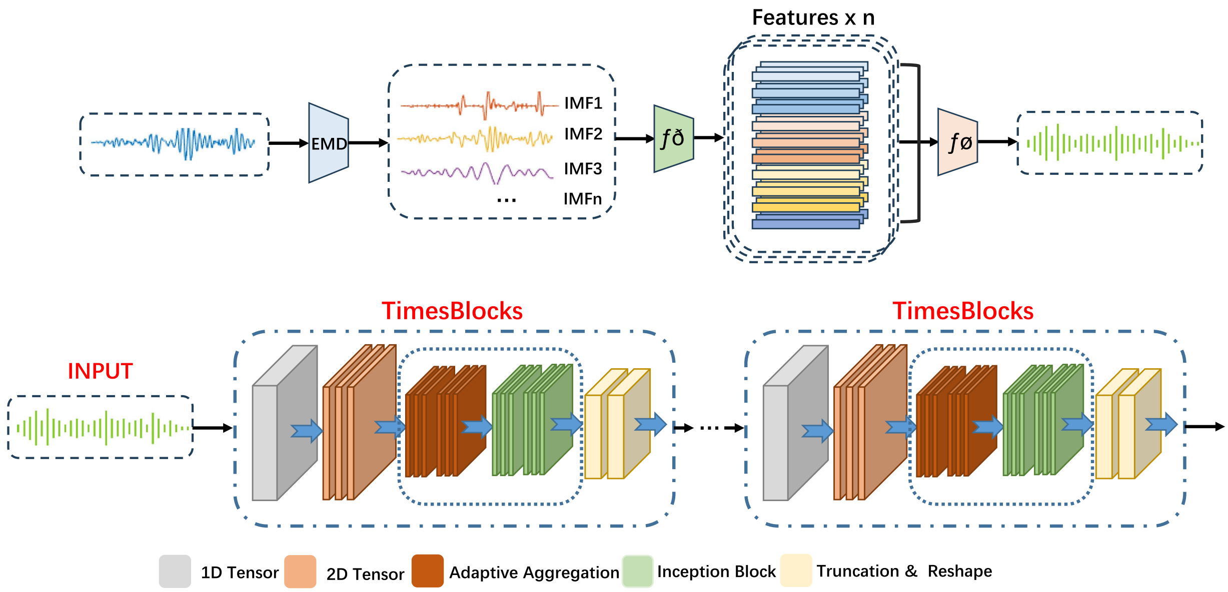


Fig. 7. Schematic network structure of TVFI-net: transforms 1D time series into 2D tensors to capture intra-period and inter-period temporal variations.

(a) Experimental scenario. (b) Bigrun Team x-26.

Fig. 8. Equipment and experimental scenario implementation.

C. Deep Learning Networks

1) Feature Extraction From Inputs: During the classification phase, the proposed neural network architecture aims to categorize the input data into four distinct classes: "No anomaly", "Single person present", "Two people present", and "Multiple people present". The input consists of the IMFs obtained from the previous decomposition step. Here, we only consider the first seven effective components. To effectively identify these data patterns, we need to extract a set of features from these IMFs components. The features we aim to extract include:

1) Means and standard deviations of each IMFs component $\mu = \frac{1}{|x|}\sum_{i=1}^{|x|} x_i$ and $\delta = \sqrt{\frac{\Sigma_{i=1}^{|x|}(x_i-\mu)^2}{|x|}}$;
2) Maximum and minimum values of IMFs;
3) The skewness (Skew$(x) = \frac{E(x-\mu)^3}{\delta^3}$;)
4) The kurtosis ($Kurtosis(x) = \frac{E(x-\mu)^4}{\delta^4}$;)
5) root sum square $\theta = \sqrt{\sum_{i=1}^{|x|} \chi^2_{\omega}}$;
6) The q-quantiles (q=0.25, 0.5, 0.75).

2) EMD Decomposition and Feature Set Construction: As show in Fig. 7, we have processed the acquired vibration data by decomposing it into seven components using the EMD algorithm, from which nine statistical features are extracted from each component, forming a feature set of size 7×9. This feature set is subsequently concatenated into a new one-dimensional time series to serve as input for the TVFI-net. Compared with directly training on raw data, this approach offers significant advantages. First, EMD can effectively separate different frequency components and reduce noise influence, allowing the model to focus on useful information in the signal. Second, this method enhances feature representation and reveals patterns hidden within complex vibration signals, thereby improving classification performance. Additionally, converting high-dimensional time series into fixed-length feature vectors simplifies the problem, reduces computational complexity, accelerates training speed, and enhances the generalization ability of the model. Since EMD is an adaptive decomposition method, it can automatically adjust its scale according to the characteristics of the signal, making it especially suitable for non-stationary and nonlinear vibration signals, thus better capturing their intrinsic changes. Meanwhile, the different IMFs obtained from EMD decomposition reflect signal features at multiple scales, aiding the model in understanding comprehensive temporal dependencies and improving prediction accuracy.

3) Advantages of the TVFI-Net Model: TVFI-net is an innovative time series analysis model specifically designed to capture and model complex multi-periodic variations. Since a single time point in vibration analysis often fails to provide sufficient semantics for analysis, much of the work focuses on temporal changes, which are more informative and can reflect the inherent characteristics of the time series. However, in practical applications, the time series involved in vibration analysis entail complex temporal patterns, primarily including the following aspects:

Multi-periodicity: Within each specific period represented by an IMF, there may exist subtle changes or fluctuations. For instance, in vibration signals caused by human heartbeat or respiration, even signals within the same period might exhibit minor time-dependent variations. There may be interactions or correlations between different IMFs, manifesting as changes between adjacent periods. For example, heartbeat and respiration are two independent periodic activities that can influence each other under certain circumstances, leading to coupling or synchronization phenomena between cycles.

Nonlinearity and Non-stationarity: In practice, vibration signals are often not simple sinusoidal waveforms but contain complex nonlinear dynamics. For example, human vital sign signals (such as heart rate variability) typically display nonlinear characteristics, making it challenging to extract these features directly from a one-dimensional time series. The statistical properties of vibration signals change over time, meaning that the signal is not constant. For instance, a concealed individual's heart rate might vary due to factors such as anxiety or ambient temperature, which adds complexity to signal analysis due to this non-stationarity.

Local and Global Features: Certain crucial information may only exist within specific time intervals or segments of a cycle. For example, a sudden movement or a pause in breathing might appear briefly in a particular IMF. On the other hand, some features span the entire time series, reflecting longer-term trends or patterns. For example, a person's overall health condition or emotional state could be mirrored in their long-term heart rate and respiratory patterns.

Multiscale Nature: Vibration signals may simultaneously encompass changes across multiple scales. For example, heartbeat is a periodic activity on a shorter timescale, whereas respiration occurs on a longer timescale.

4) Network Architecture of TVFI-Net: 2D Tensor Convert: To overcome these challenges, our TVFI-net reshapes the one-dimensional time series into two-dimensional temporal tensors. Each column of these 2D tensors contains time points within a specific time interval, while each row represents time points at the same phase across different intervals. By doing so, we transcend the limitations of one-dimensional time series and obtain a 2D representation that explicitly captures temporal changes within periods as well as across different time intervals. This approach allows for a clear depiction of both intra-period and inter-interval temporal dynamics.

Specifically, for a time series of length T with recorded variables of C, $\mathbf{x_{1D}} \in \mathbb{R}^{T\times C}$, for each of its estimated periods i, we compute the period length p_i, Eq:

$$p_i = \left\lceil \frac{T}{f_i} \right\rceil, i \in \{1,\ldots,k\}, \tag{5}$$

where T is the total length of the time series, f_i is the frequency of the i^{th} period, and K is the number of different periods estimated, which is a hyperparameter.

Based on the selected frequencies $\{f_1,\ldots,f_k\}$ and the corresponding period lengths $\{p_1,\ldots,p_k\}$, we reshape the one-dimensional time series $\mathbf{X}_{1\text{D}} \in \mathbb{R}^{T\times C}$ into multiple two-dimensional temporal tensors:

$$\mathrm{X}_{2\text{D}}^{i} = \text{Reshape}_{p_i,f_i}(\text{Padding}(\mathbf{X}_{1\text{D}})), i \in \{1,\ldots,k\}, \tag{6}$$

where Padding$(\cdot)$ serves to append a zero at the end of the data sequence along the time dimension to make it compatible with Reshape$_{p_i,f_i}(\cdot)$, where p_i and f_i represent the number of rows and columns of the transformed two-dimensional tensor, respectively. In this way, the original one-dimensional time series tensor $\mathbf{X}_{1\text{D}} \in \mathbb{R}^{T\times C}$ is converted into the i^{th} reshaped time series based on frequency f_i, resulting in a two-dimensional tensor $\mathrm{X}_{2\text{D}}^{i} \in \mathbb{R}^{p_i\times f_i\times C}$ whose columns and rows represent intra-period changes and inter-period changes under the corresponding cycle length p_i, respectively.

