



Robust Cardiac Feature Monitoring based on Millimeter-Wave Radar

Thesis submitted in accordance with the requirements of the University of Liverpool
for the degree of Doctor in Philosophy by

Yuanyuan Zhang

May 2025

Abstract

Wireless sensing brings revolutions for multiple traditional industries and empowers numerous emerging industries such as autonomous driving and device-free wellness monitoring. Among all the wireless sensors (e.g., Wi-Fi, acoustic sensors), radar-based sensing reveals outstanding performance in terms of detection range, applicable scenarios and robustness. This thesis will focus on a promising and challenging research area to realize the robust electrocardiogram (ECG) monitoring from millimeter-wave (mmWave) radar, enabling contactless vital sign monitoring for future in-cabin monitoring, elderly people caregiving and even clinical diagnosis.

Benefit from decades of development in radar-based coarse vital sign monitoring (e.g., heart rate, respiration), frequency-modulated continuous wave (FMCW) radar with high operating frequency (e.g., 66 or 77Ghz) is becoming mainstream in radar front-end design, because such configurations could capture subtle displacements caused by cardiac activities and provide enough resolution for isolating targets from specific 3D points, encouraging the related research to extract fine-grained ECG signals as the golden standard in clinical diagnosis and realize robust monitoring in the presence of real-world noises.

In this thesis, radar-based ECG recovery is thoroughly investigated to provide a contactless ECG measurement that gets rid of cumbersome wired connections and adhesive electrode patches. However, several challenges need to be solved: (a) an efficient signal processing algorithm is required to ensure a high signal-to-noise ratio (SNR) for the collected radar signal, providing enough cardiac features for the later ECG recovery; (b) the domain transformation for the cardiac activities within single-cardiac cycle from mechanical to electrical domain needs to be modeled to realize a robust ECG recovery against noises; (c) the long-term ECG recovery should be modeled and realized instead of relying on the purely data-driven methods that can hardly resist noise disturbance; (d) the developed deep learning model normally relies on the large-scale radar/ECG dataset for training, and alleviating data scarcity is an important issue especially for the deployment in the new scenarios with limited data.

Based on the challenges above, multiple novel algorithms and deep learning frameworks are proposed. Firstly, a **cardio-focusing and -tracking (CFT)** algorithm is proposed to iteratively approach the point with a high-SNR radar signal extracted. Secondly, two deep learning models, **radarODE** and **radarODE-MTL**, are designed to realize robust single-cycle and long-term ECG recovery, respectively. Thirdly, a data augmentation method **Horcrux** and a transfer learning framework **RFcardi** are proposed to jointly decrease the demand for data acquisition. Finally, extensive experiments are performed based on both public and private datasets to show the effectiveness of the proposed algorithms and frameworks. It is believed that this thesis contributes to the general development of the wireless sensing community and brings the future applications of wireless wellness monitoring closer to our daily lives.

Acknowledgements

Traveling back to 2022, I can still feel the confusion and uncertainty when I first attempt to explore my potential research directions and learn the rules of academia. At this moment, when finishing the thesis, I genuinely would like to express my gratitude to the people who taught me a lot and accompanied me during my Ph.D. study.

I would first like to thank my supervisor, Dr. Rui Yang, for the dedicated supervision. Throughout my studies, Dr. Yang offered great flexibility and patience when I wandered in various directions and also provided crucial suggestions many times when I encountered difficulties or fell into the wrong path. After four years of working with Dr. Yang and all the members of the Center for Intelligent Control and Optimization (CICO), I learned a lot about how to conduct academic research and how to pursue self-actualization.

I would also like to thank my supervisor Dr. Yutao Yue. As the leader of the Institute of Deep Perception Technology (IDPT), Dr. Yue provided a conducive environment for undisturbed study and research. During four years of working in IDPT, Dr. Yue taught me a lot about how academic research could empower industrial de-

velopment, and the bits and pieces of daily life with members of IDPT also left an indelible mark in my memories.

I also value all my friends from all the stages before my Ph.D. study. Many of you have become my co-authors, given me valuable advice, or asked me to cite your papers during these years. Since the full name list may exceed the pages of this thesis, I would like to thank my childhood friends from Huai'an, Nanjing, my undergraduate friends now spread in Shanghai, Guangzhou, HongKong, London, Seattle, Champaign, and my postgraduate friend now live in Tokyo, Fukuoka, London.

Last and most importantly, I could never thank my parent enough for the decades of love and support mentally and financially. Although I am still waiting for the day until becoming the son of a billionaire, my parents have already taught me the most important thing by precept and example, that is to be a person with honesty, integrity and kindness, because "*There is no prosthetic for an amputated spirit.*"

Publications

Journal paper:

1. **Yuanyuan Zhang**, Runwei Guan, Lingxiao Li, Rui Yang, Yutao Yue, Eng Gee Lim, “radarODE: An ODE-Embedded Deep Learning Model for Contactless ECG Reconstruction from Millimeter-Wave Radar”, *IEEE Transactions on Mobile Computing*, Apr. 2025.
2. **Yuanyuan Zhang**, Rui Yang, Yutao Yue, Eng Gee Lim, “radarODE-MTL: A Multi-Task Learning Framework with Eccentric Gradient Alignment for Robust Radar-Based ECG Reconstruction”, *IEEE Transactions on Instrumentation and Measurement*, Apr. 2025.
3. **Yuanyuan Zhang**, Rui Yang, Yutao Yue, Eng Gee Lim, Zidong Wang, “An Overview of Algorithms for Contactless Cardiac Feature Extraction From Radar Signals: Advances and Challenges”, *IEEE Transactions on Instrumentation and Measurement*, Jul. 2023.

Conference paper:

1. **Yuanyuan Zhang**, Sijie Xiong, Rui Yang, Eng Gee Lim, Yutao Yue, “Recover from Horcrux: A Spectrogram Augmentation Method for Cardiac Feature Monitoring from Radar Signal Components”, in *2025 47th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC2025)*, Jul. 2025.

Under Review:

1. **Yuanyuan Zhang**, Haocheng Zhao, Sijie Xiong, Rui Yang, Eng Gee Lim, Yutao Yue, “From High-SNR Radar Signal to ECG: A Transfer Learning Model with Cardio-Focusing Algorithm for Scenarios with Limited Data”, *IEEE Transactions on Mobile Computing*. (Under Review)
2. Sijie Xiong, **Yuanyuan Zhang**, Cheng Tang, Haoling Xiong, Yiding Li, Atsushi Shimada, “U-MA: A Unified Framework with Differential Mamba under Parallel U-Net Scheme for Time Series Forecasting”, *Engineering Applications of Artificial Intelligence*. (Under Review)
3. Sijie Xiong, Cheng Tang, **Yuanyuan Zhang**, Haoling Xiong, Youhao Xu, Atsushi Shimada, “CME-Mamba with Enhancing Nonlinear Dependencies for Time Series Forecasting”, *Applied Soft Computing*. (Under Review)

Contents

Abstract	iii
Acknowledgements	v
Publications	vii
Contents	ix
List of Figures	xv
List of Tables	xvii
List of Acronyms	xx
1 Introduction	1
1.1 Wireless Sensing for Vital Sign Monitoring	2
1.2 Radar-based Vital Sign Monitoring	3
1.3 Challenges in Radar-based ECG Recovery	5
1.4 Contributions and Thesis Outline	9

2 Background and Literature Review	11
2.1 Theoretical Background	11
2.2 Literature Review for Cardiac Feature Extraction	18
3 Efficient High-SNR Signal Acquisition	23
3.1 Introduction	23
3.2 Methodology	25
3.3 Details of Experiment and Dataset	33
3.4 Experimental Results and Evaluations	36
3.5 Conclusions	40
4 Robust Single-Cycle ECG Generation	42
4.1 Introduction	42
4.2 Methodology	45
4.3 Dataset and Implementation Details	60
4.4 Experimental Results and Evaluations	63
4.5 Conclusions	75
5 Robust Long-Term ECG Generation	77
5.1 Introduction	78
5.2 Background and Problem Statement	80
5.3 Methodology	85
5.4 Experimental Setting and Result Evaluation	92
5.5 Conclusions	108

6 Alleviate Data Scarcity for Radar-Based ECG Recovery	110
6.1 Introduction	111
6.2 Data Augmentation for Radar Spectrogram	113
6.3 Transfer Learning Framework Design	117
6.4 Experimental Results for Horcrux	120
6.5 Experimental Results for RFcardi	125
6.6 Conclusion	130
7 Conclusions and Future Work	131
7.1 Conclusions	131
7.2 Future Work	132
References	134

List of Figures

1.1	Illustration of different steps in radar-based cardiac monitoring.	3
1.2	Overview of the thesis with unsolved challenges.	6
2.1	Relationships between cardiac mechanical and electrical activities within the same cardiac cycle.	17
2.2	Illustration for spectrum-based methods: (a) raw radar signal; (b) spectrum obtained for raw radar signal from FFT, with RR, RR harmonic and HR peaks labelled; (c) raw radar signal after differentiation; (d) spectrum obtained for differentiated raw radar signal from FFT, with enhanced HR peak labelled.	19
3.1	Challenges for high-SNR signal collection: (a) and (b) Radar signals with high and low SNR for adjacent points with a distance of 0.03m.	24
3.2	Overview of using CFT for radar-based ECG recovery: (a) Rough localization of human body; (b) Use CFT to find CF point and extract high-SNR radar signals; (c) Deep learning for ECG recovery.	26
3.3	Procedures for obtaining RA map: (a) Range FFT for chirps along fast time; (b) Angle FFT along virtual channels.	27

3.4	Template for assessing SNR: (a) High-SNR radar signal; (b) Extracted signal envelope with the synthetic template; (c) (a) Low-SNR radar signal, (d) Extracted signal envelope with the synthetic template.	30
3.5	Illustration of the CFT algorithm with bold line wrapping the search region S_k : (a) Equality between γ and Γ (same as in CS algorithm); (b) Large Γ_k with refined γ_k , providing more potential points to be evaluated; (c) Jump out of the local minimum by adjusting Γ_k and γ_k	31
3.6	Indoor scenarios for data collection.	34
3.7	Visualization of the extracted radar signal for all methods: (a) - (c) If CF point is around rough body location; (d) - (f) If CF point is far from rough body location.	36
3.8	Illustration of performance in terms of peak error and MDR: (a) - (b) Scattered points with fitting curves along range axis; (c) - (d) CDF plots for all trails.	38
3.9	Impact of radar input quality on the final ECG recovery: (a) - (b) High-quality radar input and ECG recovery; (c) - (d) Low-quality radar input and ECG recovery.	40
4.1	Overview of the radarODE framework for domain transformation: (a) Single-cycle ECG length (PPI) estimation; (b) Single-cycle ECG generator; (c) Long-term ECG reconstruction.	45

4.2 Spectrograms obtained from the same radar signal: (a) Radar signal $\tilde{x}(t)$; (b) STFT result; (c) CWT result; (d) SST result with T_1 , T_2 labeled, revealing concentrated energy around vibrations and clean background.	49
4.3 Energy plot of the synchronous radar signals with different detected peaks: (a) Energy plot with high SNR and correct detection; (b) Energy plot with low SNR and wrong detection.	50
4.4 Architecture of SCEG with SST segments as input and single-cycle ECG pieces as output.	52
4.5 Structure of backbone composed of residual block and downsample block: DeformConv2d means deformable 2D convolution, BN is batch normalization, and ReLU is the rectified linear unit activation function.	55
4.6 The structure of the 9-layer TCN with dilation factor $D = 2$ and kernel size $K = 3$.	58
4.7 Environment settings for data collection from quasi-static subject [1].	61
4.8 PPI estimation error of KDE-based and mean-based method: (a) View of different subjects; (b) View of different physiological statuses; (c) CDF for the overall PPI estimation error.	65
4.9 Performance comparison for single-cycle ECG recovery: (a) - (d) CDF of RMSE, PCC, MDR and R-peak error for all trials.	66
4.10 Ablation study for training process: (a) Rigid reconstruction (ODE Recon.) from ODE decoder with ground truth (GT); (b) Training loss comparison.	68

4.11	Visualization of misalignment: (a) and (b) High- and Low-fidelity morphological references (Ref.) caused by PPI estimation error.	69
4.12	Corrupt ECG reconstruction: (a) and (b) Ideal reconstruction results for radarODE and MMECG; (c) Corrupt ECG reconstruction yield by MMECG and faithful reconstruction from radarODE; (d) Corresponding radar signal during body movements.	70
4.13	Statistical results for corrupt ECG reconstruction: (a) Overall missed detection rate (MDR); (b) MDR under different physical statuses. . .	71
4.14	Morphological accuracy comparison for all trials: (a) and (b) Overall CDF of RMSE and PCC for all trials.	72
4.15	Morphological accuracy comparison across all subjects: (a) and (b) RMSE and PCC across all subjects.	73
4.16	Morphological accuracy comparison for different physical statuses: (a) and (b) RMSE and PCC under different physical statuses.	74
4.17	CDF of the overall timing error.	75
4.18	Evaluations of fine-grained cardiac events: (a) and (b) The timing error for different peaks under various physical statuses for radarODE and MMECG.	76
5.1	The impact of strong noise and misalignment: (a) ECG recovery distorted by RBM noise [1]; (b) Misaligned ECG recovery due to the inaccurate PPI estimation [2].	79

5.2	Overview of the radarODE-MTL framework with EGA strategy: (a) Shared backbone extracts time-frequency features; (b) Morphological decoder only reconstructs the shape of the current ECG piece; (c) ECG anchor decoder estimates the time-index of anchors ; (d) Cycle length decoder estimates the cardiac cycle length; (e) The proposed EGA strategy for optimizing shared parameter space.	84
5.3	Illustration of EGA: (a) Original gradient space with gradient conflict and magnitude dominance; (b) The projection of the original gradient space into the orthogonal space with equal “learning rate”; (c) The implementation of eccentric gradient alignment to skew the joint gradient \tilde{g}_{joint} towards the hard task by introducing the eccentric vector v_{ecc} . .	88
5.4	Illustration of the recovered ECG with ground truth (GT) for different frameworks under no noise:(a) - (d) Radar signal with good quality and the recovered ECG signals.	98
5.5	Illustration of the recovered ECG with ground truth (GT) for different frameworks under constant noise:(a) - (d) Radar signal with constant noise and the recovered ECG signals.	99
5.6	Illustration of the recovered ECG with ground truth (GT) for different frameworks under abrupt noise: (a) - (d) Radar signal with abrupt noise and the recovered ECG signals.	100
5.7	Evaluations for long-term ECG recovery: (a) - (e) CDF plots of MDR, HR Error, RMSE, PCC and R^2 , with corresponding improvements. .	102

5.8	Evaluations for fine-grained ECG peaks recovery: (a) - (d) CDF plots of the timing error for QRST peaks, with corresponding improvements.	104
5.9	Noise robustness test: (a) Impact of constant noises with different intensities, (b) Impact of abrupt noises with different intensities and durations.	106
6.1	Illustration of Horcrux: (a) Traditional augmentation with zero mask; (b) Missed detected heartbeat (HB) and distorted ECG recovery; (c) Horcrux with time consistency preserved; (d) Ideal cardiac monitoring with high fidelity.	112
6.2	Pipeline of Horcrux with two branches: (a) Harmonic and percussive (H&P) components decomposition; (b) Dynamic template matching. .	114
6.3	Illustration of RFcardi with pre-text training and fine-tuning stages. .	118
6.4	Identified vibrations using DTM for radar signals with various qualities (illustrated in normalized amplitude).	122
6.5	Performance of using different scales of the dataset.	125
6.6	Results of SSR: (a) Failed SSR due to lack of data and sparse penalty; (b) Ideal SSR result with good MSE and sparsity.	126
6.7	Results of transfer learning using limited labeled data: (a) Poor ECG recovery without proper morphological feature and peak location; (b) Good ECG recovery owing to the pre-trained model.	128
6.8	Overall performance of the radar-based ECG recovery.	129

List of Tables

3.1	Parameters for data collection interface	34
3.2	Performance of supervised ECG recovery	39
4.1	Structure and Parameters for SCEG	52
4.2	Default values for η	56
4.3	Structure and Parameters for Long-term ECG Reconstruction	59
4.4	Parameters for Link Budget Analysis	62
4.5	Performance Comparison for Single-cycle ECG Recovery and Ablation Study	67
4.6	Absolute Timing Error for Reconstructed ECG Peaks	74
5.1	Comparison of different optimization strategies on radar-based ECG recovery	94
5.2	Comparison of different optimization strategies on indoor scene understanding	96
5.3	Comparison of the frameworks under different SNR	105
5.4	Comparison of the frameworks under abrupt noises	107

6.1	Parameters for Signal Model	117
6.2	Comparison of Different Augmentation Methods	123
6.3	Ablation Study for Horcrux	124
6.4	Performance of SSL with ablation study	126
6.5	Performance of ECG Recovery using different percentages of labeled data	127

Acronyms

AoA	angle of arrival
BN	batch normalization
BSS	blind source separation
CDF	cumulative distribution function
CF	cardio-focused
CFT	cardio-focusing and -tracking
CNN	convolutional neural network
CW	continuous wave
CWT	continuous wavelet transform
DFO	derivative-free optimization
DNN	deep neural network
DTM	dynamic template matching
ECG	electrocardiogram
EGA	eccentric gradient alignment
EMD	empirical mode decomposition
FFT	fast Fourier transform

FMCW	frequency-modulated continuous wave
FPGA	field-programmable gate array
H&P	harmonic and percussive
HPS	hard parameter sharing
HR	heart rate
HRV	heart rate variability
ICA	independent component analysis
IQ	in-phase and quadrature
IR-UWB	impulse-radio ultra wide band
KDE	kernel density estimation
LSTM	long short-term memory
MDR	missed detection rate
MIMO	multiple-input and multiple-out
mmWave	millimeter wave
MSE	mean square error
MTL	multi-task learning
MUSIC	multiple signal classification
ODE	ordinary differential equation
PCC	Pearson correlation coefficient
PPI	peak-to-peak interval
PSE	power spectrogram entropy

radar	radio detection and ranging
RBM	random body movement
ReLU	rectified linear unit
RMSE	root mean square error
RR	respiration rate
Rxs	receivers
SCEG	single-cycle ECG generator
SCG	seismocardiography
SNR	signal-to-noise ratio
SSL	self-supervised learning
SSR	sparse signal reconstruction/recovery
SST	synchrosqueezed wavelet transform
STFT	short-time Fourier transform
TCN	temporal convolutional network
TDM-MIMO	time division multiplexing multi-input multi-output
Txs	transmitters
VMD	variational mode decomposition

Chapter 1

Introduction

The World Health Organization reports that the global population is aging at an unprecedented rate, with one in six people globally being over 60 by 2030, and the number of elderly people over 60 will double to 2.1 billion by 2050. This demographic shift brings significant challenges, including a higher prevalence of chronic diseases, frailty, and complex health conditions among older adults [3]. Maintaining health and independence in later life is crucial not just for individuals, but for families and societies. As the world population ages, the early detection of diseases and daily monitoring of health conditions become critical to reducing the need for institutional care and improving quality of life, but reaching assistance from professional caregivers might be expensive or inconvenient especially when most of the older population will live in low- and middle-income countries [3]. Therefore, the current labor-intensive wellness monitoring needs a revolution, and wireless vital sign monitoring addresses these challenges by enabling continuous, automated monitoring of health indicators or emergency events (e.g., heart rate and fall detection) without constant human intervention.

1.1 Wireless Sensing for Vital Sign Monitoring

In the recent decade, wireless sensing has emerged as a transformative paradigm that leverages electromagnetic waves to remotely detect changes and gather information about objects, environments, and biological entities, and this non-invasive approach has been embedded in a wide range of applications from healthcare monitoring to autonomous driving, leading the rapid evolution of information and communication technologies towards the next generation [4–6]. Different from traditional contact-based sensing methods that require physical connections to the subject or environment, wireless sensing can be seamlessly integrated into daily life, fostering innovation in smart infrastructure and personalized health systems. For example, by leveraging radio frequency signals to detect subtle physiological movements such as respiration and heartbeat, radar-based systems eliminate the need for physical contact while maintaining clinical-grade accuracy, addressing key limitations of traditional wired sensors that cause patient discomfort, skin irritation, and infection risks [4].

Various wireless sensors can be utilized for vital sign monitoring according to the information received. For example, optical or thermal cameras sense the skin color or temperature variation caused by heartbeats [7, 8], acoustic sensors can monitor the heart sound [9], and Wi-Fi routers extract cardiac features from the channel state information [10]. However, the aforementioned sensors may be blamed for privacy issue [11], low accuracy [12] or vulnerability to the changing environment (e.g., light conditions or temperature variations) [13]. In contrast, radar senses the ambient environment through reflected signals mixed by chest wall displacement (induced by respiratory and cardiac activities) and all kinds of ambient noises [14], requiring proper algorithms to further extract the latent vital features. Additionally, compared with cameras, Wi-Fi routers and acoustic sensors, radar signal propagation is neither vulnerable to illumination/temperature/sound variations nor privacy-intrusive. Therefore, radar-based vital sign monitoring is promising to realize unobtrusive sensing in most scenarios after the design of advanced signal processing algorithms.

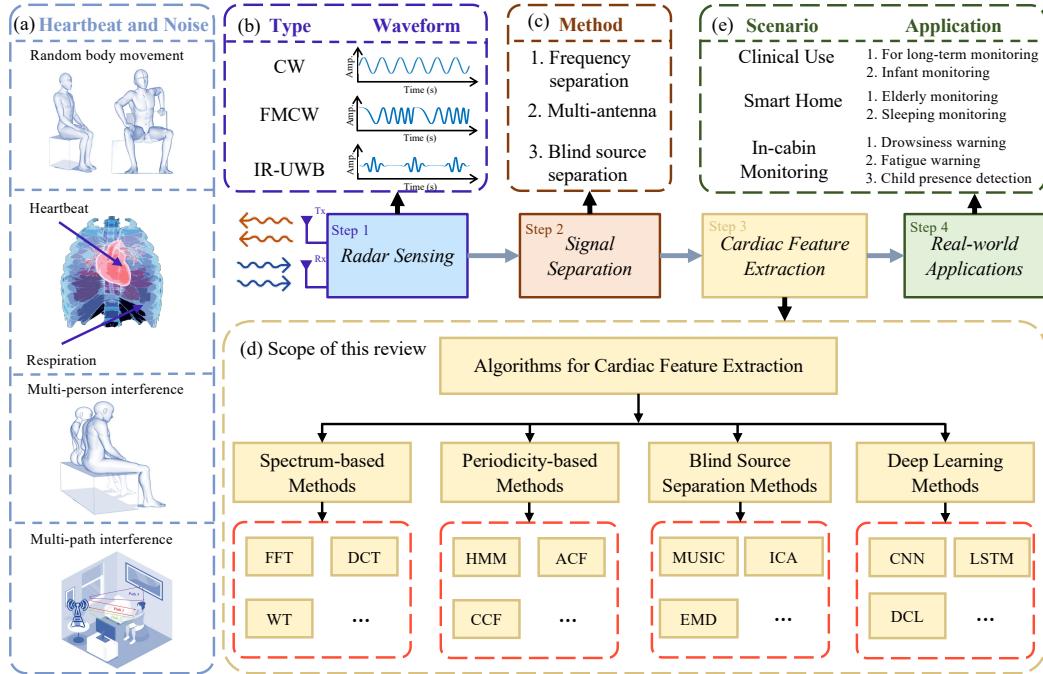


Figure 1.1: Illustration of different steps in radar-based cardiac monitoring.

1.2 Radar-based Vital Sign Monitoring

The first attempt at radar-based vital sign monitoring can be traced back to 1975 by measuring the displacement of the chest wall induced by respiration [15]. The chest wall displacement will modulate the phase component of the emitted radar signal, and the latent respiratory information can be demodulated from the phase variation [14]. Similarly, cardiac activities are small-scale displacements that also cause chest wall displacements, but such small displacements are normally ruined by respiration with orders more amplitude and need to be extracted using advanced signal processing algorithms.

The overview of the radar-based cardiac monitoring process and the downstream applications are shown in Figure 1.1. To measure the target cardiac features, the first step is to transmit the radio frequency signals with certain waveforms to sense the ambient environment as illustrated in Figure 1.1(b). After reflection, the received signals contain not only the target signal (vibration caused by heartbeats) but also

other noise signals such as random body movement (RBM), respiration and multi-person or multi-path interferences, as shown in Figure 1.1(a). The second step aims to isolate the signal reflected by a single human body from the background clutter or the interference from neighbors using specific methods, as shown in Figure 1.1(c) [16]. The third step is to extract the cardiac features from the reflected signals of a single person using various algorithms, with detailed classification provided in Figure 1.1(d). The last step is to analyze the obtained cardiac features for specific downstream applications as shown in Figure 1.1(e).

Most early studies focused on the recovery of coarse cardiac information, such as heart rate (HR), heart sound and heart rate variability (HRV), from the perspectives of radar front-end design or advanced algorithms design [4]. For example, some advanced types of radar (e.g., frequency-modulated continuous wave (FMCW) radar) are designed to enable high range-resolution or multi-person monitoring [17], and some baseband signal processing algorithms are embedded on the radar platform to realize in-phase and quadrature (IQ) modulation or accurate phase unwrapping [18]. In addition, various advanced algorithms are applied by leveraging different intrinsic characteristics of cardiac activities to robustly reconstruct cardiac features. For example, cardiac activities normally reveal strong periodicity in the time domain and have dominant peaks on the spectrum, inspiring periodicity-based methods (e.g., template matching [19], hidden Markov model [13]) and spectrum-based methods (e.g., Fourier transform [20], wavelet transform [21]) as two major categories in cardiac feature extraction algorithms.

In recent years, the emergence of commercial radar platforms with high operating frequency (i.e., millimeter wave (mmWave) radar) encourages researchers to extract fine-grained cardiac features (e.g., electrocardiogram (ECG) and seismocardiography (SCG)) from the radar signal [4]. SCG signal is measured by the accelerometer mounted on the human chest to measure the mechanical vibrations produced by heartbeats, describing the fine-grained cardiac mechanical activities such as aortic/mitral valve opening/closing and isovolumetric contraction [22]. Although these vibrations are subtle, it is still reasonable to directly map the displacements detected by radar to each fine-grained cardiac mechanical activity using high-resolution radar as proved

in [17]. However, the SCG measurement is not widely used in clinical scenarios, while the ECG signal is commonly recognized as the golden standard in cardiac monitoring because ECG describes the fine-grained cardiac activities, such as atrial/ventricular depolarization/repolarization, through the featured waveform (i.e., PQRST peaks) and is crucial to the diagnosis of cardiovascular diseases [23].

Based on the discussions above, it is believed that radar-based ECG recovery is an essential research direction to enable the realization of future wellness monitoring, even providing fine-grained cardiac features for clinical diagnosis. This thesis will thoroughly investigate all the stages in radar-based ECG recovery with multiply inventions proposed to overcome the unsolved challenges. In addition, this research has been approved by the University Ethics Committee of Xi'an Jiaotong-Liverpool University with proposal number ER-SAT-001000090020220906151929.

1.3 Challenges in Radar-based ECG Recovery

The realization of radar-based ECG recovery needs two general stages: (a) radar signal collection and pre-processing to increase signal-to-noise ratio (SNR); (b) ECG recovery from radar signal to realize domain transformation from cardiac mechanical activities to electrical activities using deep neural network (DNN), and this thesis will elaborate on four challenges figured in both stages with corresponding solutions proposed as shown in Figure 1.2.

1.3.1 The First Challenge

In the literature, many studies are dedicated to inventing advanced signal processing algorithms to enhance the signal quality, because the deep learning model for ECG recovery is vulnerable to the inputs contaminated by noises and requires high-SNR radar signals as inputs [1, 24]. The methods for capturing high-SNR radar signals can be categorized into two groups:

- The first type of method focuses on designing advanced radar front-end with multiple transmitters (Txs) and receivers (Rxs) [25, 26] or calibrating baseband

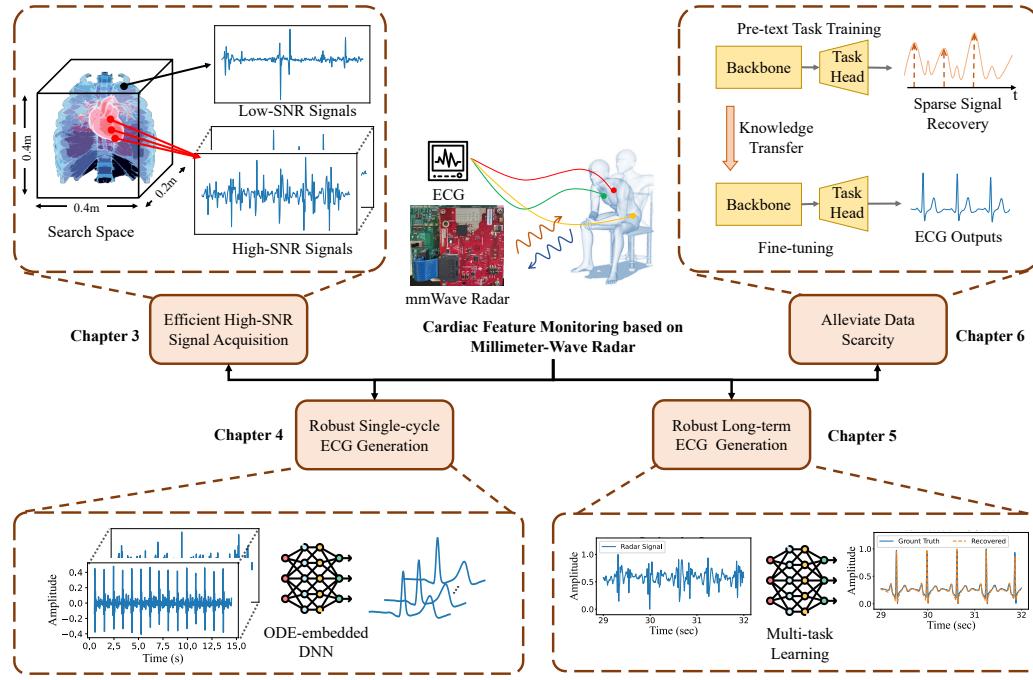


Figure 1.2: Overview of the thesis with unsolved challenges.

radar signals from in-phase and quadrature (IQ) channels to a circular shape [27–29].

- The second type of method assumes that the rough localization of human body provides accurate chest region with the majority of range bins containing useful cardiac features, and high-SNR signal can be obtained by selecting useful channels [30,31], applying clustering algorithms [1] or accumulating the signals from various dimensions (e.g., chirps, frames, antennas) [24].

The first type of method is not suitable for some commonly used frequency-modulated continuous-wave (FMCW) radar platforms (e.g., TI AWR-x radar) due to the on-board digital front-end module filtering the frequency-modulated feature of baseband signal (i.e., circular IQ plot) [32], preventing the broad applications of this approach in commercial radar. The second type of method relies on accurate localization of the chest region, while the existing methods only provide a rough location of the human body, causing a deviation of several decimeters due to different postures

of the subject [33]. Therefore, the methods based on signal accumulation may fail because only a minority of range bins contain cardiac features, hence not subjecting to the law of large numbers [24]. Although some aforementioned studies have proposed methods for selecting or clustering the useful range bins with cardiac features [1, 30], the computational cost for traversing a large objective space can be huge without an accurate cardiac location. In this case, the **first challenge** is to precisely locate and track the cardiac location during data collection to efficiently extract high-SNR radar signal.

1.3.2 The Second Challenge

To reconstruct ECG from radar signal, the researchers must deal with domain decoupling to transform the measured signal from the mechanical domain to the electrical domain to generate ECG measurement. Intuitively, it is reasonable that mechanical conduction and electrical conduction are highly correlated in describing cardiac activities because the electrical changes in cells trigger heart muscle contraction, whereas such a relationship is called excitation-contraction coupling in electrophysiology and is extremely hard to interpret or model by researchers without biological knowledge [23, 34].

In the literature, the existing studies all leverage deep learning methods to extract latent information from enormous radar/ECG pairs and try to learn domain transformation relying on the extraordinary non-linear mapping ability of the deep neural network [1, 30, 35–37]. Although these studies could successfully reconstruct the ECG signal from radar, there is no existing signal model with a compact form to describe the domain transformation for radar-based ECG reconstruction, and the well-trained deep learning model is not robust to abrupt noises such as RMB [1, 38], because these noises normally have orders of magnitude higher than cardiac-related vibrations, drowning out subtle features and ruining forward propagation of the deep neural network [39]. Therefore, the **second challenge** is to design the signal model that considers fine-grained cardiac features within a single cardiac cycle and designs a robust single-cycle ECG generation module against abrupt noises.

1.3.3 The Third Challenge

After the modeling and recovery of single-cycle ECG pieces, a follow-up issue is to model and realize the robust long-term ECG recovery. Although the model for the domain transformation between single-cycle radar/ECG pair has been proposed in [2], the long-term ECG recovery might be misaligned with ground truth due to inaccurate peak-to-peak interval (PPI) estimation, deteriorating the recovery quality even if the morphological features are well-recovered. Therefore, the **third challenge** is to model and generate the long-term ECG recovery from radar signal, and the recovery process should be robust against noises.

1.3.4 The Fourth Challenge

According to the literature, radar-based ECG recovery is only realized by deep-learning-based methods, because the domain transformation is extremely complex to be modeled mathematically while such transformation can be learned by deep learning model due to the great nonlinear mapping ability [1]. Similar to other research fields involved with deep learning, radar-based ECG recovery also asks for numerous radar signals to train the deep learning model with synchronous ECG ground truths [30, 40, 41]. According to previous research, the performance of the deep learning model degrades heavily after reducing 30% of the training data even after applying proper data augmentation techniques [42], causing difficulties for the deployment in new scenarios due to the demand for hours of ECG collection [2]. However, the method for reducing dependence on data quantity is rarely investigated for radar-based ECG recovery, and all the deep-learning-based ECG recovery models are trained in a supervised manner with large dataset containing 3 – 32 hours of synchronous radar-ECG pairs [1, 30, 37]. In this case, the **fourth challenge** is to reduce the dependency on large-scale datasets and develop appropriate transfer learning or data augmentation methods to alleviate data scarcity, especially for the deployment in new scenarios with limited data.

1.4 Contributions and Thesis Outline

The proposed solutions for the aforementioned challenges are elaborated in Chapter 3, 4, 5 and 6 with a general outline shown in Figure 1.2, and the main contributions of this thesis can be concluded as:

- Chapter 3 proposes a cardio-focusing and -tracking (CFT) algorithm based on derivative-free optimization (DFO) to find the cardio-focused (CF) point by iteratively evaluating the potential points in a discontinuous objective space, with a universal signal template designed to adaptively assess the signal SNR as costs.
- Chapter 4 designs a signal model to describe the fine-grained cardiac feature sensed by radar, enabling further domain transformation between cardiac mechanical and electrical activities. Based on the proposed signal model, an ODE-embedded module called single-cycle ECG generator (SCEG) is designed to realize the domain transformation by parameterizing the radar signal into sparse representations and hence generate the morphological ECG features as references to resist noise.
- Chapter 5 investigates the noise robustness in radar-based ECG recovery against constant or abrupt noise by modeling the cardiac domain transformation as three tasks. Then, an end-to-end multi-task learning (MTL) framework named radarODE-MTL is accordingly proposed to realize these tasks and leverage adjacent cardiac cycles to compensate for the distorted one. To assist the MTL model training, a novel optimization strategy called eccentric gradient alignment (EGA) is proposed for updating shared parameters in the MTL neural network, aiming to balance the intrinsic difficulty across tasks during network training and also prevent the negative transfer phenomenon.
- Chapter 6 tries to alleviate data scarcity in deep learning model training from two perspectives. Firstly, a transfer learning framework RFcardi is proposed following a self-supervised learning (SSL) paradigm to effectively learn the latent

representations from radar signals by leveraging an appropriate pre-text task. Secondly, a data augmentation method called Horcrux is proposed to expand the diversity of the limited training dataset without distorting the intrinsic time consistency hidden in the radar inputs.

In addition, Chapter 2 provides the necessary background for radar-based vital sign monitoring, and Chapter 7 concludes the entire thesis with several promising research directions for future investigation.

Chapter 2

Background and Literature Review

In this chapter, the necessary background for radar theory and radar-based vital sign monitoring is first provided for a better understanding of the proposed methods in the later chapters. In addition, representative signal processing and deep learning methods for cardiac feature extraction are also introduced to provide a comprehensive review of the development over the last decade.

2.1 Theoretical Background

2.1.1 Radar Types

Different radar systems transmit different types of waveforms as shown in Figure 1.1(b). For example, continuous wave (CW) radar uses continuous wave with a fixed frequency, FMCW radar uses continuous wave with linearly increased frequency, and impulse-radio ultra wide band (IR-UWB) radar uses pulses with wide frequency bandwidth. The phase components of the transmitted signals for CW and FMCW radar are modulated by the displacement composed by respiration, heartbeat and all kinds of noises in a non-linear manner as proved in [33, 43]. Then, the phase variation hidden in the raw received signal can be revealed using phase unwrapping techniques such as arctangent demodulation and extended differentiate and cross-multiply algorithm [14, 18]. For IR-UWB radar, the cardiac features are embedded in the propa-

gation time delay of the echo signal [18].

Different radar types require different architectures and are suitable for different tasks. For example, CW radar has a simple architecture and adopts the fundamental baseband signal-processing methods, but the range information cannot be extracted from the received signals due to the lack of modulation [44]. FMCW radar outperforms CW radar by leveraging the frequency modulation techniques, improving the signal-to-noise ratio (SNR) and providing the capability of range detection to further isolate the signal reflected only from the chest region [17]. Different from the continuous waveform used in CW and FMCW radar, IR-UWB radar emits widely spaced pulses with very short duration (e.g., 0.1 – 2 ns) [45]. Therefore, IR-UWB radar is more power-efficient than FMCW radar but normally cannot ensure a high SNR and range resolution [18]. In literature, IR-UWB radar is normally used for through-the-wall or long-distance monitoring [46], but the complex radar architecture (e.g., requiring internal delay calibration [44]) and signal-processing algorithms (e.g., harmonic rejection [47]) limit the relative research [43].

In addition to the different radar types, radar operating frequency (carrier frequency) is another crucial parameter affecting cardiac monitoring quality because the frequency is inversely related to the beamwidth for antennas with the same diameter [48], enabling the radar system with high operating frequency using narrow beamwidth to enhance directivity [49]. Obeid *et al.* [50] found out that high operating frequency provides a large phase difference caused by heart vibration and improves the sensitivity of cardiac monitoring. The researchers in [49] claimed that the radar with a high operating frequency, especially the mmWave range (30 – 300 GHz), can achieve a high range resolution, good noise-robustness and small antenna size.

According to the review of the trends in radar usage for the recent decade, CW radar was the most popular type before 2020 due to its simple architecture; FMCW radar receives a growing concern since 2015 because the recent-released commercial FMCW radar platforms reduce the knowledge required for designing or setting up a radar system [51]; IR-UWB radar is less popular than the other two types due to its complex radar architecture and signal-processing algorithms [44,47]. For the trend in operating frequency selection, early studies all focus on the low-frequency band for

the simplicity of radar architecture design and baseband signal-processing algorithms, whereas the recent researchers are steering toward using mmWave radar (especially 60 or 77 GHz commercial radar platform) for good performance. In summary, the FMCW with high operating frequency is becoming the mainstream for radar-based cardiac monitoring due to the balance between the performance and complexity, and the complexity for settling the radar platform is also significantly reduced due to the emergency of commercial radar.

2.1.2 Radar Theory for Vital Sign Monitoring

Cardiac Signal Extraction from CW Radar

The vanilla signal model for radar-based cardiac monitoring (e.g., heart rate monitoring) using CW radar starts from the transmitted signal expressed as

$$s_t(t) = A_t \cdot \cos(2\pi f t + \theta(t)) \quad (2.1)$$

where A_t and f are the amplitude and carrier frequency of the transmitted signal, and $\theta(t)$ is the phase noise from the signal generator with respect to time t [52]. In the ideal case, the radar signal is only reflected by a human at a fixed distance d_0 with a varying chest displacement as $x(t)$, and the received signal after propagation time $T_p(t)$ can be derived as

$$s_r(t) = A_r \cdot \cos(2\pi f(t - T_p(t)) + \theta(t - T_p(t))) \quad (2.2)$$

with

$$\begin{aligned} T_p(t) &= \frac{2d(t)}{c} \\ d(t) &= d_0 + x(t) \end{aligned} \quad (2.3)$$

where A_r is the amplitude of the received signal, c is the light speed and $2d(t)$ represents the round trip distance of the signal between the transmitter and receiver.

Then, the received signal can be expanded as

$$s_r(t) = A_r \cdot \cos(2\pi ft - \frac{4\pi d_0}{\lambda} - \frac{4\pi x(t)}{\lambda} + \theta(t - \frac{2d_0}{c} - \frac{2x(t)}{c})) \quad (2.4)$$

where λ is the wavelength that equals to $\frac{c}{f}$. According to [52,53], it is safe to eliminate changes in amplitude and phase noise term because the chest displacement is much less than the fixed distance (i.e., $x(t) \ll d_0$). Therefore, the approximate received signal is

$$s_r(t) \approx \cos(2\pi ft - \frac{4\pi d_0}{\lambda} - \frac{4\pi x(t)}{\lambda} + \theta(t - \frac{2d_0}{c})) \quad (2.5)$$

The received signal $s_r(t)$ will then pass a local oscillator with a low-pass filter to remove the frequency term, and the resultant baseband signal is

$$s_b(t) = \cos(\theta_d + \frac{4\pi x(t)}{\lambda} + \Delta\theta(t)) \quad (2.6)$$

with

$$\begin{aligned} \theta_d &= \frac{4\pi d_0}{\lambda} + \theta_0 \\ \Delta\theta(t) &= \theta(t) - \theta(t - \frac{2d_0}{c}) \end{aligned} \quad (2.7)$$

where θ_d , θ_0 and $\Delta\theta(t)$ are phase shifts affected by different factors such as d_0 , signal mixer and antenna, and can be set as constant [53]. Then, the phase signal unwrapped from the baseband signal is obtained as

$$\phi(t) = \theta_d + \frac{4\pi x(t)}{\lambda} + \Delta\theta(t) \quad (2.8)$$

Finally, the vanilla signal model derived above shows that the chest displacement $x(t)$ is involved in the phase variation of the baseband signal as

$$\Delta\phi(t) = \frac{4\pi x(t)}{\lambda} \quad (2.9)$$

The follow-up researchers have proposed various techniques to improve the accuracy of the unwrapped phase signal variation in (2.9). For example, the in-phase/quadrature

modulation is proposed to solve the null point issue [52]; the differentiate and cross-multiply algorithm is designed to avoid discontinuity in the unwrapped phase signal [4]. In addition, chest displacement $x(t)$ is a mixture of cardiac activities, respiration and noises (e.g., random body movement (RBM) [33, 54], multi-path or multi-person interference [16, 55]). Therefore, enormous advanced algorithms are proposed to decompose cardiac information from $x(t)$, as have been reviewed in [4].

Cardiac Signal Extraction from FMCW Radar

FMCW could provide advanced features that allows the extraction of the signal reflected from certain points. In the literature, FMCW radar has been widely used in nowadays glsmmWave sensing to measure the range, velocity and angle of arrival (AoA) of the objects appearing in the field of view [56], with three critical concepts that configure the transmitted waveform:

- **Chirp** is the minimum component in the FMCW signal with microsecond-level duration and is often called fast time. The waveform of a single chirp is a sinusoidal signal with frequency that changes linearly over time, with the key characteristics designated by start frequency, bandwidth and chirp duration to get range information of the object.
- **Frame** is a collection of multiple chirps that forms a complete observation window to get the velocity information based on the range bins extracted from chirps and is often referred to as slow time.
- **Virtual antenna array (channel)** is a commonly used concept in multiple-input and multiple-out (MIMO) radar systems and is able to realize complex modulations or beamforming [26]. However, this study mainly leverages the phase difference across antenna channels to estimate the AoA of the objects.

The popular commercial radar platforms have provided a convenient interface for radar configuration, signal modulation and demodulation [57], and the reflected signal

from a give point $E = (x, y, z)$ in 3D space can be expressed as:

$$R(E, t) = \sum_{v=1}^V \sum_{c=1}^C \sum_{n=1}^N s_{v,c,n}(t) \cdot e^{j2\pi \frac{2k \cdot d(E, v)}{\text{light speed}} n} \underbrace{e^{j2\pi \frac{2 \cdot d(E, v)}{\lambda}}}_{\text{phase term } \phi} \quad (2.10)$$

where (x, y, z) represents the (horizontal, radial, vertical) axis, V is the number of virtual antenna channels, C means the number of chirps within one frame, N is the total sample points within one chirp, $s_{v,c,n}(t)$ denotes the original received signal, k is the slope of frequency raising, λ means wavelength and $d(E, v)$ represents the distance between point E and virtual antenna v [1]. In FMCW processing for vital sign monitoring, the time sample t corresponds to one frame instead of the sample point n , and the signal from different chirps c and antenna channels v will be accumulated to improve SNR [24].

The interested term in (2.10) is the variation of distance $d(E, v)$, because it represents the displacements caused by respiration and heartbeat (without considering any other noise). Therefore, the chest region displacement $h(E, t)$ can be unwrapped from phase variation $\Delta\phi$ as

$$h(E, t) = \frac{\lambda \Delta\phi}{4\pi} \quad (2.11)$$

At last, some common noises, such as respiration and thermal noise, can be easily removed using a band-pass filter and differentiator to make sure that the final $h(E, t)$ mostly contains cardiac-related features from point E .

2.1.3 Radar-Based ECG Monitoring as a Domain Transformation Problem

Coarse cardiac monitoring only aims to detect a single heartbeat within one cardiac cycle, while fine-grained cardiac monitoring requires recovering subtle cardiac activities within one cardiac cycle. For example, Figure 2.1 shows the typical radar and SCG signal waveform within a single cardiac cycle that describes the cardiac mechanical activities, such as aortic valve opening/closure (AO/AC) and mitral valve opening/closure (MO/MC) [23, 34]. These mechanical activities are muscle contractions

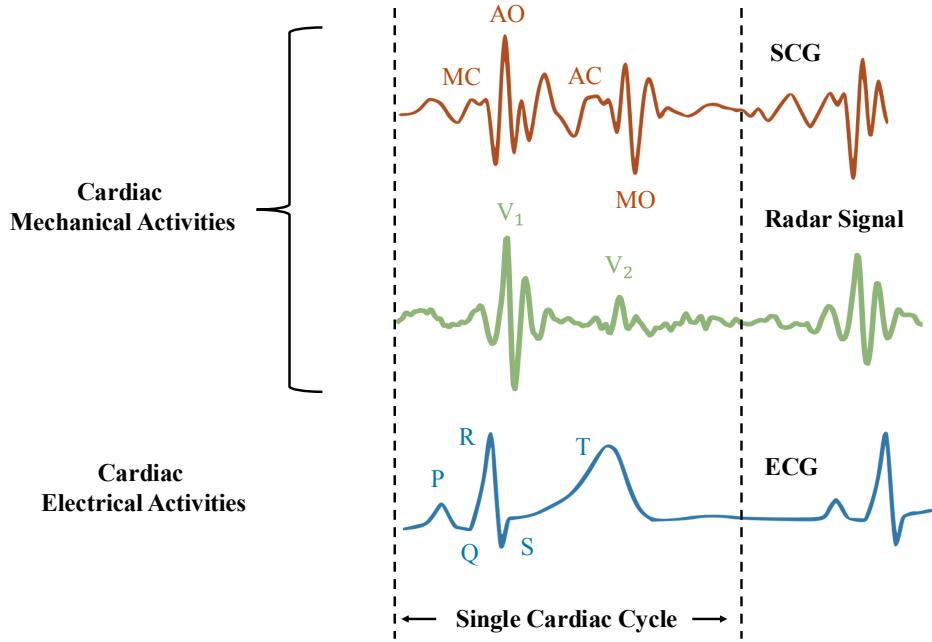


Figure 2.1: Relationships between cardiac mechanical and electrical activities within the same cardiac cycle.

stimulated by cardiac electrical events, such as P-wave, QRS-complex and P-wave in the ECG signal as shown in Figure 2.1. Therefore, radar-based ECG recovery is actually a domain transformation problem that translates cardiac mechanical activities into electrical activities and realizes fine-grained vital sign monitoring in a contactless manner. In the literature, the domain transformation is only realized by deep-learning-based methods, while a common issue of the deep learning model is not robust against large-scale noise (e.g., RBM) as reported by many previous studies [30, 35–37]. However, there is no existing work that investigates the noise-robustness of the deep learning model itself, and this thesis is motivated to develop a noise-robust deep learning model to realize the domain transformation in the presence of body movement noise.