TimesBlock: As show in Fig. 7,we arrange multiple TimesBlocks sequentially. The input sequence first passes through an embedding layer to obtain deep features $\mathbf{X}_{1\text{D}}^{0} \in \mathbb{R}^{T\times d_{\text{model}}}$. For the l^{th} layer of TimesBlock, its input is $\mathbf{X}_{\text{lD}}^{l-1} \in \mathbb{R}^{T\times d_{\text{model}}}$, and then it proceeds through two-dimensional convolution to extract two-dimensional temporal changes:

$$\mathbf{X}_{1\text{D}}^{l} = \text{TimesBlock}\left(\mathbf{X}_{1\text{D}}^{l-1}\right) + \mathbf{X}_{1\text{D}}^{l-1}. \tag{7}$$

For each TimesBlock, the entire process comprises two consecutive parts: capturing two-dimensional changes over time and adaptively aggregating representations from different periods.

From 1D to 2D representation: First, the one-dimensional temporal features $\mathbf{x}_{\text{lD}}^{l-1}$ are processed to extract periodicities, and then they are converted into two-dimensional tensors to represent two-dimensional temporal changes, as described in the previous section:

$$\begin{aligned} \mathbf{A}^{l-1}, \{f_1,\ldots,f_k\}, \{p_1,\ldots,p_k\} &= \text{Period}\left(\mathbf{X}_{1\text{D}}^{l-1}\right) \\ \mathbf{X}_{2\text{D}}^{l,i} &= \text{Reshape}_{p_l,f_i}\left(\text{Padding}(\mathbf{X}_{1\text{D}}^{l-1})\right), i \in \{1,\ldots,k\}. \end{aligned} \tag{8}$$

Extracting 2D time-varying characterizations: For the set of two-dimensional tensors $\{\mathbf{X}_{2D}^{l,1}, \mathbf{X}_{2D}^{l,2}, \ldots, \mathbf{X}_{2D}^{l,k}\}$, due to their two-dimensional locality, we can employ 2D convolutions to extract information. Here, we opt for the classic Inception model, that is:

$$\widehat{\mathrm{X}}_{2\text{D}}^{l,i} = \text{Inception}\left(\mathrm{X}_{2\text{D}}^{l,i}\right), i \in \{1,\ldots,k\}. \tag{9}$$

From 2D to 1D representation: For the extracted temporal features, we transform them back into a one-dimensional space to facilitate information aggregation.

$$\widehat{\mathbf{X}}_{\text{lD}}^{l,i} = \text{Trunc}\left(\text{Reshape}_{1,(p_i\times f_i)}(\widehat{\mathbf{X}}_{2\text{D}}^{l,i})\right), i \in \{1,\ldots,k\}, \tag{10}$$

where $\widehat{\mathbf{X}}_{1\text{D}}^{l,i} \in \mathbb{R}^{T\times d_{\text{model}}}$, Trunc$(\cdot)$ denotes the removal of the zeros added by the Padding$(\cdot)$ operation from the previous step.

Adaptive aggregation: We perform a weighted sum of the obtained one-dimensional representations $\{\widehat{\mathbf{X}}^{l,1},\ldots,\widehat{\mathbf{X}}^{l,k}\}$ according to the intensity of their corresponding frequencies, resulting in the final output:

$$\begin{aligned} \widehat{\mathbf{A}}_{f_1}^{l-1},\ldots,\widehat{\mathbf{A}}_{f_k}^{l-1} &= \text{Softmax}\left(\mathbf{A}_{f_1}^{l-1},\ldots,\mathbf{A}_{f_k}^{l-1}\right) \\ \mathrm{x}_{1\text{D}}^{l} &= \sum_{i=1}^{k} \widehat{\mathrm{A}}_{f_i}^{l-1} \times \widehat{\mathrm{X}}_{1\text{D}}^{l,i}. \end{aligned} \tag{11}$$

Through the aforementioned design, TVFI-net accomplishes the process of modeling temporal changes by "extracting two-dimensional temporal variations for each period separately and then performing adaptive fusion." Since both intra-period and inter-period changes involve multiple highly structured two-dimensional tensors, the TimesBlock can simultaneously capture multi-scale two-dimensional temporal variations comprehensively. Therefore, TVFI-net can achieve more effective representation learning compared to directly processing one-dimensional time series.

*Categorization:*The final one-dimensional tensor x_{1D}^{l} is used for the classification task. To achieve classification, we connect a fully connected layer followed by a softmax layer at the end to output the class probability distribution. The output layer consists of four nodes, corresponding to "No anomaly", "Single person present", "Two people present", and "Multiple people present".

5) Time and Space Complexity Analysis: The architecture of TVFI-net mainly consists of two components: the *feature extraction module* and the *lightweight classification network module*. Specifically:

- The *feature extraction module* generates multiple IMF components via EMD, followed by time-frequency transformation and statistical feature extraction. Its complexity is $O(N \log N)$, where N denotes the length of the input signal.
- The *classification network module* employs a modified lightweight Transformer structure, whose core operation is the multi-head self-attention mechanism. The time complexity of this part is $O(d_{\text{model}}^2 \cdot L)$, where d_{model} is the dimension of the featurefeature and L length of the sequence. L is fixed (7×9), As a result, the computational complexity becomes constant with respect to input size, i.e.,$O(1)$.

Therefore, the overall time complexity of TVFI-net can be expressed as:

$$T(N) = O(N \log N). \tag{12}$$

In comparison, the SVM training phase has a complexity of $O(N^3)$, while convolution operations in CNN models typically have a complexity of $O(C \cdot N \log N)$, where C is the number of channels. TVFI-net achieves a single detection latency of less than **0.5 seconds**, meeting the requirements for real-time detection.

The memory footprint of TVFI-net mainly comes from the following aspects:

1) *Input feature storage:* A fixed-length feature matrix (size 7×9) is generated after EMD preprocessing, which requires minimal memory.
2) *Model parameter storage:* The total number of parameters in TVFI-net is approximately **85 K**, significantly fewer than typical CNN models (often exceeding millions of parameters).
3) *Intermediate buffer overhead:* The attention weight matrices and residual connections in the Transformer architecture introduce some additional memory usage, but are kept within reasonable limits through optimization.

Thus, the space complexity of the TVFI-net is as follows:

$$S(N) = O(d_{\text{model}} \cdot L + P), \tag{13}$$

where P represents the total number of model parameters.

Due to its lightweight design, TVFI-net can run efficiently on embedded or edge computing platforms without relying on high-performance GPUs or cloud computing resources.

Through both theoretical analysis of the computational complexity of TVFI-net, we observe that the model exhibits low computational and memory overhead, offering excellent real-time performance and deployment flexibility. Compared to traditional methods, TVFI-net not only achieves significantly improved detection accuracy, but also demonstrates clear advantages in terms of resource utilization efficiency. This makes it particularly suitable for real-world applications such as concealed human detection in vehicles, where fast response times and limited hardware resources are critical constraints.

IV. Implementation

A. System Implementation

In utilizing the LDV for vibration detection on the test vehicles, we required the coordination of a laptop for data acquisition and storage. The laptop used was the Lenovo Xiaoxin 15 Pro 2021 model, equipped with an Intel Core i7 1165G7 CPU, an NVIDIA GeForce MX450 graphics card, 16 GB of RAM, and a 512 GB solid-state drive. The LDV device used was the Sunny Optical LV-S01 product [17]. In our experiments, the LV-S01 laser vibrometer showcased a velocity resolution exceeding 0.001 μm/s/$\sqrt{Hz}$; a ranging resolution above 0.31 picometers; a working frequency range spanning from DC to 3 MHz and up to DC to 25 MHz; it utilized a helium-neon laser source with a wavelength of 632.8 nanometers; and boasted a maximum measurement speed of 20 meters per second and a minimum displacement sensitivity reaching 0.04 nanometers. A high-precision wearable sensor (Bigrun Team x-26) to record ECG and respiratory effort signals as standard physiological references. Data processing and deep learning network training were carried out on a workstation paired with an RTX 3090 graphics card.