2.2 Literature Review for Cardiac Feature Extraction

2.2.1 Spectrum-based Methods

Spectrum represents the transformation of signal from time-domain to frequency-domain and can reveal the main frequency components of the signal. Normally, the respiration rate (RR)) and HR frequency at rest are in the range of 0.1–0.5 Hz and 1–1.6 Hz respectively [58], making the spectrum-based method the most straightforward way to extract cardiac features within a certain frequency range. The spectrum-based methods usually share similar ideas: (a) a piece of unwrapped phase signal taken from the dataset only shows the periodic displacement induced by strong respiration instead of subtle cardiac activities as shown in Figure 2.2a; (b) the raw signal is then transformed into a spectrum using time-frequency transformation methods (e.g., fast Fourier transform (FFT), continuous wavelet transform (CWT)) as shown in Figure 2.2b with corresponding peaks labelled; (c) for the low-noise scenario, after filtering the frequency components of RR, RR harmonic and other high-frequency noise, the remaining dominant peak of the spectrum represents the HR frequency; (d) for the noisy scenario, the researchers could use several techniques, such as noise-cancellation filter or differentiator, to suppress the noise or enhance the HR frequency component on the spectrum.

Evaluations

The spectrum-based methods are matured in the signal-processing area and can be implemented directly on the hardware with limited computational resources while achieving good real-time performance [59, 60]. In fact, several commercial data-capture boards for raw radar signal processing have already embedded these algorithms on field-programmable gate array (FPGA) [18]. For the disadvantages, the spectrum-based methods are not sensitive to the abrupt HRV, because multiple cardiac cycles are involved in a single segment truncated by the time window, but only one HR value can be estimated from each segment [61]. Another problem caused by

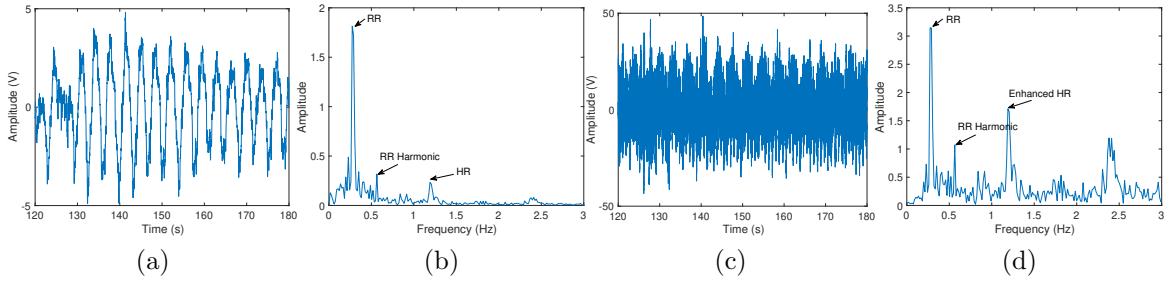


Figure 2.2: Illustration for spectrum-based methods: (a) raw radar signal; (b) spectrum obtained for raw radar signal from FFT, with RR, RR harmonic and HR peaks labelled; (c) raw radar signal after differentiation; (d) spectrum obtained for differentiated raw radar signal from FFT, with enhanced HR peak labelled.

the truncation is that each segment may not contain integral multiples of the cardiac cycles, causing the shift of the estimated HR around the true HR [62]. Lastly, various noises (such as RR harmonics and slight RBM [63]) may have the frequency components falling into the HR frequency band and hence dominate the HR components [64]. Although it is possible to extract the cardiac features using HR harmonics from the high-frequency spectrum (e.g., 1.5 – 5 Hz) to avoid the effect of strong RR harmonic components [65], the HR harmonics sometimes might be too weak and easily distorted by unexpected noises (e.g., car vibrations [66]).

2.2.2 Periodicity-based Methods

Periodicity-based methods are based on the natural periodicity of the cardiac features and leverage probability model (e.g., hidden Markov model) [13], auto/cross-correlation [61] or template determined by cardiac morphology [67] to identify the periodical patterns obscured by noise in radar signals [68]. This type of methods could identify a single heartbeat by either measuring the similarity between template and raw signals or predicting the next heartbeat based on a probabilistic model. In addition, periodicity-based methods further enable the fine cardiac event segmentation which is valuable for clinical diagnosis [17]. Lastly, periodicity-based methods are naturally immune to the noises that do not show any periodic feature.

Evaluations

The above mentioned advantages of periodicity-based methods rely on certain prior knowledge (e.g., peak-to-peak interval (PPI) or ECG waveform) to design the template/model, but such template/model may be unsuitable for the monitoring of new participants and hence require certain calibration during actual monitoring [69]. Furthermore, the optimal template/model with periodicity features learned from datasets may not fit the diverse cardiac features of other individuals [54]. In other words, the researchers need to balance the ability to detect rare cardiac events (e.g., heart diseases) against the ability to provide accurate HR estimations for most scenarios.

2.2.3 Blind Source Separation Methods

Blind source separation (BSS) is a classic problem originally in the audio processing field to separate a group of signals (e.g., voices of different people) from the mixed signal received by microphone(s) [70]. For radar-based cardiac monitoring, the received signal is also a nonlinear mixing of various sources such as RR, RBM, cardiac vibration and multi-path interference [33]. The methods introduced in this subsection aim to decompose the mixed signal and extract the cardiac vibration signal according to different criteria:

- Multiple signal classification (MUSIC): orthogonal signal spaces for cardiac signal and noises;
- Independent component analysis (ICA): statistical independence between cardiac signal and noises;
- Empirical mode decomposition (EMD): time-scale features in mixed signals;
- Variational mode decomposition (VMD) and sparse signal reconstruction/recovery (SSR): natural sparsity of heartbeats.

Evaluations

BSS methods do not require prior knowledge regarding cardiac events and can decompose the mixed signal according to specific characteristics. In addition, instead of modelling the heartbeat signal as a summation of a finite number of periodic sinusoidal signals [46], most BSS methods assume that the source signals are not simple sinusoidal functions but with narrow bandwidth, enabling the non-linear decomposition of the mixed signal.

For the drawbacks, an ideal decomposition result relies on the careful selection of the parameters (e.g., P for MUSIC, WGN for EMD, K for VMD), whereas these parameters are obtained empirically and require a re-selection after altering the monitoring environment. The inappropriate selection of parameters can cause the mode mixing or over-decomposition issue, making it impossible to extract HR from the decomposed signals [71]. Lastly, it is still a challenge to select the proper decomposed signals for HR estimation because there are multiple decomposed signals falling into the heartbeat frequency band [72].

2.2.4 Deep Learning Method

Deep learning belongs to machine learning based on artificial neural networks, but has multiple hidden layers formed by numerous neurons to achieve great performance. Each neuron receives a set of weighted inputs from other neurons and outputs the result through an activation function to introduce the non-linearity into the deep learning model [1]. Normally, the deep learning models with multiple hidden layers require complex structures and training methods, but are suitable for modelling the complex cardiac monitoring problem because the target signals are non-linearly modulated with other noises [33]. In addition, to enable the deep learning model to learn the latent information (high-level features) buried in the signal, a massive amount of data mostly needs to be fed into the deep learning model to find the optimal weights after iterative training.

In the literature, radar-based ECG waveform recovery has been achieved based on various deep-learning architectures, such as convolutional neural network (CNN)) [1,

[2], long short-term memory (LSTM) network [73], and Transformer [1, 35]. However, the noise robustness of the deep-learning framework is rarely investigated in the literature, especially for the RBM noise that is inevitable in contactless monitoring and has orders of magnitude larger than cardiac activities. The existing work either discarded the data during the RBM [36] or reported the heavy distortion as the future work [1]. Additionally, the existing deep-learning methods are also blamed for being purely data-driven as a black box and the transformation between cardiac mechanical and electrical activities lacks the theoretical explanation [2].

Evaluations

Due to the powerful fitting capability of the deep learning model, deep learning methods can model complex and non-linear projections and hence produce fine-grained cardiac signals (e.g., ECG) compared with the algorithms in other categories. In addition, with a special network design (e.g., LSTM network), the deep learning model can memorize long-term dependency from the dataset, making stable long-term cardiac monitoring possible. Lastly, compared with the other three categories of methods, the deep learning methods can potentially resist different kinds of noises by introducing these noises during the dataset collection [33].

The outstanding performance of deep learning methods generally relies on the training with a large-scale and full-featured dataset [74], whereas the well-trained deep learning model may fail to infer on the data out of distribution [69]. For example, the deep learning model trained on the dataset for normal people cannot perform accurate cardiac monitoring for patients with different cardiac features. Similarly, the deep learning model trained on the dataset collected for multi-path effect mitigation cannot resist the RBM noise. Therefore, the collected dataset should contain various real-world scenarios, increasing the complexity of establishing a full-featured dataset. Additionally, it is still a challenge to implement complex deep learning models on compact and lightweight devices with real-time monitoring ability, because effective deep learning models normally require ample computational resources and large memory.

Chapter 3

Efficient High-SNR Signal Acquisition

This chapter focuses on the first stage of radar-based ECG recovery to acquire high-SNR radar signals with ample cardiac features, because the quality of the radar inputs highly affects the fidelity of ECG recovery [1, 2]. Previous methods either require traversing a large search space to extract high-SNR signal or applying signal accumulation to suppress constant noises and enhance cardiac features, causing massive demand on computational resources and only applicable to short-range monitoring. In this chapter, a novel algorithm called cardio-focusing and -tracking (CFT) is proposed to find the cardio-focused (CF) point by iteratively evaluating the potential points in a discontinuous objective space, and the CFT algorithm is tested on the sitting subjects to provide radar measurements with better SNR compared with existing methods.

3.1 Introduction

Radio detection and ranging (radar) system is originally designed for military detection of large aircraft by emitting electromagnetic waves and evaluating the reflections. The follow-up research has investigated the civilian use of radar systems for

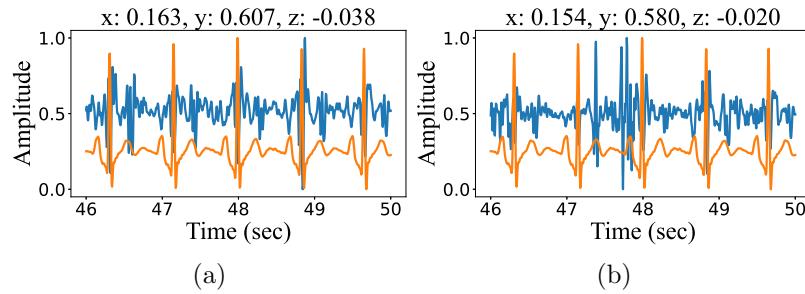


Figure 3.1: Challenges for high-SNR signal collection: (a) and (b) Radar signals with high and low SNR for adjacent points with a distance of 0.03m.

contactless sensing in various scenarios, such as autonomous driving [75] and human monitoring [4]. Over the past decade, radar-based sensing has been empowered by deep neural networks to process non-stationary reflected signals or high-dimensional data, enabling versatile applications to replace contact- or visual-based measurement for convenience or privacy concerns (e.g., vital sign monitoring [24], gesture recognition [76], fault diagnosis [77]).

Radar-based vital sign monitoring, as a popular branch of radar-based sensing, has been explored for decades to measure heart rate or respiration rate in a contactless manner [4], and some further studies leverage the deep neural network to realize domain transformation from cardiac mechanical activities (i.e., heartbeat) to electrical activities (i.e., ECG), providing a fine-grained cardiac measurement for wellness monitoring or clinical diagnosis [1, 2, 30, 35, 40, 42, 78]. In the literature, radar-based ECG recovery is only realized by deep-learning-based methods, because the domain transformation is extremely complex to be modeled mathematically while such transformation can be learned by deep learning model due to the great nonlinear mapping ability [1].

Based on the discussion Chapter 1.3.1, it is still a challenge to precisely locate and track the cardiac location during data collection to efficiently extract high-SNR radar signal. Although, it is natural to think the high-SNR radar signal can be searched in a constrained space by optimization, there is no appropriate method to assess the signal SNR in terms of cardiac features contained, and the objective space

is actually highly discontinuous with adjacent points may revealing totally different SNR as shown in Figure 3.1(a) and 3.1(b), restricting the application of common gradient-based optimization algorithms. To address these issues, the contributions of this study can be listed as:

- A CFT algorithm is proposed based on derivative-free optimization (DFO) to find the glscf point by iteratively evaluating the potential points in a discontinuous objective space, with a universal signal template designed to adaptively assess the signal SNR as costs.
- The proposed CFT algorithm has been validated on sitting subjects in various scenarios and could provide radar measurements with better SNR compared with existing methods.

The rest of this chapter is organized as follows. Section 3.2 elaborates the proposed CFT algorithm, with the experimental settings and results shown in Section 3.3 and 3.4. The final conclusion is shown in Section 3.5.

3.2 Methodology

3.2.1 Overview of CFT Algorithm for ECG Recovery

The pipeline of using CFT algorithm is shown in Figure 3.2 with three steps:

- The received radar signal will be converted into a standard format in terms of chirp, frame and virtual antenna channel to obtain the general location of the subject, as shown in Figure 3.2(a).
- The rough location acts as the initial state for the CFT algorithm, and the points within a constrained space will be evaluated to find the red CF point with best SNR as shown in Figure 3.2(b).
- Signals extracted from the ten best points will be converted into spectrograms to train the deep learning model designed in [40] and generate corresponding ECG recoveries, as shown in Figure 3.2(c).

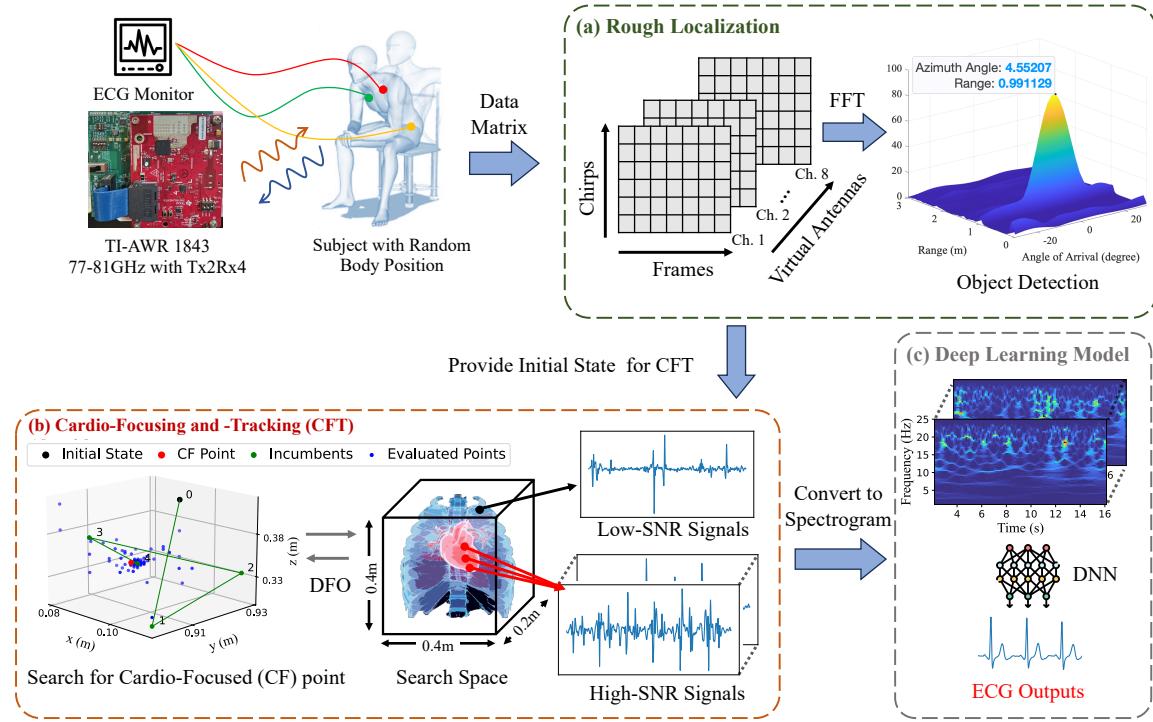


Figure 3.2: Overview of using CFT for radar-based ECG recovery: (a) Rough localization of human body; (b) Use CFT to find CF point and extract high-SNR radar signals; (c) Deep learning for ECG recovery.

In addition, TI-AWR 1843 radar operated at 77 Ghz with 2 Tx and 4 Rx will be used for data collection, enabling 8 virtual antenna channels for high-quality signal extraction. The detailed scenario descriptions and radar configurations will be provided in Section 3.3.1.

3.2.2 Rough Localization

The received radar signal is first formatted as a standard data matrix in terms of different chirps, frames and virtual channels to provide measurement of range, velocity and AoA, respectively. For the current research level of radar-based vital sign monitoring, the subjects are all quasi-static without velocity, and only the range-angle (RA) map will be calculated using fast Fourier transform (FFT) as shown in Fig-

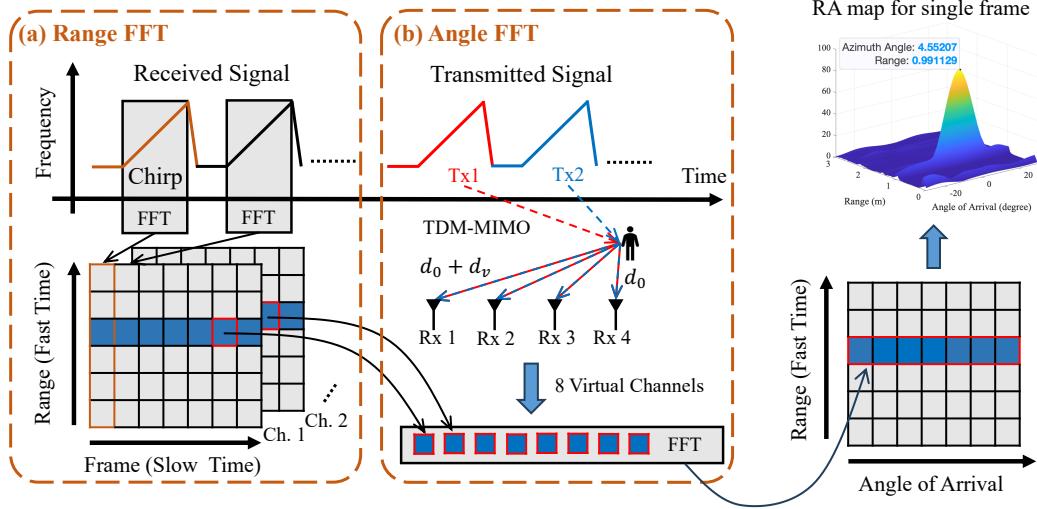


Figure 3.3: Procedures for obtaining RA map: (a) Range FFT for chirps along fast time; (b) Angle FFT along virtual channels.

ure 3.2(a), with a detailed illustration of signal waveform and processing shown in Figure 3.3.

Range FFT

According to (2.10), the signal propagation after transmitting introduces a constant phase shift ϕ_s in the received signal and is expressed as

$$\phi_s = \frac{4\pi d_0}{\lambda} \quad (3.1)$$

with d_0 representing the distance between radar and human body. Therefore, the distance d_0 can be extracted from each received signal along fast time using FFT as shown in Figure 3.3(a), and the updated data matrix now reveals the range information, i.e., a static object denoted as blue along slow time axis.

Angle FFT

The ability of AoA detection relies on the MIMO system using time division multiplexing (TDM-MIMO), with multiple Tx alternately transmitting chirp signals and the corresponding reflections can be distinguished during receiving as shown in Figure 3.3(b). Due to the physical distance varies for different Tx/Rx combinations (i.e., Tx2Rx4 creates 8 virtual channels), an extra propagation $\Delta\phi_v$ delay will be introduced as:

$$\Delta\phi_v = \frac{4\pi d_v}{\lambda} \quad (3.2)$$

$$d_v = l \sin(\theta)$$

where d_v represents the extra propagation distance, l means the distance between adjacent antenna channels and θ is the incident angle. Similar to range FFT, the phase differences across different channels can be used to extract AoA information for each range bin by performing FFT along the channel axis, as shown in red squares in Figure 3.3(b).

After combining the FFT results for all chirps and channels, the final RA map for the current time sample (frame) can be obtained as shown in Figure 3.3. The same procedure can be repeated along the slow time axis to get the rough human body location for all the time samples, but this study only requires the location obtained from the very first frame as the initial point E_0 for CFT algorithm.

3.2.3 Cardio-focusing and -tracking (CFT) Algorithm

The radar signal for any point can be extracted following (2.10) and (2.11), and the search progress from E_0 to the best point E_b (i.e., CF point with high SNR) requires: (a) an appropriate metric to assess whether the radar signal contains wanted cardiac features; (b) an optimization method that is applicable to the discontinuous objective space based on the assessed SNR values as costs.

Template Design for Assessing SNR

An explicit SNR can be calculated with the known “clean” signal, while the “clean” signal for vital signs normally reveals two prominent vibrations corresponding to the ventricular contraction and relaxation [2], as shown in Figure 3.4(a). However, considering the vibrations may have subtle differences due to different scenarios or radar configurations (e.g., noise figure and sampling frequency), a universal template h_m is designed in this study to fit the envelope of the radar signal as:

$$h_m(t) = a_1 \exp\left(-\frac{(t - b_1)^2}{2c_1^2}\right) + a_2 \exp\left(-\frac{(t - b_2)^2}{2c_2^2}\right) \quad (3.3)$$

with a_1, a_2 controlling the amplitudes of the peaks, b_1, b_2 determining the centers of the peaks and c_1, c_2 adjusting the width of the peaks. In practice, a_1 and b_1 will be fixed based on the dominant peaks detected as the red points in Figure 3.4(b), and other parameters are left to be determined as a simple curve fitting problem. Finally, the mean square error (MSE) between the radar signal envelope and the synthetic template is reckoned to be an assessment of signal SNR as shown in Figure 3.4(b), because fewer components could fit the designed template for low-SNR radar signal without obvious cardiac features, as shown in Figure 3.4(c) and 3.4(d).

Derivative-free Optimization (DFO)

The MSE values obtained from template matching for the radar pieces extracted from point E will be used as costs $\mathcal{F}(E)$ in searching CF point, but the traditional gradient-based optimization method is not applicable because there is no explicit cost function. Therefore, the CFT algorithm is developed in a derivative-free manner based on coordinate search (CS) algorithm [79], to asymptotically approach the CF point.

The definition of the DFO problem is formulated as:

$$E_b = \arg \min_{E \in \mathbb{R}^n} \{\mathcal{F}(E) : E \in \Omega\} \quad (3.4)$$

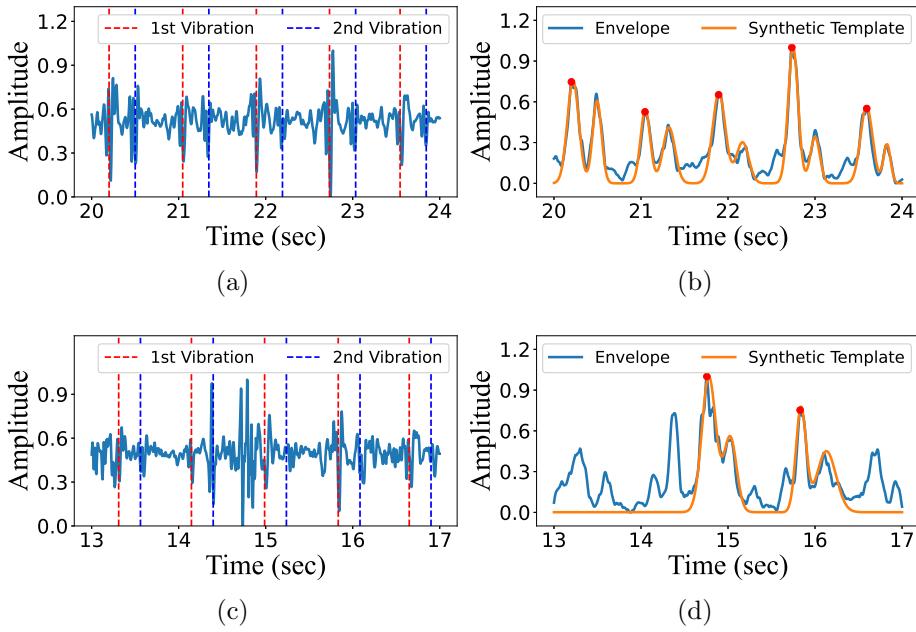


Figure 3.4: Template for assessing SNR: (a) High-SNR radar signal; (b) Extracted signal envelope with the synthetic template; (c) (a) Low-SNR radar signal, (d) Extracted signal envelope with the synthetic template.

with Ω representing a user-defined constrained n -dimensional search space near initial point E_0 as shown in Figure 3.2(b), and the cost of points out of the constraint will be set as $\mathcal{F}(E \notin \Omega) = \infty$. During each iteration k , many trial points E_k within the constraint will be evaluated to find the incumbent points E_i as the temporary best point for the next iteration.