B. Experiment Setup

We enlisted 6 subjects and prepared twelve distinct vehicle types for testing, including the Peugeot 2002, Volkswagen Magotan, Honda Odyssey (MPV), Ford Mondeo, Wuling Hongguang (Minivan, Mid-size Minivan, Small Flatbed Truck, Medium Box Truck), Jiangling (Medium Box Truck, Medium Flatbed Truck), Iveco, trailer chassis. For the sake of convenience, we'll replace the designations of these vehicle type with V1-V12 in order in the subsequent content. The experimental environments comprised an underground garage (ideal scenario) and a regular road (typical scenario),with a laser distance of 80-180 cm, including one measurement point,no special surface treatment (natural reflection). Throughout the experiment, we arranged for participants to experience a variety of configurations, including passenger cars: empty car, one person in the driver's seat, one person in the front passenger seat, one person in the back seat, two people in the back seat, and one person in both the front and back seats; trucks: empty car, one person lying/sitting in the cargo bay, two people lying/sitting in the cargo bay, three people lying/sitting in the cargo bay, etc. Subjects were instructed not to perform large-scale movements during the trials. During the formal detection phase, both the driver and passengers were required to turn off the engine and exit the vehicle, leaving only any potentially hidden individuals inside.

V. Evaluation

To evaluate the performance of the SiVe system, we conducted controlled experiments across twelve different vehicle types. The experimental scenarios encompassed a range of conditions, including vehicles with zero, one, two, and three occupants, as well as indoor and outdoor environmental settings. A total of 6 subjects were recruited for these experiments to ensure a representation of the variability in human biometric signals. Each subject underwent multiple trials across all experimental conditions to thoroughly test the system's responsiveness under diverse scenarios. Data were systematically collected from all vehicle types and environmental conditions, covering the full spectrum of vibrations produced under the specified scenarios. Subsequently, the SiVe system analyzed these datasets to determine the presence and number of individuals within the enclosed spaces.

Evaluation Metrics: To comprehensively evaluate the performance of the SiVe system, we adopted a variety of evaluation metrics to ensure its accuracy and reliability are measured from multiple perspectives. In addition to focusing on overall accuracy, which is the proportion of correctly classified samples out of the total number of samples, we also introduced precision, recall, and the F1 score. These metrics respectively measure the system's accuracy in detecting hidden individuals, its ability to identify all vehicles that actually contain hidden individuals, and the harmonic mean of these two measures. Furthermore, the false positive rate (FPR) and false negative rate (FNR) were used to assess the occurrences of false alarms and missed detections, ensuring that the system does not generate excessive unnecessary alerts or overlook critical information in real-world applications.

Considering the requirements for real-time applications, we placed special emphasis on processing time and resource utilization. Processing time refers to the average duration required for each detection, which is crucial for ensuring the system can complete tasks within a reasonable timeframe. Resource utilization, on the other hand, pertains to the use of computational resources during system operation, aiming to ensure efficient resource usage to support large-scale deployment.

Overall Performance: We first conducted a comprehensive evaluation of the SiVe system under different vehicle models and environmental conditions, and then examined various factors that might affect its performance. We have determined the following overall performance:

In different vehicle models and environmental conditions, the SiVe system can accurately determine the presence of individuals in enclosed spaces, with an accuracy rate exceeding 99%. Despite variations in the physical characteristics of the subjects, the SiVe system maintained high accuracy across all experimental conditions, with detection accuracy rates consistently above 99%.

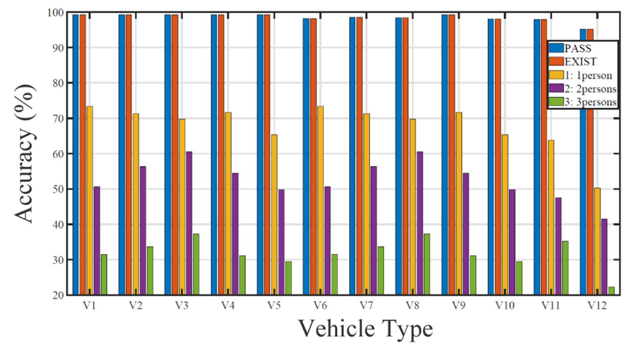


Fig. 9. Reflection conditions impact of SiVe system.

Additionally, our evaluation of laser incidence angles revealed significant impacts on the system's performance.This enhanced performance is attributed to better capture of vertical vibration signals, reduced structural interference, and more stable feature information from lower body parts.Thus, optimizing the laser incidence angle provides a significant improvement in occupant count accuracy while maintaining high overall detection reliability.The size of the time slice and the sampling rate can influence the accuracy of the SiVe system to a certain extent. Through comprehensive analysis, we selected the most appropriate time slice and sampling rate.

Reflection Conditions Impact on Detection: Due to the sensitivity of LDV signals to the reflectivity of the target surface, and given that SiVe operates by detecting vehicle exteriors, the reflectivity variation is primarily influenced by differences in surface materials across vehicle models. To evaluate system performance under varying reflectivity conditions, we conducted repeated experiments on vehicles with different reflective properties under identical testing conditions. For each vehicle, we recorded key metrics including detection accuracy, false positive rate, and false negative rate. All tests were performed at a fixed distance of 80 cm, with consistent LDV settings.

The tested vehicles exhibited three types of surface reflectivity, as illustrated in Fig. 9:

- *Normal reflectivity:* Standard painted metal surfaces (V1–V5, V9)
- *Low reflectivity:* Matte black coatings or heavily weathered old bodies (V6–V8, V10–V11)
- *Dust-covered surfaces:* A truck trailer covered with dust (V12).

Results show that the detection accuracy for the presence of individuals (Pass:Absence of individuals Exist:Presence of individuals) exceeds 99% under normal reflectivity conditions, remains above 98% under low reflectivity, and reaches 95% even under dust-covered conditions. Despite some performance degradation under lower reflectivity, SiVe achieves high accuracy and robustness through the high sensitivity of LDV, the adaptive noise suppression capability of EMD, and the generalization power of TVFI-net.

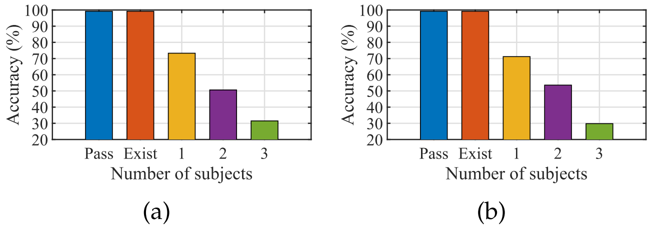


Fig. 10. Environmental Impact of SiVe System: (a) Detection Accuracy for Peugeot 2002 for Presence of Individuals (indoor); (b) Detection Accuracy for Peugeot 2002 for Presence of Individuals (outdoor). [Pass:Absence of individuals Exist:Presence of individuals].

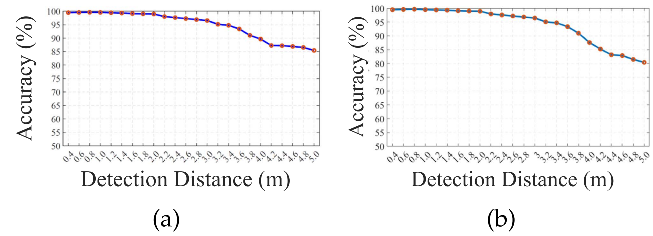


Fig. 11. Impact of Distance on Accuracy of SiVe System: (a) Classification Accuracy at Various Distances in an Enclosed Environment; (b) Classification Accuracy at Various Distances in an Open Environment.

However, the accuracy in estimating the exact number of individuals gradually declines. For example, in vehicle V1 (Peugeot 2002), the accuracy for detecting one person is 73.33%, which drops to 50.61% for two persons, and further decreases to 31.52% for three or more. This trend is consistently observed across other vehicle types, indicating increasing detection difficulty with higher occupancy.

Nonetheless, this limitation does not undermine the system's practical utility, as any positive detection can trigger manual inspection to prevent false negatives. Furthermore, accurately identifying the number of concealed individuals is an important objective we aim to improve in future work, in order to enhance the system's overall detection capability.