To perform a derivative-free search, the traditional CS algorithm starts from the initialization of grids G_k :

$$G_k := \{E_k + \gamma_k D\} \subset \mathbb{R}^n \quad (3.5)$$

where $\gamma_k > 0$ is the grid size parameter and D contains several vectors p for possible searching directions, as shown in Figure 3.5(a). The local convergence of CS is ensured by dense search directions D and a refined grid size γ_k to find better E_i compared

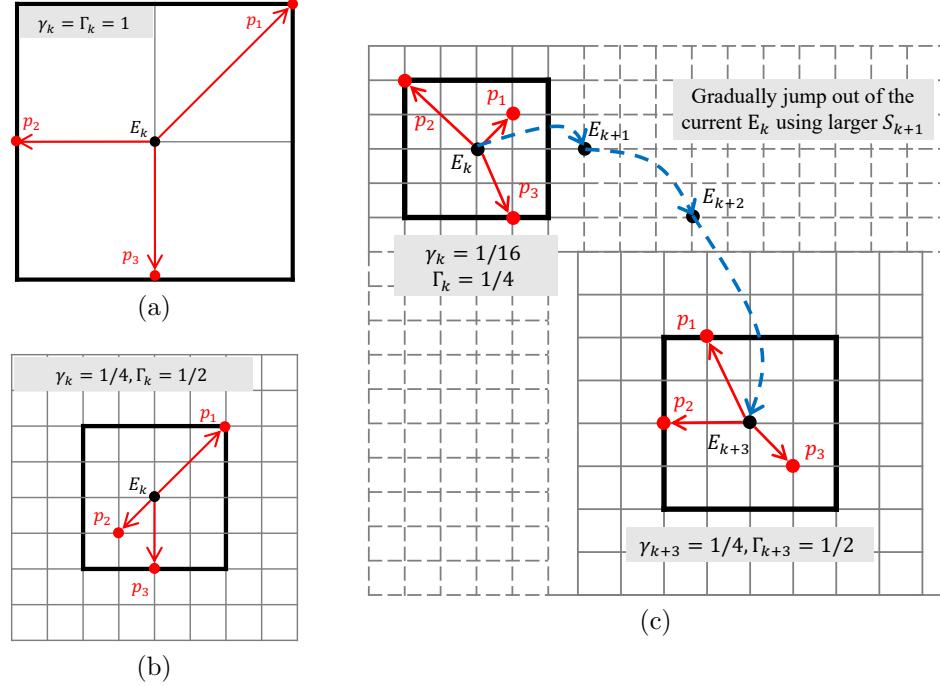


Figure 3.5: Illustration of the CFT algorithm with bold line wrapping the search region S_k : (a) Equality between γ and Γ (same as in CS algorithm); (b) Large Γ_k with refined γ_k , providing more potential points to be evaluated; (c) Jump out of the local minimum by adjusting Γ_k and γ_k .

with current E_k [79]. However, the highly discontinuous objective space for radar-monitored vital signs may have numerous local minima that distract the optimization algorithm, i.e., the signal SNR of the adjacent points might be very different, as shown in Figure 3.1(a) and 3.1(b).

To jump out of the potential local minimum, CFT algorithm is proposed by introducing search region S_k to restrict the possible search directions p , alleviating the difficulty of searching in numerous dense grid points and allowing the adjustment of search region and grid size iteratively to break the local minimum. The detailed procedures of CFT are shown in Algorithm 1, with an illustration of \mathbb{R}^2 space shown in Figure 3.5.

In CFT, the grids G_k is still expressed as in (3.5) and the newly introduced search

region S_k is expressed as:

$$S_k := \{E \in G_k : \|E - E_k\|_\infty \leq \Gamma_k a\} \quad (3.6)$$

with $a = \max \{\|a'\|_\infty : a' \in D\}$ and Γ_k as the size parameter for the search region. An intuitive interpretation of S_k is the point set that contains grid points inside and on the boundary of the bold line controlled by Γ_k , as shown in Figure 3.5(b).

Based on the well-constructed grids G_k and search region S_k , the remaining CFT algorithm is performed with searching and resizing stages:

Searching: The searching stage simply asks for the evaluation of $\mathcal{F}(E)$ on a subset of grids G_k based on any sampling algorithm (e.g., Latin hypercube sampling [79]), as indicated in line 6 in Algorithm 1.

Resizing: The resizing stage depends on the result of searching stage:

- If a new incumbent point E_i is found with better SNR, the search region will be doubled as $\Gamma_{k+1} = 2\Gamma_k$ (line 8), and the grid size will be empirically set as $\gamma_{k+1} = \min(\Gamma_k, \Gamma_k^2)$ (line 16), enabling to search in broader space in the next iteration.
- If there is no better point than the current E_k on the current grids G_k , another searching stage will be performed only within the search region S_k (line 9). Then, If a better point E_i is found, Γ_{k+1} and γ_{k+1} is obtained as above (line 11), otherwise, the search region will be halved as $\Gamma_{k+1} = \Gamma_k/2$ (line 14) for a finer search with $\gamma_{k+1} = \min(\Gamma_k, \Gamma_k^2)$ (line 16).

The searching step enables the finding of better points E_i in a broad space, and the resizing step either refines the grid if the current γ_k is not enough or enlarges the search space when stalling at the local minimum, as shown in Figure 3.5(c). Finally, the CFT algorithm will be terminated after achieving a desired SNR_d or iteration limit k_{max} .

The visualization of the CFT algorithm in Figure 3.2(b) shows that initial iterations search in a large space, and the algorithm could jump out of the green local minima to find the red CF point within the fine blue grid points. In addition, the

Algorithm 1 CFT Algorithm

```

1: Input:  $E_0$ ,  $\text{SNR}_d$ ,  $k_{max}$ 
2: Output:  $E_b$ ,  $\text{SNR}_b$ 
   OBJECTIVE:
3:  $E_b = \arg \min_{E \in \mathbb{R}^n} \{\mathcal{F}(E) : E \in \Omega\}$ 
4: Initialize  $k = 0$ ,  $\Gamma_k = \gamma_k = 1$ ,  $\text{SNR}_b = \mathcal{F}(E_k)$ 
5: while  $\text{SNR}_b > \text{SNR}_d$  and  $k < k_{max}$  do
6:   if  $\mathcal{F}(E) < \text{SNR}_b$  for some  $E \in G_k$  then
7:      $E_{k+1} \leftarrow E$ ,  $\text{SNR}_b \leftarrow \mathcal{F}(E)$ 
8:      $\Gamma_{k+1} \leftarrow 2\Gamma_k$ 
9:   else if  $\mathcal{F}(E) < \text{SNR}_b$  for some  $E \in S_k$  then
10:     $E_{k+1} \leftarrow E$ ,  $\text{SNR}_b \leftarrow \mathcal{F}(E)$ 
11:     $\Gamma_{k+1} \leftarrow 2\Gamma_k$ 
12:   else
13:      $E_{k+1} \leftarrow E_k$ 
14:      $\Gamma_{k+1} \leftarrow \Gamma_k/2$ 
15:   end if
16:    $\gamma_{k+1} \leftarrow \min(\Gamma_k, \Gamma_k^2)$ 
17:    $k \leftarrow k + 1$ ,  $E_b \leftarrow E_{k+1}$ 
18: end while

```

tracking of the CF points along time can be naturally realized by repeating Algorithm 1 with previous E_b as the new E_0 , and the SNR evaluated on the previous point might have already achieved SNR_d due to the quasi-static human body, saving a huge amount of time for calculating useless channel information for filtering or clustering [1, 24, 30].

3.3 Details of Experiment and Dataset

3.3.1 Dataset Collection and Preparation

The dataset contains a total of 80-minute synchronous radar-ECG pairs collected for 5 healthy subjects (3 men, 2 women) in 2 indoor scenarios as shown in Figure 3.6. The subjects are asked to sit causally and are allowed to change postures during data



Figure 3.6: Indoor scenarios for data collection.

Table 3.1: Parameters for data collection interface

Parameter	Value	Parameter	Value
Start Frequency	77GHz	Frequency Slope	65MHz/ μ s
Idle Time	10 μ s	Tx Start Time	1 μ s
ADC Start Time	6 μ s	ADC Samples	256
Sample Rate	5000kbps	Ramp End Time	60 μ s
Start/End Chirp Tx	0/2	No. of Chirp Loops	2
No. of Frames	12000	Frame Periodicity	5ms

collection, and each data trial lasts for 1 minute. The distance between radar and human body varies from 0.5 – 1.2m, and a longer distance causes the decrease of signal SNR with a smaller portion of the space points containing useful cardiac features.

TI-AWR 1843 radar with 2 Tx and 4 Rx is used for data collection with 8 virtual antenna channels created [57], and the radar configurations are listed in Table 3.1 with the name provided in TI mmWave-Studio interface. The signal will be sampled as 200Hz, and only a band-pass filter from 0.5 to 50Hz and a differentiator are used for removing respiration noise because the radar signal extracted from CF points already has high SNR. Lastly, the ECG ground truth is collected using TI ADS1292, and the related ECG processing (e.g., smoothing and peak finding) is realized by NeuroKit2 python package [80].

3.3.2 Implementation Details

Parameters for CFT Algorithm

The constraint Ω for the CF point search is centered at the initial state E_0 with a range of $0.4 \times 0.2 \times 0.4$ m as illustrated in Figure 3.2(b). In addition, the initial grid and search region size should be adjusted to fit the real-life physical unit as $\Gamma_k = \gamma_k = 0.1$ m, and the size will be limited as $\Gamma_k \geq \gamma_k \geq 0.001$ m to prevent an exhaustive search within a meaningless small space. At last, SNR_d is set to 0.01 for the desired MSE between normalized synthetic template and signal envelope, and k_{max} is set to 100.

Deep Learning Model Training

The deep learning model adopts the same backbone, ECG decoder and hyperparameters as in our previous open-sourced work [40] coded in PyTorch and trained on NVIDIA RTX 4090 (24GB). The total training epoch is set to 100 with batch size 8, and a 5-fold cross-validation training strategy is adopted to split the dataset to make the most of the limited dataset while excluding the testing data from the training phase.

3.3.3 Methods for Comparison

The comparison is performed with the representative methods based on accumulation and clustering to extract high-SNR radar signal:

- De-ViMo [24] is proposed for heart rate monitoring and is based on the accumulation of signals from various dimensions (e.g., chirps, antennas, spatial points) to enhance cardiac features while mitigating noise. In addition, De-ViMo also improves the rough localization by identifying the peaks in micro-motion frequency bands instead of the entire FMCW bands.
- MMECG [1] requires the calculation of numerous points in 3D space and applies clustering algorithm to improve SNR. Then, a pattern-matching process is

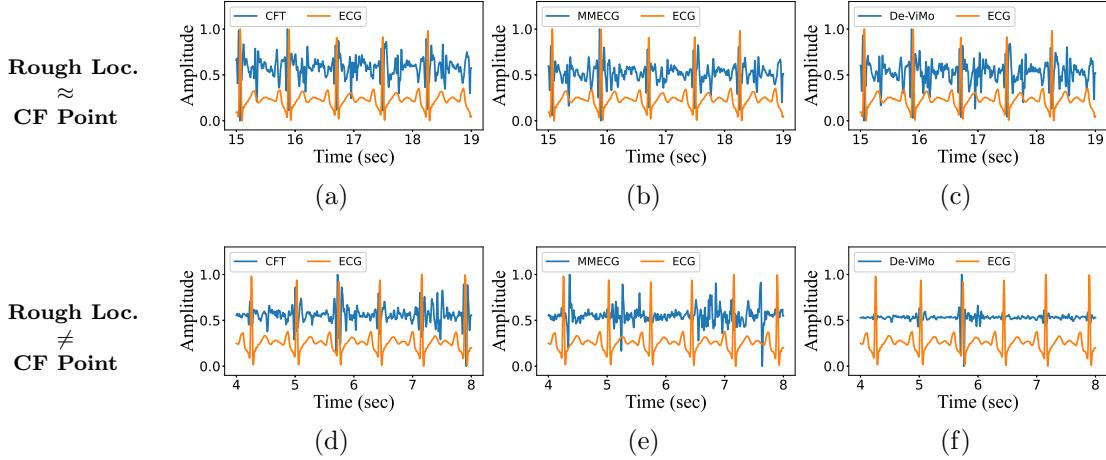


Figure 3.7: Visualization of the extracted radar signal for all methods: (a) - (c) If CF point is around rough body location; (d) - (f) If CF point is far from rough body location.

performed to learn the common pattern from the clustered result and select the best radar signal(s).

3.4 Experimental Results and Evaluations

3.4.1 Effectiveness of CFT Algorithm

The examples of the extracted radar signal for different methods are shown in Figure 3.7, illustrating that precise cardiac localization has a huge effect on the signal quality. For example, if the rough body location is around CF point, all three methods can obtain high-SNR signals with clear first and second vibrations using either space search (CFT, Figure 3.7(a)), clustering (MMECG, Figure 3.7(b)) or accumulation (De-ViMo, Figure 3.7(c)).

In contrast, only a few range bins will contain useful cardiac features if the rough body location is far from CF points, especially when increasing the monitoring range. Therefore, the signal accumulation may enhance the noises as shown in Figure 3.7(f)

while the signal clustering may also encounter a failure due to the lack of homogeneous cardiac signals as shown in Figure 3.7(e). However, The proposed CFT could precisely locate the CF point with good SNR subject to the designed signal template and DFO searching strategy, and the extracted radar signal still shows clear peaks as shown in Figure 3.7(d).

During the data collection of this study, the subjects are allowed to change postures to alleviate discomfort, with a resultant CF point deviation of several decimeters, while the rough location provided by FMCW signal processing is still unchanged. Therefore, the proposed CFT algorithm is essential because the posture change is inevitable, and a thorough evaluation in terms of different monitoring ranges will be performed in the next part.

3.4.2 Impact of Monitoring Range

To evaluate the performance of different methods when increasing the monitoring range, the quality of extracted radar signals for all trials is evaluated in terms of:

- Absolute Peak error between ECG R peaks and the dominant peaks for the first vibrations in radar signal.
- Missed detection rate (MDR) to count the cardiac cycles with no peak detected or with the absolute peak error larger than 150ms [1].

Figure 3.8(a) and 3.8(b) illustrate the peak error and MDR for all 80 trials with corresponding fitting curves indicating the mean peak error or MDR. All the methods show similar performance in the short range and experience certain degradation with respect to the increasing range. In particular, MMECG shows larger degradation and variance for longer-range cases because the rough localization based on FMCW signal processing cannot provide accurate cardiac location, and the resultant evaluated points may not capture useful information for clustering. In contrast, De-ViMo could provide a better cardiac location, and the accumulated results show better accuracy than MMECG, while the long-range monitoring still affects the quality because the accumulation is not robust to non-gaussian noise. At last, the proposed CFT could

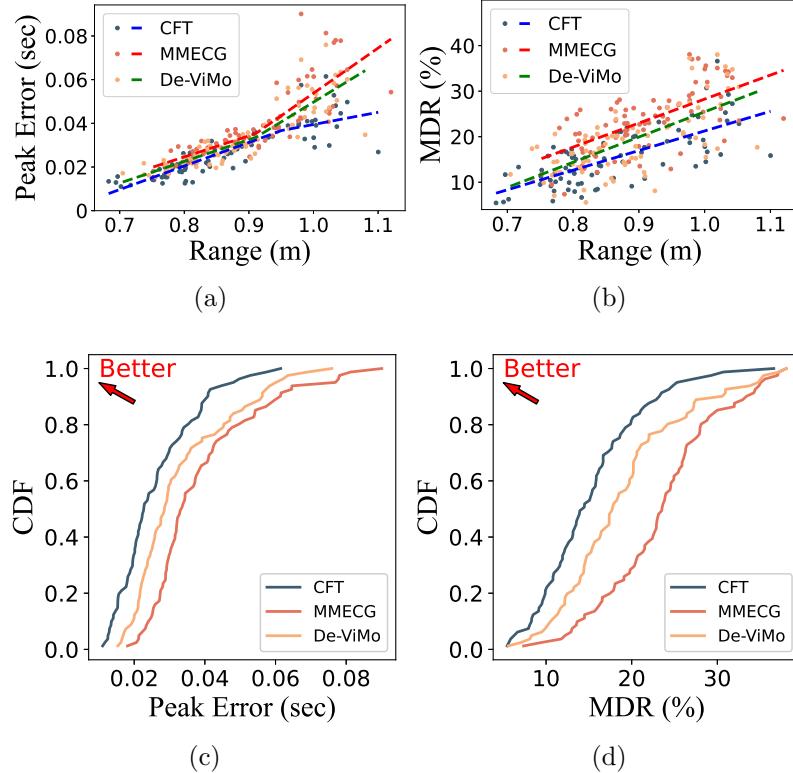


Figure 3.8: Illustration of performance in terms of peak error and MDR: (a) - (b) Scattered points with fitting curves along range axis; (c) - (d) CDF plots for all trials.

precisely focus on the CF point and keep tracking the high-SNR points during data collection, providing the best results with small variance for both peak error and MDR.

In addition, the cumulative distribution function (CDF) plots for all trials are shown in Figure 3.8(c) and 3.8(d). The proposed CFT algorithm achieves the best peak error with a median value of 0.022 sec, while DE-ViMo and MMECG have worse performances with larger median values of 0.028 sec and 0.033 sec, respectively. Similarly, the precise localization and tracking of CF point also reduces the MDR for CFT results with a median value of 14%, while DE-ViMo and MMECG may be affected by the accumulated noise or inaccurate cardiac localization with the median MDR of 17% and 23%, respectively.

Table 3.2: Performance of supervised ECG recovery

Methods	MSE ($\times 10^{-2}$) ↓	PCC ↑	Peak Error (ms) ↓	MDR ↓
MMECG [1]	0.93	80.36%	9.74	7.96%
De-ViMo [24]	0.88	83.83%	8.93	7.32%
CFT	0.82	85.47%	7.61	6.85%

3.4.3 Impact of Signal Quality on ECG Recovery

The signals extracted using different methods are used for supervised training to verify the impact of different input qualities on the ECG recovery task. The quality of the recovered ECG signal is assessed in terms of:

- The morphological accuracy is measured using MSE and Pearson correlation coefficient (PCC), with MSE sensitive to the peak deviation and PCC focusing on the similarity between the ECG patterns.
- The accurate recovery of ECG R peaks is crucial to coarse cardiac features calculation (e.g., heart rate variability) and is measured by absolute R peak error and MDR.

Table 3.2 shows the performance of the deep learning model trained with datasets yielded by different methods. The training based on CFT dataset achieves the best results on both morphological accuracy (MSE= 0.0082 and PCC= 85.47%) and R peak recovery (Peak Error= 7.61ms and MDR= 6.85%), because the high-SNR inputs provide accurate peak locations with minor noise that affects the ECG pattern generation, as shown in Figure 3.9(a) and 3.9(b).

In contrast, MMECG and De-ViMo cannot preserve the signal quality especially for long-distance cases, and the noisy inputs will prevent the deep learning mode from identifying the accurate position of ECG pieces, causing large peak error and MDR, as shown in Figure 3.9(c) and 3.9(d). It is worth noticing that poor signal SNR causes more degradation in peak error than morphological accuracy, because the ECG patterns share a similar shape and can be learned from other cardiac cycles, while the

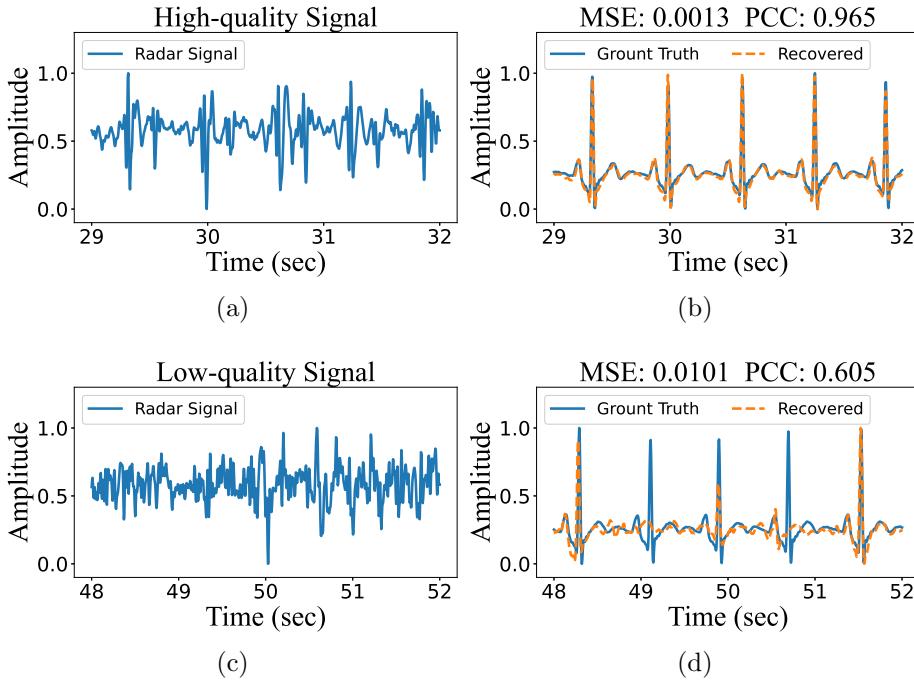


Figure 3.9: Impact of radar input quality on the final ECG recovery: (a) - (b) High-quality radar input and ECG recovery; (c) - (d) Low-quality radar input and ECG recovery.

peak recovery (detection) fully relies on the current radar input and can be ruined by noises.

3.5 Conclusions

This chapter investigates the efficient collection of high-SNR radar signals with ample cardiac features for deep-learning-based ECG recovery. Previous methods adopted signal accumulation or clustering to suppress the noises, while the rough localization based on FMCW radar cannot accurately reveal the chest region, requiring a time-consuming traverse among a 3D space for compensation. In this chapter, a novel CFT algorithm is proposed to dynamically articulate the points with the best SNR and could track the cardiac location over time if the subjects change posture. The

experiments performed in different scenarios prove the feasibility of the CFT algorithm in radar signal collection for ECG recovery, enabling a convenient deployment in new scenarios for future contactless wellness monitoring.

Chapter 4

Robust Single-Cycle ECG Generation

This chapter is based on the high-SNR radar measurements that contain ample cardiac features and aim to realize the robust ECG recovery to resist inevitable noises (e.g., random body movement (RBM) noise). Previous studies leverage the powerful fitting capability of the deep neural network (DNN) to model complex and non-linear projections from radar to ECG signal, but the inference results can be easily distorted under abrupt noises [1,37,78]. In this chapter, the traditional DNN is integrated with ordinary differential equation (ODE) with prior morphological features that constrain the ECG recovery, and a signal model is also designed to explain the critical features in the transformation from radar to ECG signals. The proposed ODE-embedded deep learning model is validated on the dataset with RBM noise, and the recovered ECG pieces still show high fidelity even when the subtle cardiac features are ruined by the noises with large amplitude.

4.1 Introduction

In the recent research about cardiac monitoring, the emergence of commercial radar platforms with high operating frequency encourages researchers to extract fine-grained

cardiac features (e.g., ECG and SCG) from the radar signal [4]. SCG signal is measured by the accelerometer mounted on the human chest to measure the mechanical vibrations produced by heartbeats, describing the fine-grained cardiac mechanical activities such as aortic/mitral valve opening/closing and isovolumetric contraction [22]. Although these vibrations are subtle, it is still reasonable to directly map the displacements detected by radar to each fine-grained cardiac mechanical activity using high-resolution radar as proved in [17]. However, radar-based SCG recovery is not widely investigated compared with ECG, because ECG provides more comprehensive information (e.g., atrial/ventricular depolarization [81]) for clinical diagnosis.