Environmental Impact on Detection: Under varying environmental conditions, using the Peugeot 2002 as an example, the detection accuracy of SiVe for the presence of personnel, as shown in Fig. 10(a) and (b) for indoor and outdoor conditions respectively, reveals consistent reliability. Indoors, the system maintains an accuracy rate above 99% in determining whether the space contains individuals. Regarding the number of occupants, the precision is 73.33% for one person, 50. 61% for two and 31. 52% for three or more. Similarly, outdoors, the system also achieves over 99% accuracy in identifying whether individuals are present, with specific occupant counts having accuracies of 71.24% for one person, 53.62% for two, and 29.83% for three or more. It is evident that, regardless of whether indoors or outdoors, SiVe can accurately determine if there are individuals within the confined space, showing remarkable consistency and resilience to external environmental factors. However, in terms of precisely identifying the number of individuals, the system's performance is somewhat inferior, particularly as the number of occupants increases, leading to a decline in recognition accuracy. This indicates that, across different environments, the core strength of SiVe lies in its high-accuracy determination of whether individuals are present, whereas identifying the exact number of individuals presents a limitation.

Impact of Distance on Accuracy: When assessing the performance of the SiVe system, we particularly examined the impact of the distance between the LDV and the detection target on the accuracy of the system. This parameter is a critical factor affecting non-contact biometric signal detection, especially under varying environmental conditions. To fully understand how changes in distance influence detection accuracy, experiments were conducted in both open and enclosed environments, collecting vibration data at different distances.

To evaluate the impact of the distance between the LDV and the detection target on system accuracy, we conducted experiments in both open and enclosed environments. In these experiments, we collected vibration data at distances ranging from 40 cm to 500 cm in increments of 20 cm and assessed the average accuracy of these data using classification algorithms.

As shown in Fig. 11(a) and (b), in an enclosed environment, the system's accuracy remains high even as the distance between the LDV device and the detection target increases. This is primarily due to the reduced background noise and fewer reflective interferences in enclosed spaces, allowing the LDV to more clearly capture weak vibration signals caused by human activities. In an open environment, despite the presence of more background noise and other sources of interference, the system maintains high accuracy at short distances. However, in both environments, accuracy gradually decreases as the distance increases. Specifically, at short distances, the accuracy can reach 99%; however, as the distance extends beyond a certain limit, the accuracy diminishes, particularly in complex backgrounds where background noise has a more significant impact.

In summary, the distance between the LDV and the detection target significantly influences the accuracy of the SiVe system. At shorter distances, regardless of whether the environment is open or enclosed, the system can maintain high accuracy; however, as the distance increases, the the accuracy gradually declines. Therefore, to ensure the reliability and efficiency of the SiVe system across various practical applications, it is most appropriate to maintain a distance of 80 cm to 180 cm between the device and the target.

Impact of EMD Processing on Accuracy: To further investigate the impact of EMD processing on detection accuracy, we compared the performance of EMD-processed and unprocessed data in a classification algorithm within an enclosed environment.

Fig. 12(a) and (b) show that EMD preprocessing significantly improves the accuracy and stability of classification. At close distances, the SiVe system's accuracy with EMD processing can exceed 99%, and even at longer distances, it maintains an accuracy above 85%. In contrast, without EMD processing, the system's accuracy remains around only 50% regardless of distance. This is because EMD can extract Intrinsic Mode Functions related to human vital signs from complex vibration

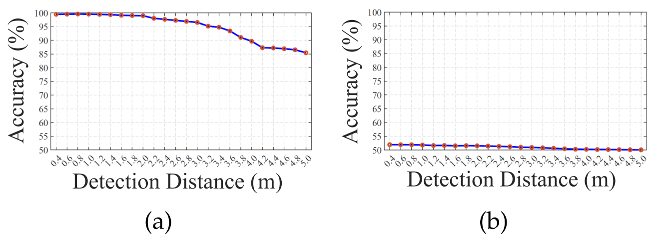


Fig. 12. Impact of EMD Processing on Accuracy of SiVe System: (a) Classification Accuracy of EMD-Processed Data at Various Distances in an Enclosed Environment; (b) Classification Accuracy of unEMD-Processed Data at Various Distances in an Enclosed Environment.

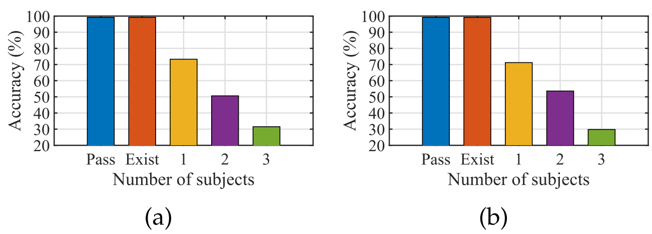


Fig. 13. Detection Position Impact of SiVe system: (a) Detection from Perpendicular Position on Surface; (b) Detection from Below the Vehicle.[Pass:Absence of individuals Exist:Presence of individuals].

signals, and these IMFs contain crucial information about low-frequency body movements such as breathing and heartbeats. In comparison, unprocessed raw data contain substantial noise and irrelevant information, leading to significantly lower classification accuracy.

In summary, EMD processing substantially enhances the detection accuracy of the SiVe system in complex environments. To ensure the reliability and efficiency of the SiVe system across various practical applications, especially for long-distance detection, employing advanced signal processing techniques like EMD is essential.

Impact of Detection Position on System Performance: To evaluate the impact of detection position on detection performance, we conducted tests using a Peugeot 2002 vehicle at two different positions: one where the laser detects from a perpendicular position on the vehicle's surface and another where it detects upward from beneath the vehicle to the chassis. The experimental results show that both methods achieve over 99% accuracy in determining whether individuals are present within the space. However, in terms of accurately identifying the specific number of occupants, the method involving detection from below the vehicle exhibits superior performance, as shown in Fig. 13(a) and (b).

Detection from Perpendicular Position on Surface: When the laser detects from a perpendicular position on the vehicle's surface, the system achieves accuracies of 73.33%, 50.61%, and 31.52% for detecting one person, two people, and three or more individuals, respectively.

Detection from Below the Vehicle: In contrast, when the laser detects upward from beneath the vehicle to the chassis, the system's accuracy significantly improves. The detection rates for one person, two people, and three or more individuals increase to 79.02%, 64.74%, and 40.18%, respectively.

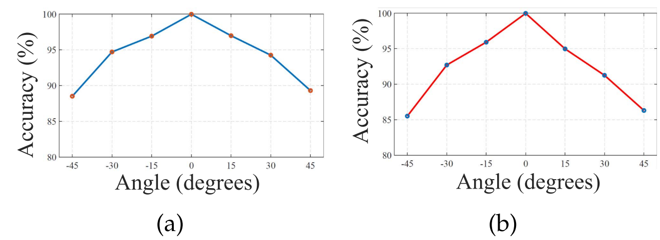


Fig. 14. Laser incidence angle impact of SiVe system: (a) Detection accuracy of SiVe system at various incidence angles in a closed environment; (b) Detection accuracy of SiVe system at various incidence angles in an open environment.

The primary reason for this difference lies in the fact that the human body's center of gravity and major sources of vibration (thoracic/abdominal regions) exert their weight directly on the bottom surface. Abdominal movements caused by respiration and heartbeat-generated vibrations are transmitted to the container floor through the feet or torso, with minimal signal attenuation. At the same time, vibration signals detected from the side walls of the vehicle must propagate through complex structural paths before reaching the measurement point, during which vibration information is prone to attenuation due to the effects of material damping. In contrast, bottom detection allows for direct acquisition of raw vibration signals without structural interference. This provides more stable information on biological characteristics, helping to distinguish between different numbers of individuals.

Impact of laser Incidence Angle: To evaluate the impact of laser incidence angle on the detection accuracy of the SiVe system, we set up relevant experiments to collect data at different incidence angles in both closed and open environments. These angles ranged from -45 degrees to +45 degrees, with a step size of 15 degrees. The experimental results are shown in Fig. 14(a) and (b), allowing us to analyze how changes in the incidence angle affect the system's detection performance.

In the closed environment, due to relatively less background noise and more controlled environmental conditions, the reflection characteristics of the laser should theoretically be more stable, which helps improve detection accuracy. However, as the incidence angle changes, the angle between the laser beam and the normal of the surface being measured alters, potentially affecting the intensity and quality of the reflected light, thereby impacting the clarity of the vibration signal received by the LDV. Specifically, At an incidence angle of 0 degrees, where the laser beam is perpendicular to the surface, the strongest and clearest reflection signal can be obtained, leading to the highest detection accuracy of the system. As the incidence angle increases (whether positively or negatively), the effective reflective area decreases, the signal strength weakens, and more scattering phenomena may be introduced, leading to signal distortion or increased noise, thus reducing detection accuracy. The situation in an open environment is more complex because there are other external vibration sources and environmental factors that influence the propagation path of the laser and its reflection characteristics. Therefore, when conducting detection in an open environment, the effect of the incidence angle may

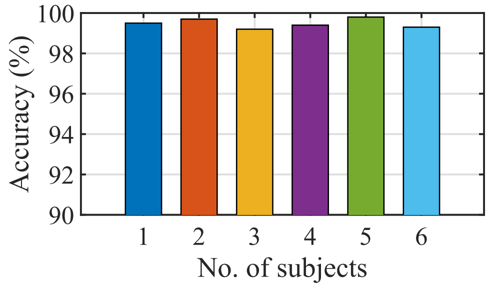


Fig. 15. Impact of Subject Characteristics of SiVe system.

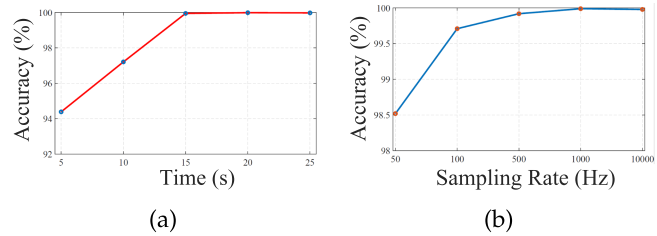


Fig. 16. Performance evaluation of SiVe system.

be more pronounced. In addition to the aforementioned effects of the incidence angle on the reflection signal, there may be additional interferences in open environments, such as sunlight or interference from other light sources, all of which could lead to a further decline in detection accuracy.

For both environments, an incidence angle close to 0 degrees generally provides the highest detection accuracy, as at this angle the laser can most effectively illuminate the target and acquire its vibration information. To ensure that the SiVe system achieves optimal performance across various application scenarios, especially when used in open environments, it is advisable to maintain the laser incidence angle as close to 0 degrees as possible and consider environmental conditions to optimize the detection setup.

Impact of Subject Characteristics: To assess the impact of different subject characteristics on the performance of the SiVe system, we conducted specialized analyses on six subjects. The height and weight combinations of the subjects were as follows: Subject 1, 165 cm tall and 60 kg; Subject 2, 175 cm tall and 75 kg; Subject 3, 180 cm tall and 80 kg; Subject 4, 155 cm tall and 55 kg; Subject 5, 170 cm tall and 65 kg; and Subject 6, 185 cm tall and 90 kg.During the experiment, each subject underwent individual testing, with multiple trials conducted for each subject, changing their position in the vehicle during each trial and comparing the results to empty vehicle test samples. As shown in the Fig. 15, the detection accuracy for all six subjects was above 99%. Despite the significant differences in height and weight among the subjects, the SiVe system maintained high accuracy across all experimental conditions. This means that the system is capable of effectively adapting to individuals of different body types, enhancing the reliability and practical applicability of the SiVe system.

Impact of Time Slices: To optimize the performance of our neural network for individual detection within enclosed spaces, we conducted experiments with varying time slice durations. The tested time slices included intervals of 5 seconds, 10 seconds, 15 seconds, 20 seconds, and 25 seconds. We evaluated model accuracy under each time slice to determine the optimal setting that balances detection efficacy with operational efficiency.The experimental results, as illustrated in Fig. 16(a), show a significant improvement in model accuracy as the time slice increases from 5 seconds up to 20 seconds, reaching the highest accuracy of 99.9% at 20 seconds. Notably, the increase in accuracy from a 15-second to a 20-second time slice is marginal, and the enhancement from 20 seconds to 25 seconds is even less pronounced. This suggests that beyond a certain threshold, further extending the duration of the time slice does not lead to substantial improvements in model performance.

Considering the working environment of our detection system, timely response and minimal latency are critical factors. Therefore, a 20-second time slice emerged as the most suitable choice. It provides near-optimal accuracy of 99.9% while maintaining the system's responsiveness. Theoretically, although longer time slices might offer slightly higher accuracy, they would introduce unnecessary delays that could compromise the system's practicality and effectiveness in real-world applications. Based on this empirical analysis and comprehensive evaluation considering practical deployment, we have determined that a 20-second time slice will be the standard configuration for implementing the neural network.

Impact of Sampling Rate: To investigate the impact of the sampling rate on the performance of the SiVe system, we conducted experiments at various sampling rates. The selected sampling rates were 50 Hz, 100 Hz, 500 Hz, 1000 Hz, and 10000 Hz. We evaluated the accuracy of the model at each sampling rate to determine the optimal selection, as shown in Fig. 16(b). The detection accuracy increased with the increase in sampling rate, reaching 99% at 100 Hz. This phenomenon can be explained from a signal processing perspective: as the sampling rate increases, the system captures more detailed information, which helps improve detection accuracy. Particularly in the case of low-frequency vibrations, a higher sampling rate can better capture these subtle vibration features, which is crucial for accurately determining the presence of individuals. However, once the sampling rate reaches a certain level, further increases in sampling rate result in limited improvements in accuracy. Excessively high sampling rates may not bring significant performance improvements but can increase computational load and data storage requirements. Based on this analysis, we ultimately chose a sampling rate of 500 Hz for the system.

Impact of Vehicle Movement: To evaluate the impact of vehicle movement on the detection performance of the SiVe system, we conducted experiments comparing the detection accuracy under stationary and slowly moving (idling) conditions. A passenger car was selected as the test vehicle, and the average detection accuracy across multiple trials was recorded under different scenarios.

The experimental results show that the system detection accuracy drops significantly when the vehicle is in motion, with an accuracy of only around 60%. The main reasons for the

TABLE I
AVERAGE ACCURACY OF THE MODEL

TVFI-net	CNN	SVM	Transformer
99.72%	95.34%	90.26%	93.43%

degraded performance during vehicle movement include: The high-frequency noise caused by road bumps and engine vibrations during driving overlaps with the physiological signals, increasing the difficulty of signal separation. The transmission path of the vibration signal becomes unstable during vehicle movement, weakening the continuity of the signal. The relative displacement caused by vehicle movement leads to frequency shifts in the reflected laser signals, affecting both signal strength and phase stability. Despite the observed decline in detection performance during vehicle movement, it does not compromise the practical applicability of the system. In real-world scenarios such as border security checkpoints or logistics inspection stations, vehicles are required to come to a complete stop before inspection, eliminating interference caused by movement. In addition, in cases where the objective is to detect concealed individuals, passengers can be asked to exit the vehicle, thereby reducing interference from complex vibration sources.

Model Comparison: To capture and model complex multi-periodic variations, we therefore designed TVFI-net. As mentioned earlier, the time series involved in vibration analysis encompass complex temporal patterns, and traditional classification networks often struggle to extract sufficient semantics from such data for effective analysis. By conducting classification experiments on data sets collected from various vehicle models using several models, we compared their average accuracy. Our results, as shown in Table.I demonstrate that TVFI-net outperforms CNN, SVM, and Transformer. The experimental out comes confirm that TVFI-net can better extract features related to human presence and body movements from complex time series data.

Physiological Signal Synchronization and Validation: To validate whether the IMFs extracted by EMD genuinely reflect human physiological activities rather than environmental noise or decomposition artifacts, we conducted a synchronized physiological baseline recording experiment using high-precision wearable sensors.

Multiple human subjects were recruited to rest inside a stationary vehicle under three representative postures: seated (static), supine (static), and lateral (with mild motion), simulating realistic concealed occupant scenarios. A LDV was positioned outside the vehicle to non-invasively measure micro-vibrations on the exterior surface (e.g., door panel). Simultaneously, each subject wore a Bigrun Team x-26 wearable sensor to record high-fidelity ECG and respiratory belt signals, serving as reference standards for heart rate (HR) and respiration rate (RR).

The LDV-acquired signal was decomposed into IMFs using Ensemble EMD. IMFs corresponding to cardiac (1–2 Hz) and respiratory (0.1–0.5 Hz) frequency bands were identified through spectral analysis and cross-correlation matching. Instantaneous HR and RR were then extracted from the dominant IMFs and quantitatively compared with the reference measurements from the Bigrun Team x-26.

As shown in Table II, the LDV-based HR estimates exhibited excellent agreement with the reference across all postures: $R^2 > 0.93$, RMSE < 1.05 bpm, MAE < 0.83 bpm. Bland-Altman analysis revealed that 95% of differences fell within ± 0.72 bpm, indicating strong consistency. In contrast, no periodic oscillations were observed in the empty-vehicle tests (no human present), and the correlation between the extracted signal and a virtual baseline was negligible (< 0.01), effectively ruling out false positives induced by environmental noise or structural resonance.

These results demonstrate that the IMFs extracted from vehicle surface vibrations are physiologically meaningful and directly linked to human vital signs. This validation constitutes an independent, objective, and quantitative evidence chain for the physiological origin of the extracted features, going beyond mere spectral observation.