To reconstruct ECG from radar signal, the most straightforward approach is to directly sense the variation in the scattered electromagnetic field through frequency shift of mm-wave response, and the ECG signal can then be decoupled from the scattered electromagnetic field based on the dynamic model in the form of partial differential equations deduced from cardiac electrophysiology (i.e., ionic concentration in cardiac cells) [82]. However, the solutions of the entire model are extremely hard to obtain either numerically or analytically, and the constructed models will be changed with respect to different environments and noises due to the Green's function [83], causing difficulty in adapting the model in various real-life scenarios.

The second approach, which is also the most adopted approach, only uses radar to sense the chest region displacement through the reflected radar signal as in coarse cardiac monitoring, but then the researchers must deal with domain decoupling to transform the measured signal from the mechanical domain to the electrical domain to generate ECG measurement. Intuitively, it is reasonable that mechanical conduction and electrical conduction are highly correlated in describing cardiac activities because the electrical changes in cells trigger heart muscle contraction, whereas such a relationship is called excitation-contraction coupling in electrophysiology and is extremely hard to interpret or model by researchers without biological knowledge [23, 34].

In the literature, the existing studies all leverage deep learning methods to extract latent information from enormous radar/ECG pairs and try to learn domain transformation relying on the extraordinary non-linear mapping ability of the deep neural network [1, 30, 35–37]. Although these studies could successfully reconstruct the ECG

signal from radar, three drawbacks still need to be improved:

- There is no existing signal model with a compact form to describe the domain transformation for radar-based ECG reconstruction.
- The purely data-driven method could learn the domain transformation as a black box, but researchers can hardly intervene in the learning process to enhance the characteristic peaks of ECG or explain the intrinsic correspondence in domain transformation.
- The well-trained deep learning model is not robust to abrupt noises such as body movement [1,38], because these noises normally have orders of magnitude higher than cardiac-related vibrations, drowning out subtle features and ruining forward propagation of the deep neural network [39].

Inspired by the above discussions, this chapter aims to design a framework for radar-based ECG reconstruction with ordinary differential equations (ODEs) embedded to provide prior knowledge on domain transformation and resist abrupt noises. The contributions of this research can be concluded as:

- This study proposes an ODE-embedded framework called radarODE to produce robust long-term ECG recovery against abrupt noises with the aid of morphological ECG features as the reference.
- A signal model is designed to describe the fine-grained cardiac feature sensed by radar, enabling further domain transformation between cardiac mechanical and electrical activities, instead of using purely data-driven approaches without any explanation as in the literature.
- Based on the proposed signal model, an ODE-embedded module called single-cycle ECG generator (SCEG) is designed to realize the domain transformation by parameterizing the radar signal into sparse representations and hence generate the morphological ECG features as references to resist noise.

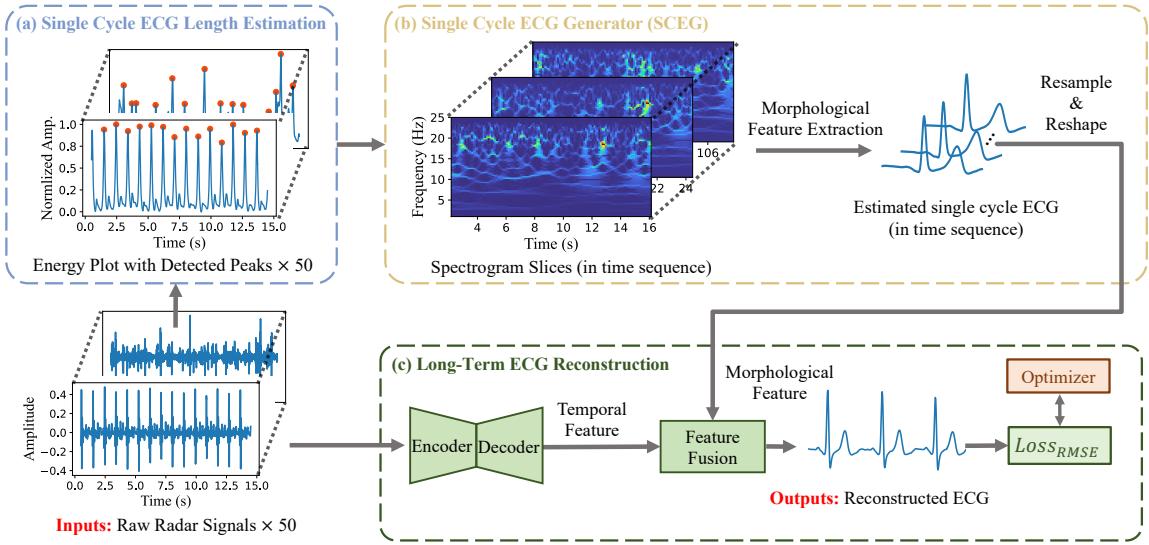


Figure 4.1: Overview of the radarODE framework for domain transformation: (a) Single-cycle ECG length (PPI) estimation; (b) Single-cycle ECG generator; (c) Long-term ECG reconstruction.

The rest of the chapter is organized as follows. Section 4.2 explains the proposed model for radar signal and the structure of radarODE framework. Section 4.4 and 4.3 introduce the public dataset used for validation in this research and then present the results obtained with corresponding comparisons and evaluations. Finally, Section 4.5 concludes this chapter.

4.2 Methodology

4.2.1 Overview

In order to realize a robust radar-to-ECG reconstruction, the proposed radarODE framework is designed based on the domain transformation within a single cardiac cycle to generate faithful ECG reference that aids the domain transformation in long-term ECG recovery. The inputs of radarODE are 50 synchronous radar signals that represent the measurements from 50 spatial points within the chest region [1], as shown in Figure 4.1, and the entire domain transformation is realized by three modules:

- The first module estimates the peak-to-peak interval (PPI) for consecutive heartbeats based on the proposed peak detection algorithm and provides information for single cardiac cycle segmentation, as shown in Figure 4.1(a).
- The second module first transforms the time-domain radar signal within a single cardiac cycle into a spectrogram using synchrosqueezed wavelet transform (SST). Then, the temporal and ODE decoder within SCEG will generate the faithful ECG pieces as morphological references while resisting noise disruption, as shown in Figure 4.1(b).
- The third module is used to fuse the extracted morphological and temporal features hidden in the original radar signal to generate the final ECG recovery, as shown in Figure 4.1(c).

4.2.2 Model for Radar Signal and Pre-Processing

Fine-Grained Model for Radar Signal

According to the literature in Chapter 2, the chest displacement $x(t)$ can be further decomposed as

$$x(t) = x_c(t) + x_r(t) + x_n(t) \quad (4.1)$$

where $x_c(t)$ means cardiac mechanical activities, $x_r(t)$ is respiration induced displacement and $x_n(t)$ is noise term. After the pre-processing, the respiration term has been filtered out, and the actual radar signal $\tilde{x}(t)$ provided in the dataset [1] can be expressed as

$$\tilde{x}(t) = x_c(t) + x_n(t) \quad (4.2)$$

Furthermore, the pre-processed radar signal $\tilde{x}(t)$ has two prominent vibrations v_1 and v_2 as shown in Figure 2.1, corresponding to the fine-grained cardiac mechanical activities shown in SCG. According to the previous work on SCG modeling, the heart muscle contraction has a pulsatile nature, and the bones/tissues in chest area introduce the extra damping into the pulse [20]. Inspired by the natural characteristics, the radar signal with two prominent vibrations measured in a single cardiac cycle is

innovatively modeled as the Gaussian pulses with certain central frequencies as

$$\tilde{x}(t) = v_1(t) + v_2(t) + x_n(t) \quad (4.3)$$

with

$$\begin{aligned} v_1 &= a_1 \cos(2\pi f_1 t) \exp\left(-\frac{(t - T_1)^2}{b_1^2}\right) \\ v_2 &= a_2 \cos(2\pi f_2 t) \exp\left(-\frac{(t - T_2)^2}{b_1^2}\right) \end{aligned} \quad (4.4)$$

where a_1, b_1 and a_2, b_2 jointly contribute to the amplitudes and length of the first and second prominent vibrations, f_1, f_2 are the corresponding central frequencies, T_1, T_2 determine when the vibrations happen, and $x_n(t)$ represents all the noises.

The aim of proposing the model in (4.3) is not to perform the curve fitting but to provide the explanation for the later radarODE design, because the positions of the prominent vibrations (T_1, T_2) are crucial to the precise reconstruction of ECG peaks using deep neural network.

Signal Pre-Processing with Synchrosqueezed Wavelet Transform (SST)

Based on the proposed radar signal model in (4.3), the next step is to enhance the prominent vibrations (i.e., T_1, T_2). Figure 2.1 shows that the high SNR radar signal could reveal prominent peaks of v_1 and v_2 in the time domain, but in most cases, these two peaks (especially v_2) could be ruined by noise. Therefore, this research decides to extract the time-frequency domain information from the spectrogram obtained by synchrosqueezed wavelet transform (SST) [84], and the two vibrations can then be localized by the SCEG module proposed later in Section 4.2.3.

SST evolves from continuous wavelet transform (CWT) but with concentrated energy distribution along the frequency axis, providing a sparser time-frequency representation with enhanced prominent vibrations compared with other tools such as short-time Fourier transform (STFT) and CWT.

The first step of SST is to calculate the CWT of radar signal $\tilde{x}(t)$ as

$$W_{\tilde{x}}(a, b) = \int \tilde{x}(t) a^{-1/2} \psi^* \left(\frac{t-b}{a} \right) dt \quad (4.5)$$

where ψ^* is the complex conjugate of the chosen mother wavelet, and a, b are the adjustable scaling and translation factors for the wavelet ψ to extract frequency- and time-domain information, respectively. In this research, the Morlet wavelet is selected as mother wavelet because it is widely used for vibration signal processing, especially for time-frequency localization [85].

The second step is to calculate the candidate instantaneous frequency for $W_{\tilde{x}}(a, b) \neq 0$ as

$$f_{\tilde{x}}(a, b) = -2\pi i (W_{\tilde{x}}(a, b))^{-1} \frac{\partial W_{\tilde{x}}(a, b)}{\partial b} \quad (4.6)$$

The final step is to concentrate the energy along the candidate instantaneous frequency as

$$T_{\tilde{x}}(2\pi f, b) = \int_{A(b)} W_{\tilde{x}}(a, b) a^{-3/2} \delta(2\pi f_{\tilde{x}}(a, b) - 2\pi f) df \quad (4.7)$$

where $A(b) = \{a; W_{\tilde{x}}(a, b) \neq 0\}$, and δ represents the Dirac-delta function in a distribution version to smoothly squeeze the spread-out energy into a narrow band around the instantaneous frequency [84].

The quality of the resultant spectrograms can be evaluated using power spectrogram entropy (PSE) [38], with a small value indicating that the energy is concentrated around a certain frequency. The calculated PSE for the spectrogram produced by STFT, CWT and SST is 0.94, 0.90 and 0.76 respectively, with the visualized results shown in Figure 4.2(b), 4.2(c) and 4.2(d). It is clear that the spectrogram obtained by STFT only shows the rough positions of each v_1 , while CWT gives a sharp position for both vibrations, but the energy is still spread out. By further concentrating the energy distribution, SST produces the spectrogram with a relatively clean background and sharp peaks for the vibrations, reducing the burden of the deep-learning-based SCEG module in extracting latent features (i.e., T_1, T_2).

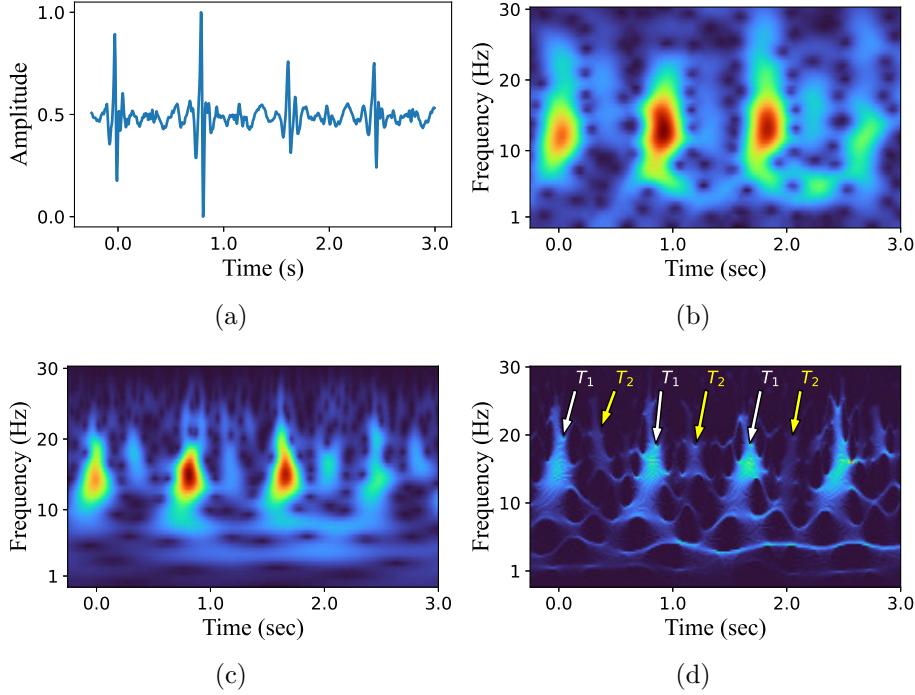


Figure 4.2: Spectrograms obtained from the same radar signal: (a) Radar signal $\tilde{x}(t)$; (b) STFT result; (c) CWT result; (d) SST result with T_1 , T_2 labeled, revealing concentrated energy around vibrations and clean background.

4.2.3 radarODE Framework Design

Single-cycle ECG Length Estimation

The first module of radarODE aims to estimate the length of each single-cycle ECG piece by calculating the interval between two consecutive heartbeats (i.e., PPI) from the energy plots \hat{x} of the radar signal \tilde{x} (omit (t) for simplicity) as shown in Figure 4.3(a). The energy plot is obtained by simply adding the spectrogram along frequency axis, but the peak detection results may not be promising due to low SNR signals as shown in Figure 4.3(b). Therefore, a new algorithm is proposed as in Algorithm 2 to eliminate the wrong detection obtained from 50 radar energy plots $X_{\mathcal{L}}$, with the length of each $\hat{x}_i \in X_{\mathcal{L}}$ equal to l .

The design of Algorithm 2 is based on the fact that the PPI for healthy people tends

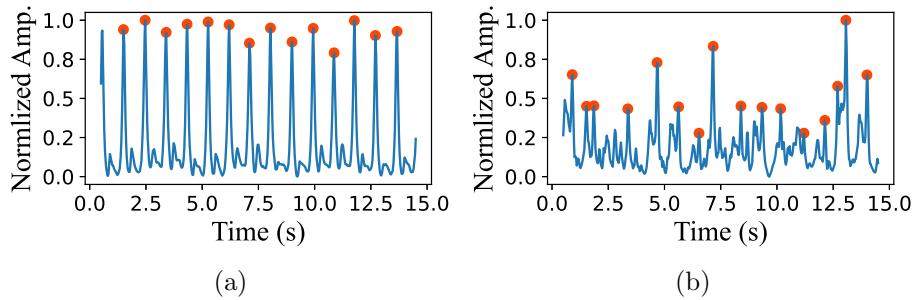


Figure 4.3: Energy plot of the synchronous radar signals with different detected peaks: (a) Energy plot with high SNR and correct detection; (b) Energy plot with low SNR and wrong detection.

to be unchanged in adjacent cardiac cycles. In this case, the long-term radar energy plots are firstly sliced into short segments as shown in the **INITIALIZATION** stage in Algorithm 2, and then the biopeaks algorithm implemented in NeuroKit2 [80] is used for detecting all the potential peaks P from each energy plot segment \hat{x}_i^j as:

$$P = \text{biopeaks}(\hat{x}_i^j) \quad (4.8)$$

Secondly, the resultant PPI_C obtained from Line 8-10 in Algorithm 2 contains potential estimated PPI from 50 radar energy plots, with the correct estimations as the majority. Therefore, the kernel density estimation (KDE) [86] is applied on the candidate set PPI_C to calculate the probability density of different PPI values as:

$$\text{KDE} : \hat{f}(p) = \frac{1}{nh} \sum_{c=1}^n K\left(\frac{p - PPI_{c \in C}}{h}\right) \quad (4.9)$$

where $\hat{f}(p)$ means the estimated probability density function at point p , n is the number of all the estimated PPI in PPI_C , K is the Gaussian kernel function and $h = n^{-1/5}$ is the bandwidth of the kernel. Lastly, the final PPI for the current segment is selected as the argument p when $\hat{f}(p)$ achieves the maximum as in Line 13 in Algorithm 2, and the long-term PPI estimation can be obtained step by step as Algorithm 2 terminated.

Algorithm 2 PPI Estimation

```

1: Input: Radar Energy Plots  $X_{\mathcal{L}} = \{\hat{x}_1, \hat{x}_2, \dots, \hat{x}_{50}\}$ , Segments Length  $l_{seg}$  and
   Step Length  $l_{step}$ 
2: Output: Estimated  $PPI$ 
   INITIALIZATION:
3:   - Let  $X_{\mathcal{S}} = \{X_{\mathcal{L}}^1, X_{\mathcal{L}}^2, \dots, X_{\mathcal{L}}^J\}$  be an ordered list of the segment lists sliced
      from  $X_{\mathcal{L}}$  with length  $l_{seg}$  and step  $l_{step}$ , where  $J = \frac{l-l_{seg}}{l_{step}}$ .
4:   - Let  $PPI \leftarrow \emptyset$ .
   MAIN ITERATION:
5:   for each segment list  $X_{\mathcal{L}}^j \in X_{\mathcal{S}}$  do
6:     - Let  $PPI_c \leftarrow \emptyset$  to save the candidate PPI obtained from each segment.
7:     for each segment  $\hat{x}_i^j \in X_{\mathcal{L}}^j$  do
8:       1) Apply biopeaks [80] on  $\hat{x}_i^j$  to get all the detected peaks  $P$  as in (4.8).
9:       2) Get PPI for the current radar signal segment using differentiation as
           $PPI_c \leftarrow diff(P)$ .
10:      3) Update  $PPI_c \leftarrow PPI_c \cup PPI_c$ .
11:    end for
12:    - Calculate the probability density function  $\hat{f}(p)$  for  $PPI_c$  using KDE as in
        (4.9).
13:    - Determine the final PPI for the current segment list and update the set as
         $PPI \leftarrow PPI \cup \arg \max_p \hat{f}(p)$ .
14: end for

```

Single-cycle ECG Generator (SCEG)

Based on the yielded PPI , the SST spectrogram can be sliced into segments corresponding to a single cardiac cycle, and the aim of the SCEG module is to reconstruct the ECG for each single cardiac cycle, hence realizing the transformation from mechanical to electrical domain. In general, the input of the SCEG is N segments of the SST plot within [1, 25] Hz with the size of $F \times T$ on frequency and time axis, and the output is the corresponding N ECG pieces with the same length T , as shown in Figure 4.4. In practice, the deep neural network only accepts the inputs/outputs of the same size. Therefore, the actual SST segment is centered at the current cardiac cycle and expands to 4 seconds, and the corresponding ECG ground truth is resampled to

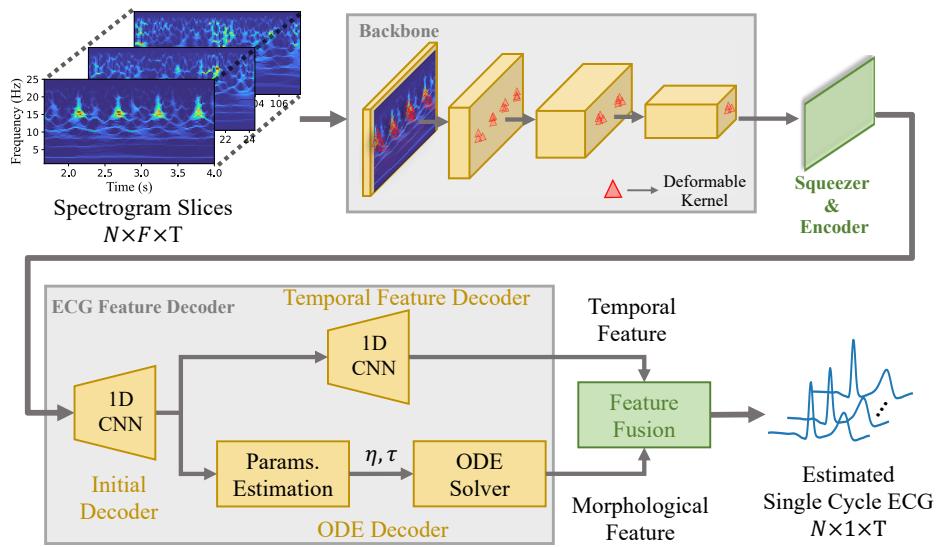


Figure 4.4: Architecture of SCEG with SST segments as input and single-cycle ECG pieces as output.

a fixed length of 200 for loss calculation.

For architecture design, the SCEG module adopts the popular backbone-encoder-decoder structure as verified by enormous image-related tasks [41, 87], with detailed parameters shown in Table 4.1. In addition, a feature fusion block is added after the decoder to fuse the temporal and morphological features and generate the final ECG reconstructions. The detailed implementation of each part in Table 4.1 with explanations can be elaborated as:

Table 4.1: Structure and Parameters for SCEG

Layers	Parameters $(C_{in}, C_{out}, K, S)^1$	Output Shape
Input SST		$N: \text{Batch Size}$
a. Backbone		
Residual Block	$(50, 128, (2, 1), (1, 1))$	$(N, 128, 72, 118)$
Downsample Block	$(128, 128, (3, 2), (2, 2))$	$(N, 128, 36, 60)$

Continued on next page

Table 4.1 – continued from previous page

Layers	Parameters $(C_{in}, C_{out}, K, S)^1$	Output Shape N : Batch Size
Residual Block	(128, 256, (2, 1), (1, 1))	(N , 256, 37, 60)
Downsample Block	(256, 256, (3, 2), (2, 2))	(N , 256, 19, 31)
Residual Block	(256, 512, (2, 1), (1, 1))	(N , 512, 20, 31)
Downsample Block	(512, 512, (3, 3), (2, 1))	(N , 512, 10, 31)
Residual Block	(512, 1024, (2, 1), (1, 1))	(N , 1024, 11, 31)
Downsample Block	(1024, 1024, (3, 3), (2, 1))	(N , 1024, 6, 31)
b. Squeeze&Encoder		
Conv2d	(1024, 1024, (6, 1), (1, 1))	(N , 1024, 31)
Transconv1d Block	(1024, 512, 5, 3)	(N , 512, 95)
Transconv1d Block	(512, 256, 5, 3)	(N , 256, 287)
Transconv1d Block	(512, 128, 5, 3)	(N , 128, 863)
c. ECG Feature Decoder		
Initial Decoder		
Conv1d Block	(128, 64, 7, 2)	(N , 64, 430)
Conv1d Block	(64, 32, 7, 2)	(N , 32, 213)
Conv1d Block	(32, 16, 7, 1)	(N , 16, 209)
Conv1d Block	(16, 8, 5, 1)	(N , 8, 207)
Temporal Feature Decoder		
Conv1d	(8, 4, 7, 1)	(N , 4, 203)
Conv1d	(4, 2, 5, 1)	(N , 2, 201)
Conv1d	(2, 1, 2, 1)	(N , 1, 200)
ODE Decoder		
Linear Block	(8 * 207, 512, -, -)	(N , 512)
Linear Block	(512, 128, -, -)	(N , 128)
Linear Block	(128, 32, -, -)	(N , 32)
Linear Block	(32, 16, -, -)	(N , 16)

Continued on next page

Table 4.1 – continued from previous page

Layers	Parameters $(C_{in}, C_{out}, K, S)^1$	Output Shape N : Batch Size
ODE Solver	–	$(N, 1, 200)$
d. Feature Fusion		
Feature Multiply	–	$(N, 1, 200)$
Stack	–	$(N, 1, 4, 200)$
Conv2d Block	$(1, 16, (5, 5), (1, 2))$	$(N, 16, 2, 98)$
Conv2d Block	$(16, 32, (3, 3), (1, 2))$	$(N, 32, 2, 48)$
Conv2d Block	$(32, 64, (3, 3), (2, 2))$	$(N, 64, 1, 23)$
Transconv1d Block	$(64, 32, 5, 2)$	$(N, 32, 52)$
Transconv1d Block	$(32, 16, 5, 2)$	$(N, 16, 106)$
Transconv1d Block	$(16, 8, 3, 2)$	$(N, 8, 10, 211)$
Transconv1d	$(8, 4, 6, 1)$	$(N, 4, 206)$
Transconv1d	$(4, 2, 5, 1)$	$(N, 2, 202)$
Transconv1d	$(2, 1, 3, 1)$	$(N, 1, 200)$
Output single-cycle ECG piece		$(N, 1, 200)$

1. C_{in} : Input channel, C_{out} : Output channel, K : Kernel size, S : Stride

- a. **Backbone:** Backbone is typically used as the first block to extract both low-level (e.g., color, edge) and high-level features (e.g., presence of specific pattern) from the input images. In the context of this research, the backbone is expected to localize the vibrations v_1, v_2 revealed as periodically appeared bright triangles within the range of [1, 25] Hz on SST plots, providing latent information of T_1, T_2 for the further ECG reconstruction.