Quantitative Validation: To quantitatively verify the physiological correspondence of iIMFs obtained via EMD, we conducted a controlled experiment to confirm that IMF1 predominantly carries cardiac information while IMF3 encodes respiratory dynamics.

In a static in-vehicle scenario, we simultaneously recorded LDV signals along with ground-truth physiological measurements: ECG for heart rate and a respiratory belt for breathing rate. The LDV signal was first low-pass filtered (0-10 Hz, fourth-order Butterworth) to suppress non-physiological artifacts, followed by EMD to yield a set of IMFs.

We then estimated physiological rates from specific IMFs using spectral peak detection:

- Heart rate ($\hat{HR}$) was estimated from IMF1 by identifying the dominant frequency peak in the range [0.8, 2.0] Hz via fast Fourier transform (FFT).
- Respiratory rate ($\hat{RR}$) was estimated from IMF3 by detecting the peak within [0.1, 0.5] Hz using the same FFT-based approach.

Let HR_{gt} and RR_{gt} denote the ground-truth heart and respiratory rates derived from the wearable sensors. We define the relative error (RE) as:

$$RE_{\text{heart}} = \frac{\left|\hat{HR} - HR_{\text{gt}}\right|}{HR_{\text{gt}}}, \quad RE_{\text{breath}} = \frac{\left|\hat{RR} - RR_{\text{gt}}\right|}{RR_{\text{gt}}}. \quad (14)$$

As shown in Fig. 18, the relative error for heart rate estimation using IMF1 is consistently low, confirming its strong fidelity to cardiac activity. Similarly, respiratory rate estimated from IMF3 exhibits minimal relative error, indicating that IMF3 is the primary carrier of breathing-related oscillations. In contrast, when IMF1 is used to estimate respiration or IMF3 to estimate heart rate, the relative errors increase dramatically (often exceeding 20%) demonstrating poor physiological relevance in these mismatched configurations.

These results provide statistical evidence that the EMD decomposition yields physiologically meaningful components:

TABLE II
QUANTITATIVE AGREEMENT BETWEEN LDV-ESTIMATED HEART RATE AND BIGRUN TEAM X-26 REFERENCE MEASUREMENTS

Condition	**Sample Size**	**RMSE (bpm)**	**MAE (bpm)**	R^2	**BA Range (bpm)**	**Corr**
Seated (Static)	50	0.82 ± 0.15	0.65 ± 0.12	0.96	-0.52 to $+0.48$	-0.03
Supine (Static)	50	0.78 ± 0.13	0.61 ± 0.11	0.97	-0.49 to $+0.45$	-0.01
Lateral (Mild Motion)	50	1.05 ± 0.18	0.83 ± 0.15	0.93	-0.72 to $+0.68$	-0.02
Empty-Vehicle Test	50	N/A	N/A	< 0.01	N/A	0.00

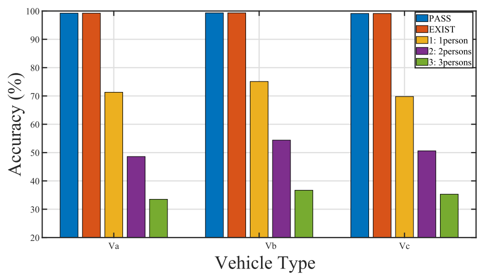
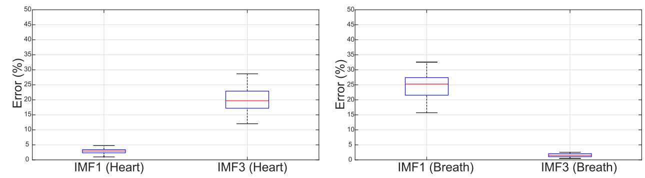


Fig. 17. External validation experiment.



(a) (b)

Fig. 18. (a): Relative Error for Heart Rate Estimation. (b): Relative Error for Breathing Rate Estimation.

IMF1 is specifically associated with heartbeat, while IMF3 corresponds to respiration.

External Validation Experiment: To systematically evaluate the potential overfitting of TVFI-net and assess its generalization capability in real-world scenarios, we conducted an external validation experiment using a completely independent test dataset. Such external dataset was collected with no overlap in vehicle models, environmental conditions, or participant population compared to the original training set. Hence it can provide a rigorous assessment of model robustness and cross-scenario transferability.

Specifically, a new dataset was acquired at different times (three months apart) and distinct geographical locations , the test road segments were distinct from those used in the primary data collection, to ensure environmental diversity. The experiments were conducted on three widely-used passenger vehicle models in the market, representing different body structures and dynamic characteristics: FAW-Volkswagen Sagitar, a compact sedan (Va); GAC-Toyota Highlander, a mid-size SUV (Vb); Geely Xingyue L, a domestically produced intelligent SUV (Vc).

A total of 12 healthy adult participants (6 male, 6 female) were recruited for this external test, none of whom had participated in the original training data collection. The experimental setup remains consistent with previous experiments.

The TVFI-net model trained on the original dataset was applied directly to this unseen dataset without any fine-tuning or parameter adaptation (i.e., zero-shot inference). Experimental results in Fig. 17 show that the model achieved an average accuracy of 99.23% on the external dataset. Compared to the performance on the original test set (exceeds 99%), the degradation is less than 0.5 percentage points, the model remains within the high-precision range without significant performance degradation, indicating it has not overfitted to specific vehicle models, environments, or individuals. Those results demonstrate that TVFI-net has learned universal micro-vibration patterns caused by human physiological activities, rather than specific noise or artifacts in the training data. As a result, it can exhibit strong cross-scenario generalization capabilities.

Ablation experiments based on feature sequence length: To validate the effectiveness of selecting seven Intrinsic Mode Functions (IMFs) and extracting nine statistical features from them as the input feature vector, a series of ablation studies were conducted. These experiments were designed to systematically assess the impact of varying the number of IMFs as well as different feature combinations on the performance of the TVFI-net model.

The dataset used in this study comprises vibration signal data collected from twelve different types of vehicles under various indoor and outdoor environmental conditions. In the first set of experiments, TVFI-net models were trained using 3, 5, 7, and 9 IMFs, respectively, and their performance was evaluated on the same test set. For each configuration, the average detection accuracy, recall rate, and precision were computed and recorded.

The selection of seven IMFs was based on their ability to cover different frequency components of the signal. These IMFs effectively capture low-frequency vibrations associated with human respiration and heartbeat, as well as mid-frequency components related to subtle body movements and partial environmental noise. Using more than seven IMFs may introduce excessive high-frequency noise that could degrade model performance, whereas using fewer IMFs may lead to the loss of critical signal information, thereby reducing detection accuracy.

Furthermore, the selection of nine statistical features—including mean, standard deviation, maximum and minimum values, skewness, kurtosis, root sum of squares, and the first and third quartiles—was motivated by their ability to comprehensively characterize the statistical properties of each IMF. These features have been experimentally validated to play a significant role in distinguishing between different occupancy states. They not only reflect the central tendency and dispersion of the signal but also capture its distribution shape and extreme value variations, thereby enhancing the discriminative capability of the model.

Experimental results show that the model achieved the best detection performance when using seven IMFs, with an average detection accuracy exceeding 99%. Reducing or increasing the number of IMFs led to a decline in performance, indicating that seven IMFs represent a well-balanced trade-off between information retention and noise suppression.

After comparing the performance of different feature combinations, the final set of nine features—mean, standard deviation, maximum, minimum, skewness, kurtosis, root sum of squares, and the first and third quartiles—was selected. This is because these features effectively capture the key statistical characteristics of each IMF and are crucial for improving classification accuracy.

VI. Related Work

A. Personnel Detection Techniques

Personnel presence detection systems that operate behind barriers in enclosed environments play a crucial role in major security systems such as customs and public security agencies. Currently, there are mainly two detection methods. One is based on the detection of vibration sensors, using tactile vibration sensors [18], [19], [20]. Although these can indicate human activity by capturing minute vibrations, their limitations include the potential alteration of the original vibration characteristics of the object being measured, leading to reduced reliability of the measurement results. There is also a risk of contamination or damage to the surface of the object being measured, and the reliance on physical contact limits the practicality of this method in various detection scenarios.

The second method is non-contact detection, typically through radio frequency (RF) signal monitoring systems [21], [22], [23], [24], [25], [26], [27], [28]. These systems often exhibit the ability to penetrate obstacles and utilize machine learning algorithms [29] for data analysis, allowing personnel activity monitoring without direct line of sight. For example, FreeSense [23] is a novel human detection method based on Wi-Fi signals that detects human motion by identifying phase differences among amplitude waveforms received by multiple antennas; Liang X et al. [25] use Ultra-Wide Band (UWB) signals to detect human vital signs through walls; Pa-Count [26] proposes an effective real-time passenger counting system deployed within vehicles using Wi-Fi Channel State Information (CSI). However, the inherent properties of RF signals, such as longer wavelengths and limited spatial resolution, restrict their accuracy in complex environments. Moreover, due to the lack of directional specificity in RF signals, especially in close-range monitoring, distinguishing signals from closely located objects becomes difficult, adding complexity to signal processing.

The personnel detection technology proposed in this study, based on LDV, stands out distinctly from traditional contact-based and RF signal detection methods. The extremely short wavelength of lasers allows for submicron to nanometer-level high spatial resolution, ensuring precise non-contact measurement of minute vibrations, avoiding any interference caused by contact. The high directionality and focusing ability of lasers enable precise targeting and measurement of vibrations in specific areas, effectively suppressing background noise and significantly improving the signal-to-noise ratio and accuracy of the measurements.

B. Applications of LDV

LDV stand out for their non-contact measurement capabilities, high resolution, and wide dynamic range, showcasing remarkable utility across interdisciplinary fields [30]. In mechanical engineering, LDV play a central role in modal analysis, precisely determining structural vibration characteristics such as natural frequencies, damping ratios, and mode shapes, providing a powerful tool for assessing structural dynamic performance [31], [32]. Specifically, in the complex analysis of mechanical vibrations, LDV offer a more accurate and comprehensive means of diagnosing vibrations in engines and rotating machinery, revolutionizing environmental vibration health monitoring in industries [33], [34].In human health monitoring systems, the non-intrusive measurement ability of LDV is extensively utilized to monitor the minute vibrations of human tissues [35], opening new avenues in medical diagnostics and biomechanical studies. For example, LDV have been applied in both clinical and non-clinical settings to monitor cardiac activity, revealing subtle changes in heart function [10].Due to the high directionality and coherence of lasers, maintaining good focus and signal strength even at a distance, LDV have been applied to the field of remote speech signal acquisition [36], [37], enabling non-contact measurement of micro-vibrations produced by speech and subsequent recovery of intelligible voice signals.Moreover, LDV have a specific and advanced application in the safety aspects of autonomous vehicles – adversarial attacks and defense strategies [38], [39]. Benefiting from the high-precision vibration measurement characteristics of LDV and their consistency and accuracy, we have utilized LDV for the detection of hidden individuals in vehicles, demonstrating excellent results.

VII. Conclusion

In this paper we propose SiVe, an advanced personnel presence detection system that employs LDV technology to achieve high-precision detection of individuals concealed within enclosed spaces under non-contact conditions. Through a series of meticulously designed and rigorously controlled experiments, we tested the performance of the SiVe system across different vehicle models and various environmental settings. The results demonstrate its exceptional capability in accurately determining the presence of individuals with detection accuracy rates exceeding 99% in both indoor and outdoor environments.The SiVe system exhibits outstanding reliability and stability, performing consistently regardless of variations in vehicle model or external conditions. Its core strength lies in its high-precision judgment of whether individuals are present.

Although the SiVe system demonstrates high accuracy and robustness in non-contact detection, there are still several limitations in its current implementation that need to be considered for real-world deployment:

- *Cost and Availability of Laser Equipment:* The current system relies on high-precision LDV, such as the LV-S01, which costs over \$50,000. This high cost limits the scalability and widespread adoption of the system. In addition, the calibration and maintenance of LDV devices require specialized technical expertise, increasing operational complexity.
- *Deployment Constraints:* For optimal performance, the system requires the vehicle to be turned off, with all known occupants exiting the vehicle, leaving only potential concealed individuals inside. While this condition is feasible in controlled environments such as border checkpoints, it may not always be practical in dynamic or time-sensitive scenarios.
- *Laser Incidence Angle and Detection Distance Limitations:* The system performs best when the laser is incident vertically and at a distance of no more than 180 cm. In real-world settings, however, vehicle parking angles, obstructions, or terrain variations may prevent achieving ideal detection conditions, directly affecting system stability and reliability.
- *Limitation in Counting Capability:* While the system achieves high accuracy in detecting a single individual, its performance degrades significantly when multiple people are present. This is primarily due to overlapping vibration signals generated by multiple people, making it difficult for the model to distinguish physiological features such as breathing patterns. Additionally, complex interior structures and object placements within the vehicle further hinder accurate person counting.

Although these limitations exist in the current implementation, we believe they can be progressively addressed through continued research and development. Future work will focus on improving system performance, reducing costs, enhancing adaptability, and refining detection accuracy, particularly in multi-person scenarios, to enable broader and more reliable applications in real-world settings.

References

[1] Newsweek, "Illegal immigration costs US billions: Biden administration policy impact on taxpayer burden," 2023. Accessed: Jul. 12, 2024. [Online]. Available: https://www.newsweek.com

[2] T. W. Rogers, N. Jaccard, E. J. Morton, and L. D. Griffin, "Automated x-ray image analysis for Cargo security: Critical review and future promise," *J. X-ray Sci. Technol.*, vol. 25, no. 1, pp. 33–56, 2017.

[3] O. Hupe and U. Ankerhold, "X-ray security scanners for personnel and vehicle control: Dose quantities and dose values," *Eur. J. Radiol.*, vol. 63, no. 2, pp. 237–241, 2007.

[4] S. Pan, M. Mirshekari, P. Zhang, and H. Y. Noh, "Occupant traffic estimation through structural vibration sensing," in *Proc. Sensors Smart Structures Technol. Civil, Mech., Aerosp. Syst.*, 2016, pp. 41–52.

[5] M. Mirshekari, S. Pan, J. Fagert, E. M. Schooler, P. Zhang, and H. Y. Noh, "Occupant localization using footstep-induced structural vibration," *Mech. Syst. Signal Process.*, vol. 112, pp. 77–97, 2018.

[6] T. Wang et al., "Perspective on advances in laser-based high-resolution mass spectrometry imaging," *Anal. Chem.*, vol. 92, no. 1, pp. 543–553, 2019.

[7] M. Schewe, D. Kohlmann, H. Wulfmeier, H. Fritze, and C. Rembe, "Differential laser Doppler vibrometry for displacement measurements down to 1 mHz with 1 nm amplitude resolution in harsh environments," *Measurement*, vol. 210, 2023, Art. no. 112576.

[8] G. Petrie and C. K. Toth, "Introduction to laser ranging, profiling, and scanning," in *Topographic Laser Ranging Scanning*, Boca Raton, FL, USA: CRC Press, 2018, pp. 1–28.

[9] W. Sun et al., "Vibrosense: Recognizing home activities by deep learning subtle vibrations on an interior surface of a house from a single point using laser Doppler vibrometry," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 4, no. 3, pp. 1–28, 2020.

[10] L. Antognoli, S. Moccia, L. Migliorelli, S. Casaccia, L. Scalise, and E. Frontoni, "Heartbeat detection by laser Doppler vibrometry and machine learning," *Sensors*, vol. 20, no. 18, 2020, Art. no. 5362.

[11] H. Aygun and A. Apolskis, "Tracheal sound acquisition using laser Doppler vibrometer," in *Proc. Inst. Acoust.*, U.K., vol. 40, no. Pt. 1, 2018, pp. 362–369.

[12] K. Kroschel, A. Luik, and J. Metzler, "Contactless measurement of heart sounds and heart rate with laser Doppler vibrometry," in *Proc. Konferenz Elektronische Sprachsignalverarbeitung*, 2020, pp. 1–8.

[13] N. Huang, "Computer implicated empirical mode decomposition method," *Appar. Artic. Manuf. US Provisional Appl.*, vol. 60, 1996, Art. no. 822.

[14] O. Singh et al., "ECG signal denoising based on empirical mode decomposition and moving average filter," in *Proc. 2013 IEEE Int. Conf. Signal Process., Comput. Control*, 2013, pp. 1–6.

[15] M.-G. Murariu, F.-R. Dorobanu, and D. Tărniceriu, "A novel automated empirical mode decomposition (EMD) based method and spectral feature extraction for epilepsy EEG signals classification," *Electronics*, vol. 12, no. 9, 2023, Art. no. 1958.

[16] K. V. Madhav, M. R. Ram, E. H. Krishna, N. R. Komalla, and K. A. Reddy, "Estimation of respiration rate from ECG, BP and PPG signals using empirical mode decomposition," in *Proc. 2011 IEEE Int. Instrum. Meas. Technol. Conf.*, 2011, pp. 1–4.