According to the literature, ResNet with residual blocks is widely used as the backbone for feature extraction in many fields [88], and this research will use a similar structure with 4 layers of residual blocks and downsample blocks as shown in Figure 4.5, with the key parameters listed in Table 4.1. In addition, the traditional 2D convolution is all replaced by deformable 2D convolution [89] with deformable

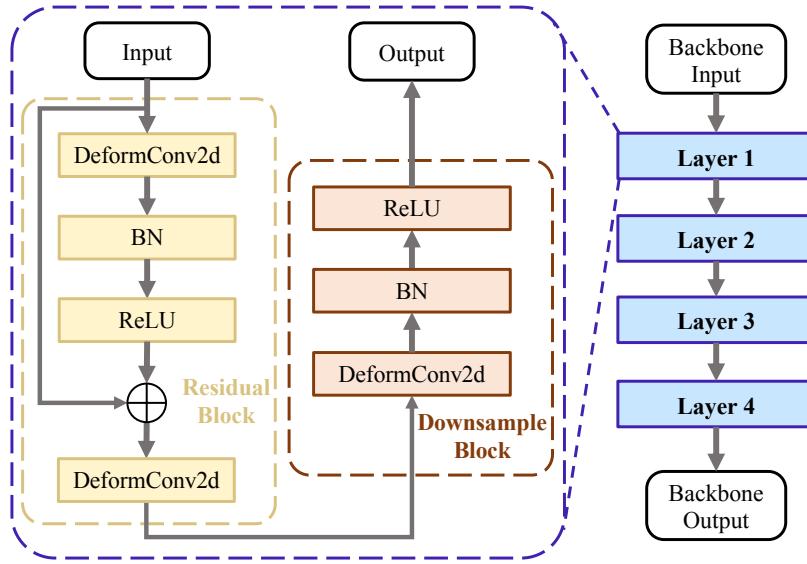


Figure 4.5: Structure of backbone composed of residual block and downsample block: DeformConv2d means deformable 2D convolution, BN is batch normalization, and ReLU is the rectified linear unit activation function.

kernels (instead of square kernels) to fit the irregular shape of the target patterns, as shown in red triangles in Figure 4.4.

- b. **Squeezer and Encoder:** The output feature map from the backbone should involve both frequency- and time-domain features, and the squeezer is simply used to squeeze the frequency-domain feature using 2D convolution (Conv2d) and output the feature map with the temporal feature only. Then, the encoder assembled by three 1D transposed convolution (Transconv1d) blocks is used to further extract the temporal feature, with each block comprising Transconv1d, BN and ReLU. Finally, the output feature map from the encoder should contain latent information about v_1, v_2 (i.e., (T_1, T_2)).
- c. **ECG Feature Decoder:** The decoder is an essential part of SCEG to extract the temporal and morphological features separately from the latent information in the SST feature map, as shown in two branches in Figure 4.4. At first, an initial decoder is shared by the latter two decoders and comprises four 1D convolution (Conv1d)

Table 4.2: Default values for η

$\eta \setminus \mathcal{F}$	e_P	e_Q	e_R	e_S	e_T
a_{e_f}	5	-100	480	-120	8
b_{e_f}	0.25	0.1	0.1	0.1	0.4
θ_{e_f}	$\frac{-15\pi}{180}$	$\frac{25\pi}{180}$	$\frac{40\pi}{180}$	$\frac{60\pi}{180}$	$\frac{135\pi}{180}$

blocks with Conv1d, BN, and ReLU inside. Similarly, the temporal feature decoder is assembled by three Conv1d, and the output feature should contain prominent peaks at the position of T_1, T_2 . However, these two peaks may still have deviations from the peaks in ECG ground truth, because obviously the mechanical vibrations v_1, v_2 lag behind QRS-complex and T-peaks, as shown in Figure 2.1. Another problem with the temporal feature decoder is that the ECG pieces have an entirely different shape with radar measurements, and the decoder needs to ‘memorize’ the unique pattern of ECG. Although the previous work shows the deep neural network could learn the patterns after training, the whole process of ECG reconstruction lacks supervision and is vulnerable to noise in radar signals [1, 90].

In this case, the ODE decoder is designed as a branch to assist the transformation between cardiac mechanical and electrical activities. The main obstacles in modeling such domain transformation are the lack of (a) A compact model for ECG signal and (b) A corresponding explanation between parameters in describing radar signal and ECG signal, e.g., what is the relationship between R peak and v_1 in Figure 2.1. In this work, the aforementioned two obstacles can be solved from the following perspectives:

- The shape of the ECG piece can be modeled morphologically using ODEs in a compact form without any biological/chemical knowledge [91].
- The measurements of mechanical activities generally lag behind those of electrical activities with a short time delay τ [23, 34].

Inspired by the above facts, the ODE decoder is designed as a parameter estimation part and an ODE solver as shown in Figure 4.4, and the solution of the ODEs

will be shifted to the left with time τ . In this manner, the latent information in describing radar signals is first transformed to the parameters for the ECG signal, and then the ODE solver will generate morphological-prior to accelerate the convergence of the model training process and provide extra robustness against noises.

To be specific, the parameters estimation part contains four linear blocks (Linear Layer, BN, Tanh) to project the latent space yielded by the initial decoder into parameters η , τ , and η will be sent to an ODE solver to solve a 3D trajectory denoted by (x, y, z) as

$$\begin{cases} \frac{dx}{dt} = \alpha(x, y)x - \omega y \\ \frac{dy}{dt} = \alpha(x, y)y + \omega x \\ \frac{dz}{dt} = -\sum_{e_f \in \mathcal{F}} a_{e_f} \Delta \theta_{e_f}(x, y) e^{-\Delta \theta_{e_f}(x, y)^2 / 2b_{e_f}^2} - z \end{cases} \quad (4.10)$$

with

$$\begin{aligned} \alpha(x, y) &= 1 - \sqrt{x^2 + y^2} \\ \Delta \theta_{e_f}(x, y) &= (\theta(x, y) - \theta_{e_f}) \mod 2\pi \\ \theta(x, y) &= \text{atan } 2(y, x) \in [-\pi, \pi] \\ e_f &\in \mathcal{F} = \{e_P, e_Q, e_R, e_S, e_T\} \end{aligned} \quad (4.11)$$

where \mathcal{F} represents five characteristic peaks (PQRST) in a single-cycle ECG signal, and the whole ODEs can be interpreted as manipulating each peak along a unit circle by varying the value of $\eta = \{a_{e_f}, b_{e_f}, \theta_{e_f}\}$ to adjust corresponding amplitude, width and position of each peak. After specifying all 15 parameters η (3 for each peak) and the initial conditions of (x, y, z) , the value of z can be solved by the ODE solver using the Euler method to get the final single-cycle ECG signal as the morphological feature. In practice, the default values for η are provided in advance as in Table 4.2, and the estimated parameters within the range of $[-1, 1]$ are used to scale the default values.

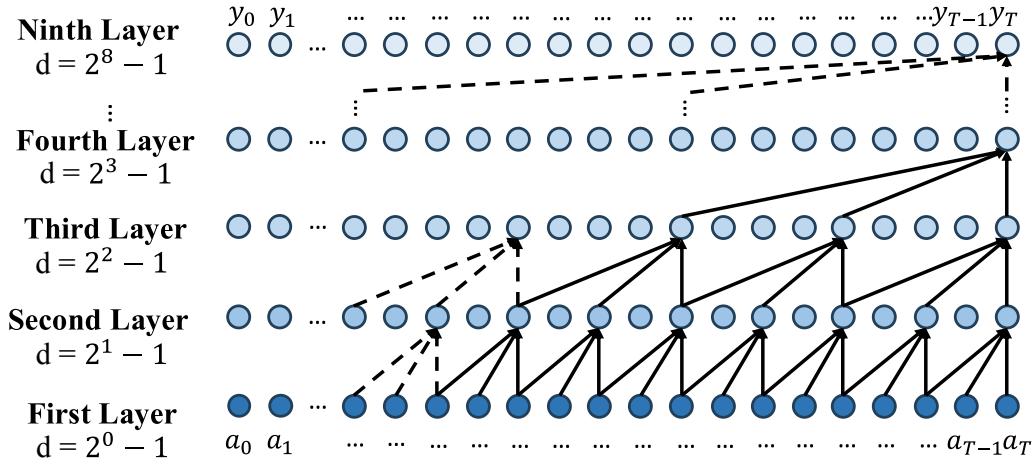


Figure 4.6: The structure of the 9-layer TCN with dilation factor $D = 2$ and kernel size $K = 3$.

- d. **Feature Fusion:** The feature fusion module could leverage respective advantages of the morphological and temporal features and generate the final ECG signal for loss calculation, because the morphological feature only focuses on five peaks to provide a rough shape of ECG with calibrated peaks, and the temporal feature reserves all the other feature neglected in (4.10) to help the final reconstruction [5]. Therefore, two features are first fused together by multiplication (Mul.) into one and then stacked four times by itself. Then, the stacked feature is encoded and decoded as in Table 4.1 to produce the final single-cycle ECG piece.

The last step of SCEG is to resample all the ECG pieces generated after the feature fusion part with respect to the previous PPI estimation, and the resampled ECG pieces are concatenated in a time sequence to form the final morphological reference for long-term ECG reconstruction.

Long-Term ECG Reconstruction

The long-term ECG reconstruction network adopts a similar encoder-decoder-fusion structure with SCEG as shown in Figure 4.1(c), with the detailed structure and

Table 4.3: Structure and Parameters for Long-term ECG Reconstruction

Layers	Parameters (C_{in} , C_{out} , K , S) ¹	Output Shape N : Batch Size
Input raw radar Signal		(N , 50, 800)
a. Encoder		
Residual Block	(50, 128, 5, 1)	(N , 128, 800)
Downsample Block	(128, 128, 5, 2)	(N , 128, 400)
Residual Block	(128, 256, 5, 1)	(N , 256, 400)
Downsample Block	(256, 256, 5, 2)	(N , 256, 200)
Residual Block	(256, 512, 5, 1)	(N , 512, 200)
Downsample Block	(512, 512, 5, 2)	(N , 512, 100)
b. Decoder		
Transconv1d Block	(512, 128, 5, 2)	(N , 128, 200)
Transconv1d Block	(128, 16, 5, 2)	(N , 16, 400)
Transconv1d Block	(16, 1, 5, 2)	(N , 1, 800)
c. Feature Fusion (TCN)		
Feature Stack ²	—	(N , 2, 800)
Dilated Conv1d $\times 9$	$K = 3$, $D = 2$	(N , 1, 800)
Output long-term ECG		(N , 1, 800)

1. C_{in} : Input channel, C_{out} : Output channel, K : Kernel size, S : Stride
 2. Stack with the morphological feature as in Figure 4.1(c).

parameters shown in Table 4.3 and Figure 4.6. The encoder takes 50 series of 4-sec time-domain radar signals as input and is composed of three groups of residual and downsample blocks with the deformable 2D convolution in Figure 4.5 replaced by 1D convolution. Then, the decoder is realized by three Transconv1d blocks to generate the temporal feature.

The temporal feature generated from decoder in Table 4.3 will be stacked with morphological feature as illustrated in Figure 4.1(c), and these two features act as two channels for later feature fusion by the temporal convolutional network (TCN). In general, TCN adopts dilation 1D convolution (Dilated Conv1d) to process multi-channel data structure using expanded receptive field with gaps between elements as shown in Figure 4.6, and the feature fusion is achieved during channel reduction as

a common technique in traditional convolution neural network [85]. Specifically, the solid line in Figure 4.6 shows the connection of a 4-layer TCN with the gaps for each layer l as $d = D^{l-1} - 1$, and the output y_T is predicted based on a receptive field of 15 input feature points $\{a_{T-14}, \dots, a_T\}$. In this chapter, 9-layer Dilated Conv1d is adopted with dilation factor $D = 2$ and kernel size $K = 3$, and the receptive field is 511 to make the most of contextual information contained in the temporal and morphological features.

4.3 Dataset and Implementation Details

4.3.1 Hardware and Environment Settings for Data Collection

The public dataset can be requested from [1] and is collected by the TI AWR-1843 radar with 77 GHz start frequency and 3.8 GHz bandwidth, providing good SNR and resolution to detect subtle vibrations that fit the proposed signal model and framework. To realize the 3D beamforming to extract cardiac features from real 3D space, 3 transmitters (Tx) and 4 receivers (Rx) are enabled with time division multiplexing multi-input multi-output (TDM-MIMO) applied in chirp transmitting and receiving.

The data collection is performed for subjects lying on the bed with quasi-static status to ensure good SNR with the least RBM noise. In addition, radar is placed right above the human chest region in a range of 0.4 – 0.5m with minor propagation attenuation as shown in Figure 4.7, and hence the large- or small-scale signal variations (e.g., path loss, multi-path fading) are not considered in [1].

4.3.2 Link Budget Analysis

Link budget analysis is a common evaluation for the performance of a radar system by accounting for all gains and losses from the transmitter to the receiver [21,53], but such analysis is not provided in [1]. For a radar system, the received power P_R can

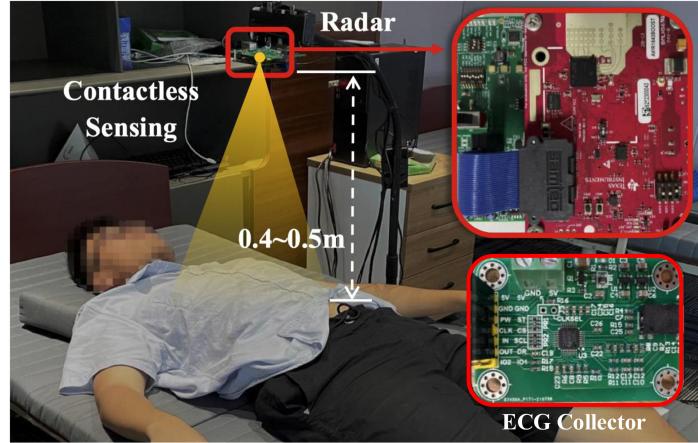


Figure 4.7: Environment settings for data collection from quasi-static subject [1].

be expressed with respect to the transmitted power P_T as:

$$P_R = \frac{G_T G_R \lambda^2 \sigma}{(4\pi)^3 R^4} P_T \quad (4.12)$$

where G_T/G_R is the gain of the transmitter/receiver, σ is the effective radar cross section (RCS) for the human chest region, and R is the distance between radar and chest [21].

In addition, the lowest detectable signal power of the receiver is defined in [53] as:

$$P_{r,\min} = -174 \text{ dBm} + 10 \log_{10}(B) + \text{NF} + \text{SNR}_{\min} \quad (4.13)$$

where -174 represents the thermal-noise level, NF means the noise figure, B is the bandwidth of the receiver and SNR_{\min} is the desired SNR considering the afterward signal processing.

By combining (4.12) and (4.13), the maximum detectable range can be calculated as:

$$R_{\max} = \sqrt[4]{\frac{G_T G_R \lambda^2 \sigma P_T}{(4\pi)^3 P_{r,\min}}} \quad (4.14)$$

All the radar-related values in (4.14) are shown in Table 4.4 according to the

Table 4.4: Parameters for Link Budget Analysis

Parameter	Value	Parameter	Value
Tx Gain (G_T)	10 dBi	Rx Gain (G_R)	30 dBi
Tx power (P_T)	12 dBm	Noise Figure (NF)	15 dB
Wavelength (λ)	3.9 mm	Bandwidth (B)	3.8 GHz
RCS (σ)	-20 dBsm	Desired SNR (SNR_{\min})	10 dB

datasheet of TI AWR-1843 radar [57]. According to the previous work [1,53], the RCS value can be estimated as -20 dBsm for the quasi-static subjects wearing electrically thin cloth (i.e., the thickness is much smaller than the wavelength), with respiration noise filtered during the pre-processing stage. After substituting the values, the lowest detectable signal power is $P_{r,\min} = -53$ dBm, and the maximum detectable range is $R_{\max} = 4$ m, revealing that the parameter setting used in [1] could provide the received signal with good SNR.

4.3.3 Dataset Description

Half of the actual dataset is released with 91 trials for 11 subjects (with subject ID 1, 2, 5, 9, 10, 13, 14, 16, 17, 29, 30), and each trial contains 3 minutes of data (radar measurements and ECG ground truth) with 200 Hz sampling rate collected under 4 physiological statuses (i.e., normal breath (NB, 43 trials), irregular breath (IB, 18 trials), sleep (SP, 18 trials) and post exercise (PE, 12 trials)). In addition, the work in [1] has pre-processed the radar signal using several techniques, such as 3D beamforming, dynamic time wrapping, to remove respiration noise and enhance cardiac activities. Lastly, no existing study is found for ECG reconstruction in literature based on the same dataset, and the proposed framework MMECG in [1] will be used as the only benchmark to make a comparison with our radarODE.

4.3.4 Implementation Details and Compared Frameworks

The proposed radarODE network is coded using PyTorch and trained for 200 epochs with batch size 32 on the NVIDIA RTX A4000 (16 GB) using stochastic gradient

descent optimizer with early stop function [92] and learning rate 0.001 based on a cosine annealing schedule [93]. The dataset is split into training and testing sets based on 11-fold cross-validation with 1 fixed subject for testing and the other 10 subjects alternatively selected for training or validation, ensuring to make the most of all the trials while excluding the testing data from the training phase. In addition, all the ground truth characteristic peaks, PPI, and cardiac cycles are obtained by the NeuroKit2 from ECG signals [80]. Furthermore, the 4-seconds-long input SST segments only contain the frequency component within [1, 25] Hz and are down-sampled to 30 Hz in the time-axis for saving memory usage in backbone design.

In addition, three frameworks are selected for comparison with the following brief introduction of the architecture:

- MMECG [1] receives multiple 1D radar signals as input and utilizes Conv1d and Transformer as decoder to simultaneously extract temporal and spatial features. The encoded features are further fused by multiplication and then decoded by Transconv1d and TCN to produce ECG recovery.
- RSSRNet [35] takes spectrogram (STFT) as input with a Conv2d backbone and Transformer encoder. The adopted decoder is Transconv2d, and the output is still a spectrogram and needs to be converted to the ECG signal via inverse STFT.
- RadarNet [30] adopts 1D radar signal as input and directly generates coarse ECG signal using Conv1d. Then, several layers of ResNet are adopted to refine the ECG waveform.

4.4 Experimental Results and Evaluations

This section provides the experimental results and evaluations in terms of three core modules as depicted in Figure 4.1, with the first module providing PPI estimation for input/output slicing and reshaping, the second module robustly generating ECG pieces for the single cardiac cycle, and the third module yielding the final long-term ECG recovery.

4.4.1 Evaluations of PPI Estimation

The PPI estimation is the first module of radarODE, and the accuracy of the estimated PPI directly affects the fidelity of the concatenated morphological reference. Therefore, Figure 4.8(a) shows the PPI estimation error obtained by Algorithm 2 for all subjects based on KDE defined in (4.9) or directly averaging the candidate PPI values. The large PPI error (e.g., for subject 13, 17) is normally caused by body movements or residual respiration noise due to IB or PE status, as also shown in Figure 4.8(b) with large median error and variation for both methods. In contrast, the subjects in NB and SP statuses tend to be stable with less body movement, and the respiration noise can be well eliminated, achieving low median PPI errors as 0.03s and 0.02s using the KDE-based method for each status.

Overall, it is clear that the KDE-based PPI estimation is more accurate than the mean-based estimation for each subject, as shown in the cumulative distribution function (CDF) in Figure 4.8(c), because the KDE-based method is robust to the outliers caused by noises and could figure out the correct PPI estimation near the majority of candidate values.

4.4.2 Evaluations of SCEG Module in radarODE

SCEG is the core module that realizes the domain transformation and ensures the robust long-term ECG recovery in the next stage, and the performance is evaluated on the generated single-cycle ECG pieces in terms of morphological accuracy, corrupt ECG reconstruction and absolute R-peak error as shown in Figure 4.9 and Table 4.5.

Comparison of Morphological Accuracy

The morphological accuracy is shown in Table 4.5 as the median value of root mean square error (RMSE) and Pearson correlation coefficient (PCC), with RMSE sensitive to the deviation of peaks and PCC focusing on the general shape. The overall performances for all trials are shown as CDF plots in Figure 4.9(a) and 4.9(b). The results indicate that radarODE could generate high-fidelity ECG signals with good

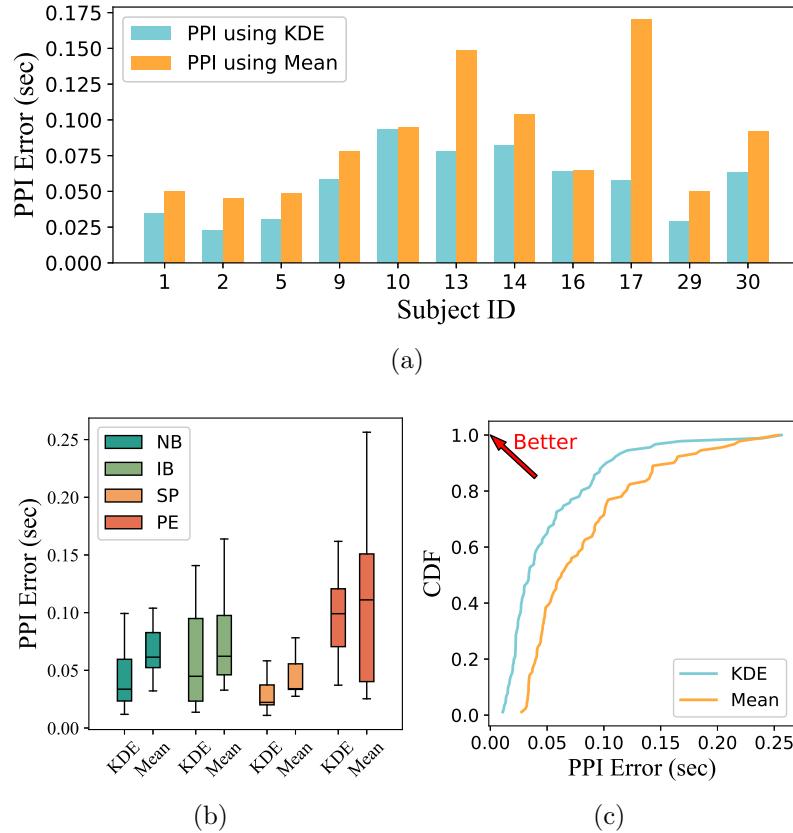


Figure 4.8: PPI estimation error of KDE-based and mean-based method: (a) View of different subjects; (b) View of different physiological statuses; (c) CDF for the overall PPI estimation error.

RMSE and PCC owing to the prior knowledge provided by the ODE decoder, while MMECG achieves the second-best result and shows a domain transformation ability for the majority of radar inputs with good SNR. In contrast, RSSRNet and Radar-Net achieve a similar performance because they only accept single-channel inputs and cannot effectively leverage the features of 50 channels provided in the dataset.