[17] Sunny Optical Technology Co., Ltd., "LV-S01 single-point laser doppler vibrometer," 2023, Accessed: Jul. 22, 2024. [Online]. Available: https://sunnyinnova.com/products/13526.html

[18] R. Damarla and D. Ufford, "Personnel detection using ground sensors," in *Unattended Ground, Sea, and Air Sensor Technologies and Applications IX*, vol. 6562. Bellingham, WA, USA: SPIE, 2007, pp. 28–37.

[19] P.-S. Huang, T. Damarla, and M. Hasegawa-Johnson, "Multi-sensory features for personnel detection at border crossings," in *Proc. IEEE 14th Int. Conf. Inf. Fusion*, 2011, pp. 1–8.

[20] X. Jin, S. Gupta, A. Ray, and T. Damarla, "Multimodal sensor fusion for personnel detection," in *Proc. IEEE 14th Int. Conf. Inf. Fusion*, 2011, pp. 1–8.

[21] S. Di Domenico, M. De Sanctis, E. Cianca, and M. Ruggieri, "WiFi-based through-the-wall presence detection of stationary and moving humans analyzing the Doppler spectrum," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 33, no. 5/6, pp. 14–19, May/Jun. 2018.

[22] W. Li, B. Tan, and R. J. Piechocki, "WiFi-based passive sensing system for human presence and activity event classification," *IET Wireless Sensor Syst.*, vol. 8, no. 6, pp. 276–283, 2018.

[23] T. Xin et al., "FreeSense: A robust approach for indoor human detection using Wi-Fi signals," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 2, no. 3, pp. 1–23, 2018.

[24] X. Li et al., "IndoTrack: Device-free indoor human tracking with commodity Wi-Fi," *Proc. ACM Interactive, Mobile, Wearable Ubiquitous Technol.*, vol. 1, no. 3, pp. 1–22, 2017.

[25] X. Liang, T. Lv, H. Zhang, Y. Gao, and G. Fang, "Through-wall human being detection using UWB impulse radar," *EURASIP J. Wireless Commun. Netw.*, vol. 2018, no. 1, pp. 1–17, 2018.

[26] H. Jiang, S. Chen, Z. Xiao, J. Hu, J. Liu, and S. Dustdar, "Pa-count: Passenger counting in vehicles using Wi-Fi signals," *IEEE Trans. Mobile Comput.*, vol. 23, no. 4, pp. 2684–2697, Apr. 2024.

[27] S. Muhammad et al., "A contactless and non-intrusive system for driver's stress detection," in *Proc. Adjunct ACM Int. Joint Conf. Pervasive Ubiquitous Comput. 2023 ACM Int. Symp. Wearable Comput.*, 2023, pp. 58–62.

[28] S. Chen et al., "BeamCount: Indoor crowd counting using Wi-Fi beamforming feedback information," in *Proc. 25th Int. Symp. Theory, Algorithmic Found., Protoc. Des. Mobile Netw. Mobile Comput.*, 2024, pp. 1–10.

[29] Y. Liu, T. Wang, Y. Jiang, and B. Chen, "Harvesting ambient RF for presence detection through deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 4, pp. 1571–1583, Apr. 2020.

[30] S. J. Rothberg et al., "An international review of laser Doppler vibrometry: Making light work of vibration measurement," *Opt. Lasers Eng.*, vol. 99, pp. 11–22, 2017.

[31] A. Stanbridge and D. Ewins, “Modal testing using a scanning laser Doppler vibrometer,” *Mech. Syst. Signal Process.*, vol. 13, no. 2, pp. 255–270, 1999.
[32] M. S. Allen and M.W. Sracic, “A new method for processing impact excited continuous-scan laser Doppler vibrometer measurements,” *Mech. Syst. Signal Process.*, vol. 24, no. 3, pp. 721–735, 2010.
[33] J. R. Bell and S. Rothberg, “Rotational vibration measurements using laser Doppler vibrometry: Comprehensive theory and practical application,” *J. Sound Vib.*, vol. 238, no. 4, pp. 673–690, 2000.
[34] A. Malekjafarian, D. Martinez, and E. J. OBrien, “The feasibility of using laser Doppler vibrometer measurements from a passing vehicle for bridge damage detection,” *Shock Vib.*, vol. 2018, no. 1, 2018, Art. no. 9385171.
[35] L. Mignanelli and C. Rembe, “Non-contact health monitoring with LDV,” in *Laser Doppler Vibrometry Non-Contact Diagnostics*, Berlin, Germany: Springer, 2020, pp. 1–8.
[36] Y. Wang, W. Zhang, Z. Wu, X. Kong, Y. Wang, and H. Zhang, “Highly-accurate and real-time speech measurement for laser Doppler vibrometers,” *IEICE TRANSACTIONS Inf. Syst.*, vol. 105, no. 9, pp. 1568–1580, 2022.
[37] C. Cai, K. Iwai, and T. Nishiura, “Speech enhancement for laser Doppler vibrometer dealing with unknown irradiated objects,” *IEICE Trans. Fund. Electron., Commun. Comput. Sci.*, vol. 106, no. 4, pp. 647–656, 2023.
[38] C. Yan et al., “Rolling colors: Adversarial laser exploits against traffic light recognition,” in *Proc. 31st USENIX Secur. Symp.*, 2022, pp. 1957–1974.
[39] Z. Jin, X. Ji, Y. Cheng, B. Yang, C. Yan, and W. Xu, “PLA-LiDAR: Physical laser attacks against LiDAR-based 3D object detection in autonomous vehicle,” in *Proc. 2023 IEEE Symp. Secur. Privacy*, 2023, pp. 1822–1839.

Zhu Xiao (Senior Member, IEEE) received the MS and PhD degrees in communication and information systems from Xidian University, China, in 2007 and 2009, respectively. He is currently a full professor with the College of Computer Science and Electronic Engineering, Hunan University, China. His research interests include Internet of Vehicles and mobile computing. He is serving as an associated editor for *IEEE Transactions on Intelligent Transportation Systems*.

Shirong Guan is currently working toward the MS degree in computer science and electrical engineering with Hunan University, China. His research interests include mobile computing and signal processing.

Jingyang Hu received the PhD degree from Hunan University, in 2025. He is currently a postdoctor with the University of Science and Technology of China. From 2022 to 2023, he works as a joint PhD student with the School of Computer Science and Engineering, Nanyang Technological University (NTU). His research interests include mobile and pervasive computing, the Internet of Things, and machine learning.

Siyu Chen is currently working toward the PhD degree with the College of Computer Science and Electronic Engineering, Hunan University, China. From 2024 to 2025, he works as a joint PhD student with the College of Computing and Data Science at Nanyang Technological University (NTU), Singapore. His research interests lie in the Internet of Things and large language models.

Hongbo Jiang (Senior Member, IEEE) received the PhD degree from Case Western Reserve University, in 2008. He is a full professor in the College of Computer Science and Electronic Engineering, Hunan University. He was a professor with the Huazhong University of Science and Technology. His research concerns computer networking and mobile networks. He is serving as the editor for *IEEE/ACM Transactions on Networking*, the associate editor for *IEEE Transactions on Mobile Computing*.

Keqing Li (Fellow, IEEE) received the BS degree in computer science from Tsinghua University, in 1985 and the PhD degree in computer science from the University of Houston, in 1990. He is a SUNY distinguished professor with the State University of New York and a National Distinguished Professor, Hunan University (China). He has authored or co-authored more than 1230 journal articles, book chapters, and refereed conference papers. He holds nearly 80 patents announced or authorized by the Chinese National Intellectual Property Administration. He is among the world’s top few most influential scientists in parallel and distributed computing,regarding single-year impact (ranked 2) and career-long impact(ranked 3)based on a composite indicator of the Scopus citation database. He received the IEEE TCCLD Research Impact Award from the IEEE CS Technical Committee on Cloud Computing, in 2022 and the IEEE TCSVC Research Innovation Award from the IEEE CS Technical Community on Services Computing, in 2023. He won the IEEE Region 1 Technological Innovation Award (Academic), in 2023. He was a recipient of the 2022-2023 International Science and Technology Cooperation Award and the 2023 Xiaoxiang Friendship Award of Hunan Province, China. He is a member of the SUNY Distinguished Academy.He is an AAAS Fellow, an AAIA Fellow, an ACIS Fellow, and an AIIA Fellow. He is a member of the European Academy of Sciences and Arts. He is a member of Academia Europaea (Academician of the Academy of Europe).