Comparison of Corrupt ECG Recovery

The ability to recover ECG pieces from corrupt radar signals reveals the noise robustness of each framework, and missed detection rate (MDR) is adopted to count failed

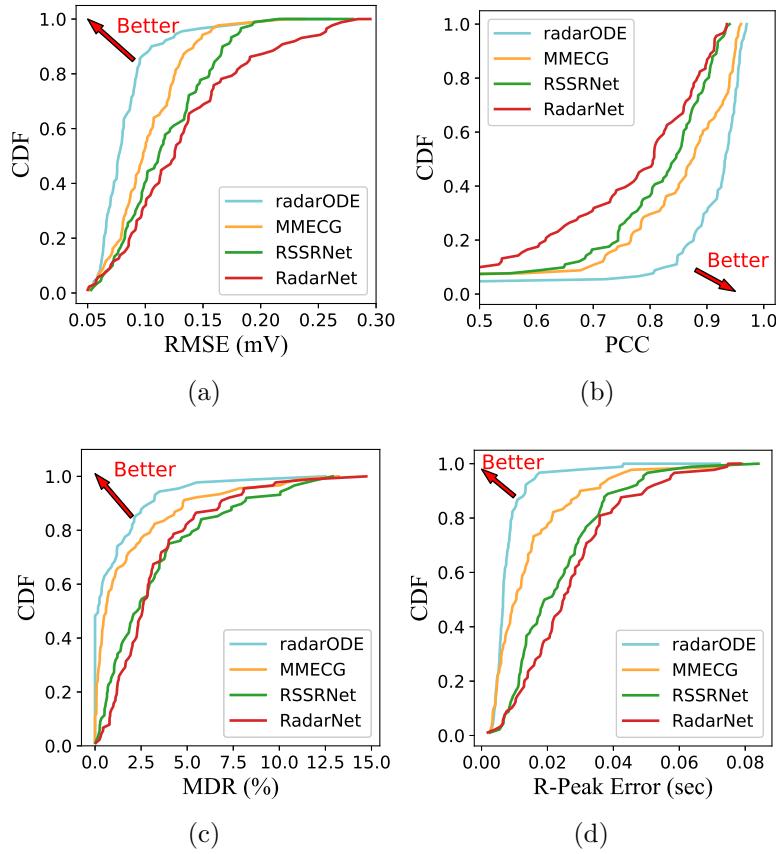


Figure 4.9: Performance comparison for single-cycle ECG recovery: (a) - (d) CDF of RMSE, PCC, MDR and R-peak error for all trials.

recoveries without showing characteristic ECG patterns (R peaks). Three rules are made to define the failed ECG recoveries:

- The deviation of the corrupt R peak from the ground truth exceeds the absolute tolerance of 0.15s [1].
- The corrupt R peak has one or more neighboring R peaks closer than 0.3s.
- The amplitude of the corrupt R peak is 30% lower than that of the ground truth R peak.

Table 4.5: Performance Comparison for Single-cycle ECG Recovery and Ablation Study

Framework	Backbone	Encoder	Decoder	RMSE (mV)	PCC	MDR	R Error (sec)
MMECG [1]	-	Conv1d + Transformer	Transconv1d + TCN	0.091	87.9%	1.24%	0.012
RSSRNet [35]	Conv2d	Transformer	Transconv2d	0.100	86.0%	2.17%	0.019
RadarNet [30]	-	Conv1d	Conv1d (ResNet)	0.113	80.2%	2.58%	0.024
radarODE	Deform Conv2d	Conv1d	Initial + Temporal	0.086	89.4%	1.53%	0.012
			Initial + ODE	0.092	85.5%	0.14%	0.005
			Initial + Temporal + ODE	0.077	92.6%	0.18%	0.006

The results of MDR are shown in Table 4.5 with CDF plots of all trials shown in Figure 4.9(c). It is evident that radarODE achieves better noise robustness compared with previous work, owing to the constraint brought by ODE decoder. MMECG still has a better performance compared with RSSRNet and RadarNet due to the powerful backbone to extract features from undistorted channels. The overall performance of MDR in Figure 4.9(c) coincides with the morphological accuracy in Figure 4.9(a) and 4.9(b), because the corrupt ECG recoveries also significantly affect the RMSE and PCC.

Comparison of R-peak Timing Error

After filtering the corrupt ECG pieces, the absolute timing error of R peaks is calculated to evaluate the quality of fine-grained ECG features, with the median value and CDF plots shown in Table 4.5 and Figure 4.9(d) respectively. The results are still similar to previous evaluations, with radarODE recovering the most accurate R peaks and the other three frameworks showing larger R-peak deviation. In addition to the benefits brought by ODE decoder, the adopted SST inputs also provide necessary time-frequency features to help the deep learning model identify indistinctive vibrations under low-SNR scenarios.

Ablation Study

The ablation study is performed to further evaluate the contributions of temporal and ODE decoders. The results in Table 4.5 reveal that both decoders can work individually and achieve reasonable results, but the ODE decoder has lower accuracy

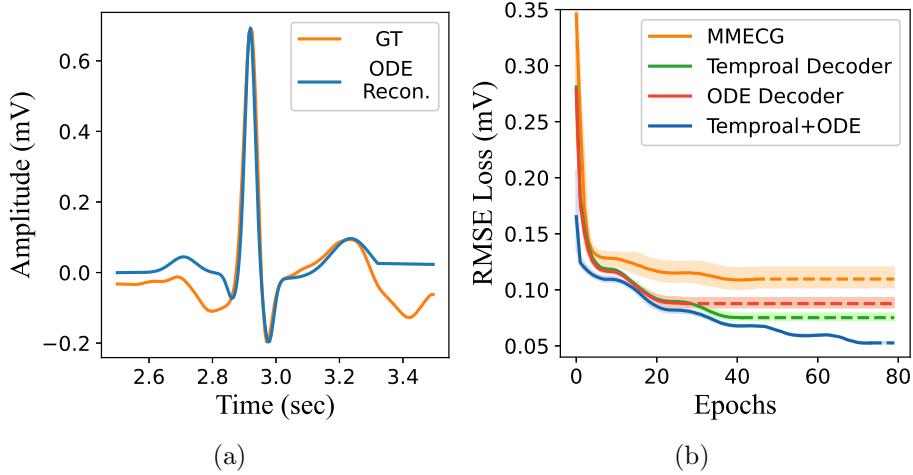


Figure 4.10: Ablation study for training process: (a) Rigid reconstruction (ODE Recon.) from ODE decoder with ground truth (GT); (b) Training loss comparison.

because it neglects the subtle ECG feature while only focusing on the characteristic peaks with rigid connections elsewhere, as shown in Figure 4.10(a). On the contrary, the introduced ODE model could resist strong noise and achieve the lowest MDR as 0.14%, while the temporal decoder only gets a similar MDR (1.53%) with MMECG (1.24%). In radarODE, the outputs of temporal and ODE decoder are fused together to achieve noise robustness with MDR= 0.18%, while maintaining a faithful ECG shape (i.e., good RMSE and PCC).

Visualization of Training Process

Figure 4.10(b) illustrates the training loss for benchmark and ablation study with the dotted line representing the early stop of the training process and the shaded area indicating the repetition of the training process for five times. From the ablation study, SCEG with only ODE decoder cannot provide a very accurate reconstruction due to the rigid shape as shown in Figure 4.10(a), while the temporal decode could achieve the second-best result. After combining ODE and temporal decoders, the outputs from ODE decoder act as the morphological-prior to accelerate the convergence, achieving

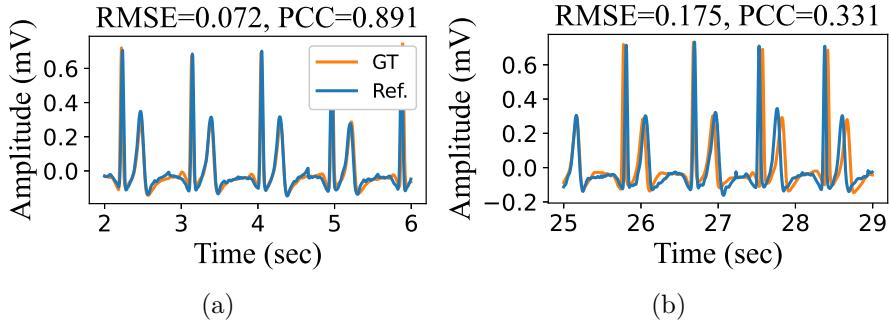


Figure 4.11: Visualization of misalignment: (a) and (b) High- and Low-fidelity morphological references (Ref.) caused by PPI estimation error.

the RMSE of 0.17mV after the first epoch. In addition, the morphological reference will not be destroyed by noises and could stabilize the training process, contributing to the lowest training loss as shown in Figure 4.10(b).

Summary of SCEG Evaluation

The proposed SCEG module in radarODE achieves better accuracy in generating single-cycle ECG pieces with the improvements brought by the ODE decoder and SST inputs, enabling successful recoveries even under abrupt noises and achieving the best RMSE, PCC and R-peak accuracy compared with previous frameworks.

The generated ECG pieces can be resized and concatenated based on PPI estimation, but the PPI estimation error may accumulate and degrade the accuracy as shown in Figure 4.11(a) and 4.11(b), because slight deviations of the peaks ruin the overall RMSE/PCC, hence requiring long-term reconstruction module to refine the concatenated ECG pieces in the next step. To provide a comprehensive evaluation for long-term ECG recovery from various dimensions (e.g., trials, subjects, and physical statuses), MMECG is selected as the only benchmark in the next section for a clear result visualization.

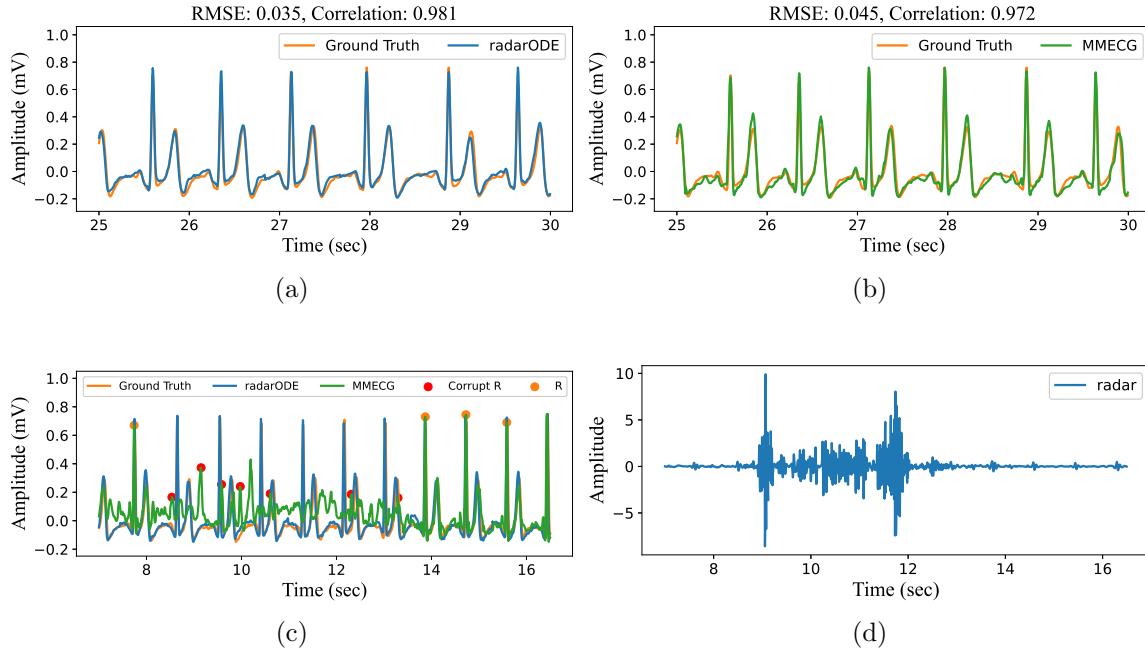


Figure 4.12: Corrupt ECG reconstruction: (a) and (b) Ideal reconstruction results for radarODE and MMECG; (c) Corrupt ECG reconstruction yield by MMECG and faithful reconstruction from radarODE; (d) Corresponding radar signal during body movements.

4.4.3 Overall Evaluations of radarODE

The long-term ECG reconstruction module finally generates 3-minute-long ECG signals for the evaluations of the entire radarODE framework in terms of corrupt ECG reconstruction, morphological accuracy, and fine-grained cardiac feature accuracy.

Corrupt ECG Reconstruction

The ideal reconstructed ECG signals are shown in Figure 4.12(a) and 4.12(b) with corresponding RMSE/PCC labeled. However, the long-term ECG signal may contain corrupt parts due to the presence of body movements (especially in IB and PE). Figure 4.12(c) shows the corrupted ECG reconstruction yield by MMECG with the falsely detected R peaks noted as red dots, and Figure 4.12(d) is the raw radar signal with

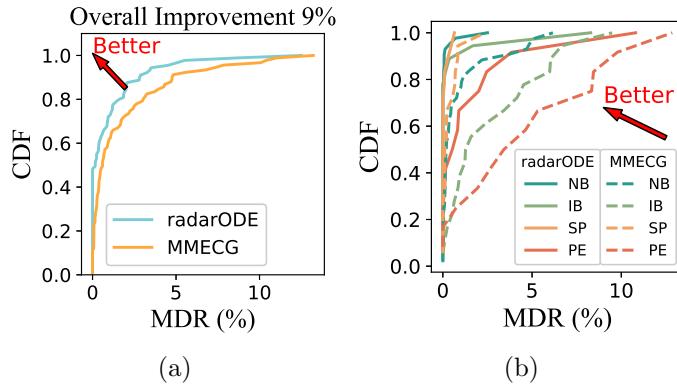


Figure 4.13: Statistical results for corrupt ECG reconstruction: (a) Overall missed detection rate (MDR); (b) MDR under different physical statuses.

extensive distortion induced by body movements. In contrast, the radarODE could still provide faithful ECG reconstructions under body movements due to the introduction of prior knowledge (ODE model) about ECG as the morphological reference, hence gaining certain robustness in resisting noises.

The overall MDR can be calculated from the recovered ECG signal with the CDF plot shown in Figure 4.13(a), and the overall improvement achieved by radarODE is 9%. In addition, the CDF plots for different physical statuses are plotted in Figure 4.13(b) with the 90-percentile MDR of 0.12%, 0.85%, 0.32%, 3.71% during NB, IB, SP, PE for radarODE and 3.36%, 6.22%, 0.83%, 9.64% for MMECG. The result shows that different physical statuses have noticeable impacts on the quality of the reconstructed long-term ECG, and radarODE could provide a lower MDR than MMECG in all statuses owing to the prior knowledge in the ODE decoder.

Morphological Accuracy

The morphological accuracy measures the similarity between reconstructed and ground truth ECG signals by calculating RMSE and PCC. Figure 4.14(a) and 4.14(b) show the overall performance of radarODE and MMECG in CDF with the median RMSE and PCC of 0.097mV/89.6% and 0.120mV/81.2%. It is worth noticing that the overall improvement of radarODE compared to MMECG is 16% and 19% for RMSE and

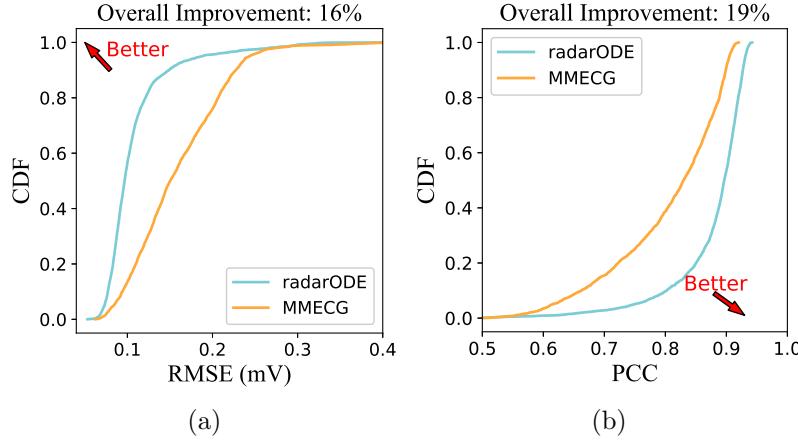


Figure 4.14: Morphological accuracy comparison for all trials: (a) and (b) Overall CDF of RMSE and PCC for all trials.

PCC across 91 trials respectively, indicating that the morphological-prior is more helpful in generalizing the typical ECG pattern than calibrating the peaks.

In addition, Figure 4.15(a) and 4.15(b) illustrate the RMSE/PCC across all subjects, with radarODE always achieving better results than MMECG. It is worth noting that the results of the long-term reconstruction show certain consistency with the previous PPI estimation error, because the fidelity of the morphological reference is directly affected by the PPI error. For example, subjects 10, 13, 14, and 30 get worse results than others in either RMSE or PCC evaluation due to the large PPI estimation error as shown in Figure 4.8(a).

Lastly, Figure 4.16(a) and 4.16(b) illustrate the RMSE/PCC for all trials in terms of different physical statuses, and the box plots show that the stable statuses (i.e., NB, SP) guarantee the reconstruction with small variance. In contrast, unstable statuses (i.e., IB, PE) can severely ruin the radar signal due to body movements, causing an inconsistent quality of the reconstructed ECG. However, radarODE could still provide the reconstructions with a smaller variance than MMECG, especially for unstable statuses because of the morphological-prior embedded in the ODE decoder.

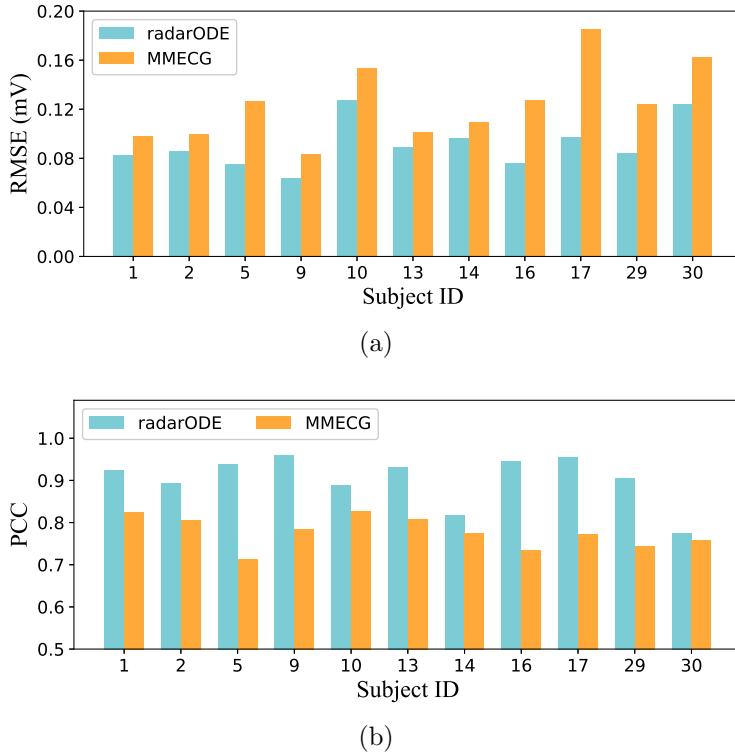


Figure 4.15: Morphological accuracy comparison across all subjects: (a) and (b) RMSE and PCC across all subjects.

Fine-Grained Cardiac Events Reconstruction

The evaluation of fine-grained cardiac features aims to analyze the timing accuracy of the QRST peaks and the P peak is not considered in this evaluation as also suggested in the benchmark paper [1], because the P peak is inconspicuous and even unable to be detected for some ground truth ECG signal. The overall result is shown in Figure 4.17 as CDF with the median/90-percentile absolute timing error for QRST peaks shown in Table 4.6. The improvement owes to the use of SST spectrogram with more evident patterns for the prominent heart vibrations, and the ODE decoder contributes to calibrating the peak positions according to η and τ , hence improving the overall peak accuracy.

Figure 4.18(a) and 4.18(b) demonstrate the improvement of radarODE in terms of

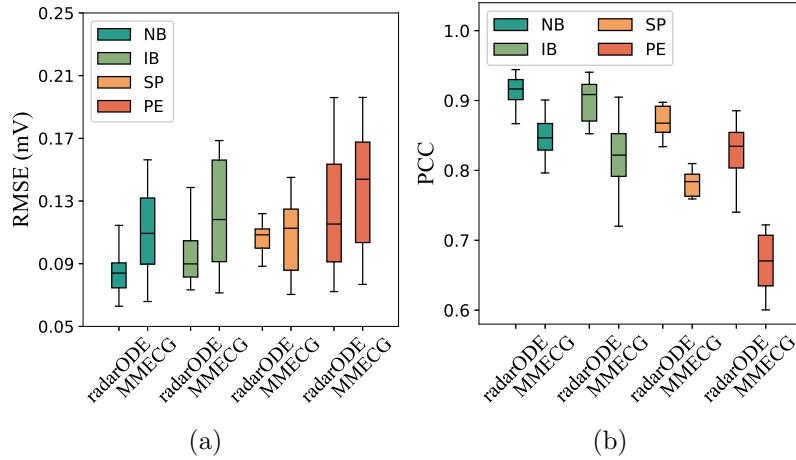


Figure 4.16: Morphological accuracy comparison for different physical statuses: (a) and (b) RMSE and PCC under different physical statuses.

Table 4.6: Absolute Timing Error for Reconstructed ECG Peaks

Framework	Percentile	Q	R	S	T
MMECG [1]	Median	0.021	0.012	0.019	0.018
	90-percentile	0.048	0.023	0.029	0.035
radarODE	Median	0.015	0.007	0.009	0.014
	90-percentile	0.027	0.015	0.020	0.023

unit: second

different physical statuses, and the results coincide with many previous evaluations. Firstly, the radarODE outperforms MMECG for all statuses with the R peak always achieving the best accuracy. Secondly, stable physical statuses tend to yield accurate reconstruction with small variance, but the difference between statuses is smaller than that of the morphological accuracy analysis, because the corrupt reconstructions have been filtered for peak accuracy evaluation.

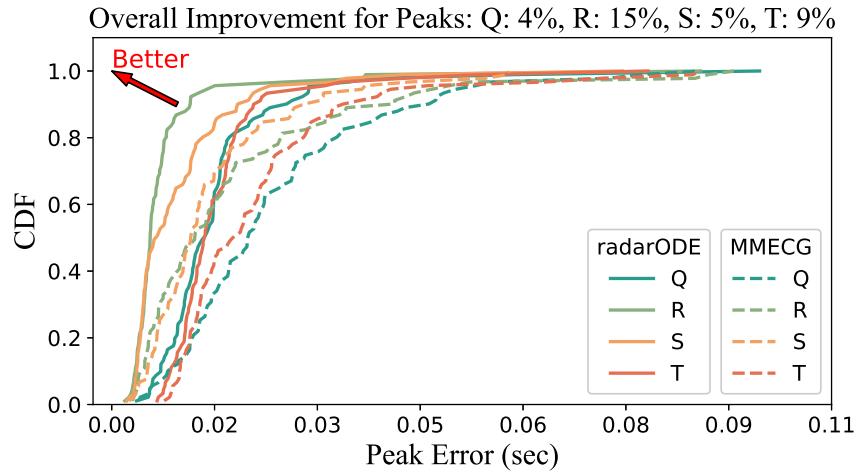


Figure 4.17: CDF of the overall timing error.

Summary of Long-term ECG Recovery Evaluation

The experimental results illustrated that the proposed radarODE could yield high-quality long-term ECG recovery even under extensive body movement noise, with better RMSE and PCC compared with the benchmark. In addition, the introduction of the SST spectrogram and the ODE decoder further improves the accuracy of the fine-grained cardiac features that are crucial to the potential applications in clinical diagnosis.

4.5 Conclusions

Radar-based ECG reconstruction is highly reliant on purely data-driven approaches and lacks theoretical support regarding the transformation between mechanical activities measured by radar and electrical activities described as ECG. This research aims to bridge the gap to realize the transformation from the mechanical domain to the electrical domain by proposing the signal model with fine-grained features considered and further designing a deep learning framework radarODE with morphological prior embedding as ODEs. The radarODE framework is validated on the public dataset containing 4.5 hours of radar measurements with corresponding ablation study and

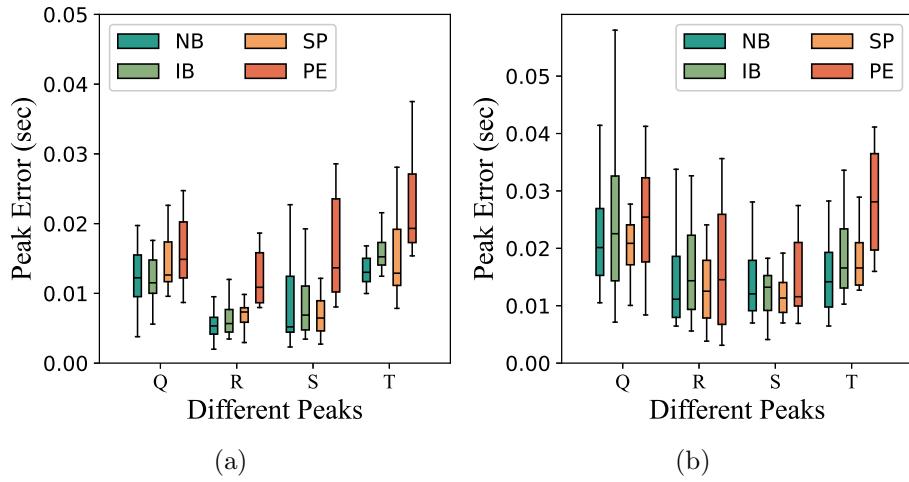


Figure 4.18: Evaluations of fine-grained cardiac events: (a) and (b) The timing error for different peaks under various physical statuses for radarODE and MMECG.

comparisons with the benchmark. The experimental result shows that radarODE could achieve a better MDR, morphological accuracy and peak accuracy than the benchmark, proving the rationality of the proposed signal model and the effectiveness of the radarODE under various physical statuses and random body movement.

Chapter 5

Robust Long-Term ECG Generation

This chapter further extends the robust single-cycle ECG recovery to the robust long-term ECG recovery by leveraging multi-task learning (MTL) as a popular paradigm to help the concatenation of the well-recovery ECG pieces. According to the experiments in Chapter 4, the misalignment cause by inaccurate peak-to-peak interval (PPI) estimation heavily ruin the final results, and it is necessary to design the long-term ECG recovery based on single-cycle ECG pieces, instead of realizing the recovery process in terms of arbitrary divided radar/ECG pairs. Therefore, a novel MTL structure designed for long-term ECG recovery is designed with three sub-tasks responsible for recovering different ECG features. In addition, a optimization strategy is also designed to balance the training process for the sub-tasks with great disparity in terms of intrinsic difficulties. Sufficient experiments show that the proposed radarODE-MTL with eccentric gradient alignment (EGA) optimization strategy outperforms other frameworks and optimization strategies under various noise conditions and datasets, and the deconstructed tasks in radarODE-MTL could further improve the interpretability in radar-based ECG recovery.

5.1 Introduction

In the literature, radar-based ECG waveform recovery has been achieved based on various deep-learning architectures, such as convolutional neural network (CNN) [1,2], long short-term memory (LSTM) network [73], and Transformer [1,35]. However, the noise robustness of the deep-learning framework is rarely investigated in the literature, especially for the random body movement (RBM) noise that is inevitable in contactless monitoring and has orders of magnitude larger than cardiac activities. The existing work either discarded the data during the RBM [36] or reported the heavy distortion as the future work [1]. Additionally, the existing deep-learning methods are also blamed for being purely data-driven as a black box and the transformation between cardiac mechanical and electrical activities lacks the theoretical explanation [2]. The main problem in the existing domain transformation methods can be summarized as follows:

- The transformation between arbitrary radar/ECG pairs is hard to model, and hence the ECG recovery process is vulnerable to the noises with bad root mean square error (RMSE) and Pearson correlation coefficient (PCC) as shown in Figure 5.1(a).
- Although the model for the domain transformation between single-cycle radar/ECG pair has been proposed in [2], the long-term ECG recovery might be misaligned with ground truth due to inaccurate PPI estimation [2], deteriorating the RMSE/PCC even if the morphological features are well-recovered as shown in Figure 5.1(b).

Based on the limitations of the existing methods, it is necessary to provide a feasible model that explains the transformation inside radar-based long-term ECG recovery and is also robust to real-life noises. Therefore, this work proposes to deconstruct the radar-based ECG reconstruction into three individual tasks as a multi-task learning (MTL) problem to extract cardiac features with different levels of granularity, i.e., coarse features: heartbeat detection and cardiac cycle timing; fine-grained feature: ECG waveform. However, another consequent problem is to simultaneously

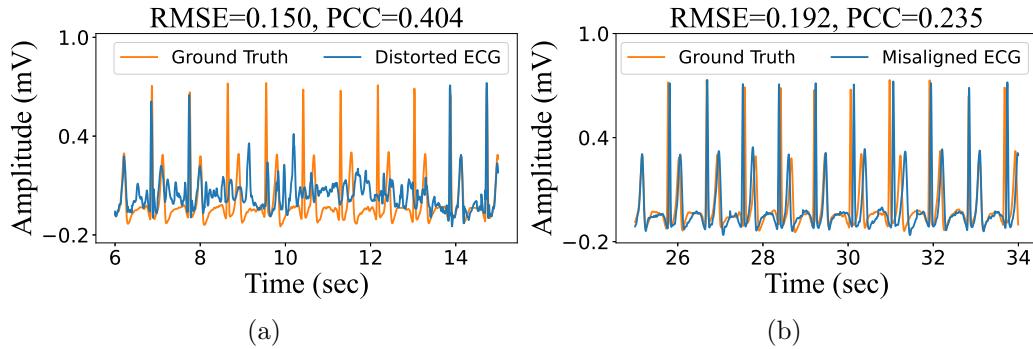


Figure 5.1: The impact of strong noise and misalignment: (a) ECG recovery distorted by RBM noise [1]; (b) Misaligned ECG recovery due to the inaccurate PPI estimation [2].

optimize three individual tasks under the MTL paradigm, because the optimization of one task may degrade the performance of the others [94, 95].

In the literature, MTL is a widely-used deep learning paradigm in various fields such as scene understanding [96, 97], autonomous driving [75] and speech/text processing [98]. However, the MTL paradigm has never been applied in radar-based ECG recovery, and the existing MTL optimization strategies cannot fairly optimize all the tasks due to the imbalanced task difficulties [99]. In this work, the difficulty of extracting the ECG waveform is much higher than the other two, and simply applying the existing optimization strategies cannot achieve an ideal result with fair improvements on all tasks according to our initial experiments.

Inspired by the above discussion, the contributions of this work can be concluded as:

- A novel optimization strategy called EGA is proposed for updating shared parameters in the MTL neural network, aiming to balance the intrinsic difficulty across tasks during network training and also prevent the negative transfer phenomenon.
- To the best of our knowledge, this is the first work that investigates the noise robustness in radar-based ECG recovery against constant or abrupt noise by

modeling the cardiac domain transformation as three tasks. An end-to-end MTL framework named radarODE-MTL is accordingly proposed to realize these tasks and leverage adjacent cardiac cycles to compensate for the distorted one.

- Sufficient experiments show that the proposed radarODE-MTL with EGA optimization strategy outperforms other frameworks and optimization strategies under various noise conditions and datasets, and the deconstructed tasks in radarODE-MTL could further improve the interpretability in radar-based ECG recovery.

The rest of the chapter is organized as follows. Section 5.2 provides the background for radar-based ECG recovery and MTL optimization. The proposed radarODE-MTL framework with EGA strategy is elaborated in Section 5.3, and the experimental settings and results are shown in Section 5.4. At last, Section 5.5 concludes this paper.

5.2 Background and Problem Statement

This section will provide compact explanations of the domain transformation in ECG recovery and the optimization problem in MTL network, with the corresponding problem statements.

5.2.1 Model for Domain Transformation and Problem Statement

Signal Model for Cardiac Mechanical Activities

In radar-based ECG recovery, the baseband signal is normally pre-processed using bandpass filter, differentiator and digital beamforming to remove the background and respiration noise to enhance cardiac-related features [1, 27, 46]. According to our previous work [2], the fine-grained cardiac mechanical activities include aortic valve opening/closure (AO/AC) and mitral valve opening/closure (MO/MC), revealed by

the corresponding prominent vibrations v_1 and v_2 as measured in radar signal $x(t)$ as depicted in Figure 2.1. Therefore, the resultant radar signal $x(t)$ can be expressed for K cardiac cycles as:

$$x(t) = \sum_{k=1}^K v_1^k(t) + \sum_{k=1}^K v_2^k(t) + n_{abr}(t) + n_{con}(t) \quad (5.1)$$

with

$$\begin{aligned} v_1^k(t) &= a_1^k \cos(2\pi f_1^k t) \exp\left(-\frac{(t - T_1^k)^2}{b_1^{k2}}\right) \\ v_2^k(t) &= a_2^k \cos(2\pi f_2^k t) \exp\left(-\frac{(t - T_2^k)^2}{b_1^{k2}}\right) \end{aligned} \quad (5.2)$$

where a_1^k , b_1^k and a_2^k , b_2^k jointly determine the amplitudes and lengths of the first and second prominent vibrations for k^{th} cardiac cycle, f_1^k , f_2^k are the corresponding central frequencies and T_1^k , T_2^k represent when the vibrations happen. In addition, $n_{abr}(t)$ represents the abrupt noises (e.g., RBM) and $n_{con}(t)$ describes many other constant noises that affect the SNR, such as thermal noise [27, 100], monitoring from random directions [24] and long-range monitoring [46].

Model of Domain Transformation

The radar signal modeled in (5.1) shares a strong temporal consistency with the ECG signal as shown in Figure 2.1, because the excitation-contraction coupling indicates that the electrical signal (ECG) triggers the heart muscle contraction (SCG) [23]. Therefore, this work proposed to deconstruct the radar-based ECG recovery into three tasks to realize the robust transformation from the measured radar signal $x(t)$ to the ECG signal, and the three tasks can be modeled as:

- Task 1: The reconstruction of the morphological features aims to map the single-cycle cardiac activities $x(t)$ to ECG with the deep neural network acting as a mapping function as $x_{ecg}(t) = \mathcal{T}(x(t))$.
- Task 2: The detection of R peaks (anchors) is equivalent to finding $\mathbf{R} =$

$\{T_1^1, T_1^2, \dots, T_1^K\}$ in (5.2) according to the central frequency f_1^k of v_1^k , as shown in Figure 2.1.

- Task 3: The prediction of the cardiac cycle length is equivalent to finding the peak-to-peak interval (PPI) used for resizing x_{ecg} obtained in Task 1, as shown in Figure 2.1.

Theoretically, PPI can be directly obtained from \mathbf{R} as $T_1^{k+1} - T_1^k$, but it is necessary to reckon the PPI estimation to be an individual task in practice, because if one R peak fails to be detected in \mathbf{R} , the resultant PPI will be extremely large, destroying the long-term ECG recovery.

Problem Statement for Domain Transformation

The fine-grained ECG recovery could only be realized by deep-learning methods, and the noise robustness of the deep-learning model has never been evaluated in the literature [1, 2, 35, 36]. Therefore, radarODE-MTL dissects the long-term ECG recovery into three tasks, and hence each decoder only focuses on extracting the cardiac feature with different granularity, aiming to improve the accuracy and noise robustness of the radar-based ECG recovery.

5.2.2 Optimization Strategies for MTL

Optimization of MTL Network

A standard definition for an MTL optimization problem with n tasks under hard parameter sharing (HPS) [101] architecture is given by:

$$\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^m} \left\{ \mathcal{F}(\boldsymbol{\theta}) \triangleq \frac{1}{n} \sum_{i=1}^n \mathcal{L}_i(\boldsymbol{\theta}) \right\} \quad (5.3)$$

where $\boldsymbol{\theta} \in \mathbb{R}^m$ denotes the shared parameter space, $\mathcal{L}_i(\boldsymbol{\theta})$ is the task-specific non-negative objective function for $\mathbb{R}^m \rightarrow \mathbb{R}_+$, and $\mathcal{F}(\boldsymbol{\theta})$ represents a mapping from the

parameter space to the objective space as $\mathbb{R}^m \rightarrow \mathbb{R}^n$. The MTL optimization strategy aims to find the optimal parameter set $\boldsymbol{\theta}^*$ that minimizes the average loss.

The dilemma in the design of MTL optimization strategies is mainly on avoiding negative transfer when the optimization of individual tasks conflicts with each other [102–109], spawning two main categories of methods, loss balancing method and gradient balancing methods, to impartially search for the optimal solution(s) subjecting to Pareto optimality [105].

The loss balancing methods add the weight to each task loss $\mathcal{L}_i(\boldsymbol{\theta})$ based on various criteria, such as learning rate [107], inherent task uncertainty [109] or the loss magnitude [104]. In contrast, gradient balancing methods address the negative transfer by balancing both magnitudes and the directions of the task-specific gradient $\mathbf{g}_i = \nabla_{\boldsymbol{\theta}} \mathcal{L}_i(\boldsymbol{\theta})$, according to certain criteria such as the cosine similarity between gradients [105], descending rate [105] or the orthogonality of the gradient system [102].

Problem Statement for Designing MTL Optimization Strategies

The existing methods perform not well on the proposed radarODE-MTL framework because most methods aim to treat all the tasks equally and pay too much attention to the easy tasks with the least achievement after convergence (e.g., slow learning rate in GradNorm [110], small singular value in Aligned-MTL [102]), while the hard task tolerates a slow convergence rate due to the limited gradient magnitudes or update frequencies [99]. Several studies in the literature proposed to increase the weight for the hard task metered the learning rate [99]. However, the forcible change of the weight may aggravate the gradient conflict and hence degrade other tasks, because the loss-balancing method can not alleviate the gradient conflict issue [105].

In addition, the slow learning rate can be interpreted in two ways: (a) The optimization stalls due to the compromise in gradients normalization, and the constraint on the hard task should be released as adopted in GradNorm [110] and DWA [107]; (b) The optimization has already achieved convergence and should be terminated as in the early stop technique [92]. Unfortunately, it is hardly investigated whether the optimization actually converges or stalls, or say, should more computational resources be

skewed towards the task with limited learning progress. Therefore, EGA is proposed in this paper to estimate the intrinsic task difficulty based on the current learning progress and dynamically alter the gradients in orthogonal space to fairly benefit all the tasks without knowing the actual optimization status (i.e., stall or convergence).

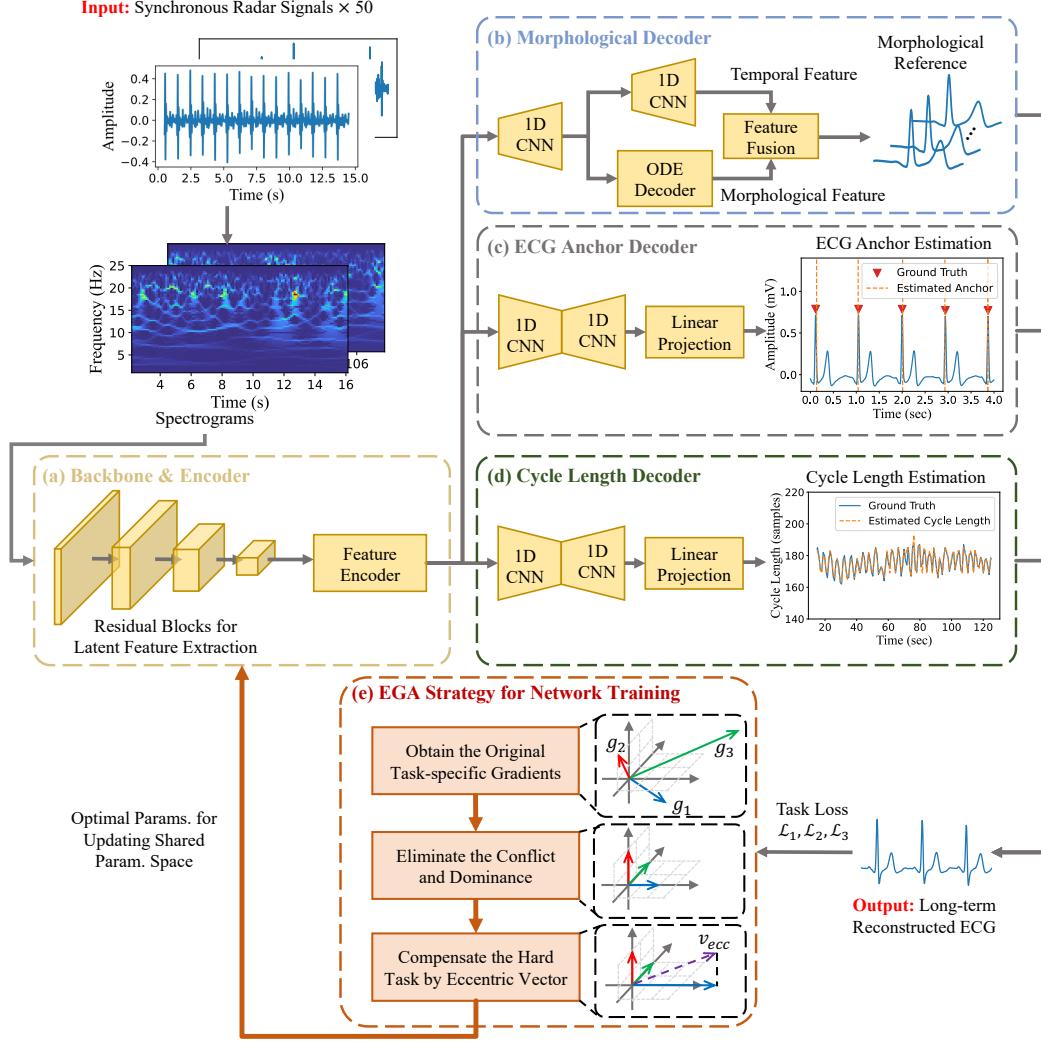


Figure 5.2: Overview of the radarODE-MTL framework with EGA strategy: (a) Shared backbone extracts time-frequency features; (b) Morphological decoder only reconstructs the shape of the current ECG piece; (c) ECG anchor decoder estimates the time-index of anchors ; (d) Cycle length decoder estimates the cardiac cycle length; (e) The proposed EGA strategy for optimizing shared parameter space.

5.3 Methodology

5.3.1 Overview of radarODE-MTL with EGA Strategy

The aforementioned three deconstructed tasks for radar-based ECG recovery can be realized by the proposed radarODE-MTL framework as shown in Figure 5.2, and the dataset used for training and validation is provided in [1]. Firstly, the 50 synchronous radar signals will be pre-processed into spectrograms by synchrosqueezed transform (SST) to highlight the central frequencies for locating the prominent vibrations v_1 and v_2 . Then, radarODE-MTL is designed to generate the long-term ECG recovery in an end-to-end manner with certain shared layers to capture the common representations for all tasks and three task-specific decoders to recover the ECG morphological features, detect ECG anchors (R peaks) and estimate single-cardiac-cycle length respectively, as shown in Figure 5.2(a)-(d).

During the training stage, the network optimization of three decoders follows the standard single-task optimization method, and the share parameter space (Backbone and Encoder) is updated using the proposed EGA strategy based on the task-specific loss $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$, as shown in Figure 5.2(e). In general, the EGA strategy first tries to eliminate the conflict and dominance among the original task-specific gradients, e.g., $\mathbf{g}_1, \mathbf{g}_2$ have opposite directions and \mathbf{g}_3 has large magnitude. Secondly, the eccentric vector (v_{ecc}) is introduced for balancing the task difficulties to fairly optimize all the tasks.

5.3.2 Backbone and Encoder

The backbone of radarODE-MTL is used to extract the latent features from the input SST spectrograms as shown in Figure 5.2(a) and is expected to figure out the remarkable patterns for vibrations v_1 and v_2 with certain central frequencies and periodicity. Specifically, four residual blocks are adopted in this work as the backbone because the ResNet has been proven to be an efficient structure in computer vision or signal processing [41, 88, 111]. Then, the encoder contains only one 2D convolutional layer to further compress the feature in the time-frequency domain into the 1D time

domain for later processing. The performance of the backbone and encoder has been verified in our previous work with the detailed structure shown in [2].

5.3.3 Morphological Decoder

The morphological decoder has been designed in our previous work radarODE [2] as the single cycle ECG generate (SCEG) module to realize the robust domain transformation in a single cardiac cycle with a fast rate of convergence, because an ordinary differential equation (ODE) model is introduced in the ODE decoder to provide morphological feature as the prior knowledge to guide/constrain the ECG recovery. Similarly, in radarODE-MTL, a morphological decoder will be used to realize the mapping function $\mathcal{T}(\cdot)$ in Task 1 and generate morphological reference by fusing both temporal and morphological features, as shown in Figure 5.2(b).

5.3.4 ECG Anchor Decoder and Cycle Length Decoder

The ECG anchor decoder and cycle length decoder are designed to identify the time-domain anchors T_1^k and single-cardiac-cycle length PPI^k in Task 2 and 3 simultaneously for the accurate alignment of ECG pieces as shown in Figure 5.2(c) and (d), avoiding the impact of error accumulation in long-term ECG recovery [2]. In addition, the prediction of the ECG anchors and cycle lengths can leverage the context information even if the current cardiac cycle is ruined by noises, because the vital signs are nearly unchanged for healthy people in successive cardiac cycles [13].

The structures of the ECG anchor decoder and cycle length decoder are the same as shown in Figure 5.2(c) and (d), with several layers of 1D CNN-based encoder/decoder followed by a linear projection block. Specifically, the encoder is assembled by four 1D CNN blocks with each block containing 1D convolution, batch normalization (BN) and rectified linear unit (ReLU) activation function; the decoder is composed of two 1D transposed CNN blocks with each block containing 1D transposed convolution, BN and ReLU; and the linear projection block is assembled by linear layer, BN and ReLU with one linear layer appended at last as the output layer.

5.3.5 Input, Output and Loss Function

The inputs of radarODE-MTL are the 4-sec segments divided from long-term radar signal with a step length of 1 sec, and the middle cardiac cycle is selected as the ground truth ECG piece. Then, to calculate the loss value, the ground truth ECG piece should be resampled as a fixed length 200 to match the output dimension, and the RMSE is used to calculate \mathcal{L}_1 . The output of the ECG anchor decoder should contain multiple predicted anchors within 4-sec segment, and the cross-entropy loss is used for \mathcal{L}_2 calculation as a multi-class classification problem (i.e., each time index acts as a possible class). Differently, the output of the cycle length decoder only represents the length of the current evaluated cardiac cycle with only one true label (value = 1), and the cross-entropy loss is used for \mathcal{L}_3 calculation as a one-class classification problem.

Eventually, the calculated $\mathcal{L}_1, \mathcal{L}_2, \mathcal{L}_3$ will be used for optimization using the later proposed EGA strategy during training, otherwise the three outputs can directly form the long-term ECG recovery by aligning the recovered ECG pieces (Task 1) with the predicted anchors (Task 2) after resampling the ECG pieces as the cycle lengths (Task 3).

5.3.6 Eccentric Gradient Alignment (EGA) Strategy

According to the discussion in Section 5.2.2, the imbalanced difficulties among three tasks will raise a new challenge to not only simultaneously optimize all the tasks without negative transfer [106], but also keep improving the hard tasks even if the easy tasks have already achieved convergence.

In this case, EGA first needs to solve the gradient conflict and magnitude dominance within the original task-specific gradients $\mathbf{g}_1, \mathbf{g}_2, \mathbf{g}_3$ as shown in Figure 5.3(a), e.g., \mathbf{g}_1 and \mathbf{g}_2 may have opposite directions hence canceling with each other, and \mathbf{g}_3 may have a large magnitude hence dominating the linear combination of all the gradients, with the resultant \mathbf{g}_{joint} leaning on \mathbf{g}_3 . A common solution is to project all the gradients into an orthogonal space to eliminate gradient conflict [102, 112], and hence the optimization based on \mathbf{g}_{joint} will not degrade any of the tasks. Then, the

magnitude of the gradients will be unified as the same value (e.g., $\tilde{\sigma}$) to obtain new task-specific gradients $\tilde{g}_1, \tilde{g}_2, \tilde{g}_3$, as shown in Figure 5.3(b).

Furthermore, instead of categorically selecting the hard task based on the learning rate and only increasing the corresponding weight, EGA creatively provides an adjustable estimation of the intrinsic task difficulty by mapping the learning rate through a softmax with hyperparameter T . In other words, suitable intrinsic task difficulty can be obtained by adjusting T without knowing the actual optimization status (i.e., stall or convergence), and the discrepancy among task difficulties can be adjusted to avoid overlooking or overrating any task. In practice, to integrate the estimated intrinsic task difficulty with MTL optimization, EGA proposed to add an eccentric vector v_{ecc} to eccentrically align the joint gradient \tilde{g}_{joint} to the hard task, as shown in Figure 5.3(c).

The detailed EGA strategy will be explained in this section in terms of the preparation stage, gradient projection and normalization, and eccentric gradient alignment.

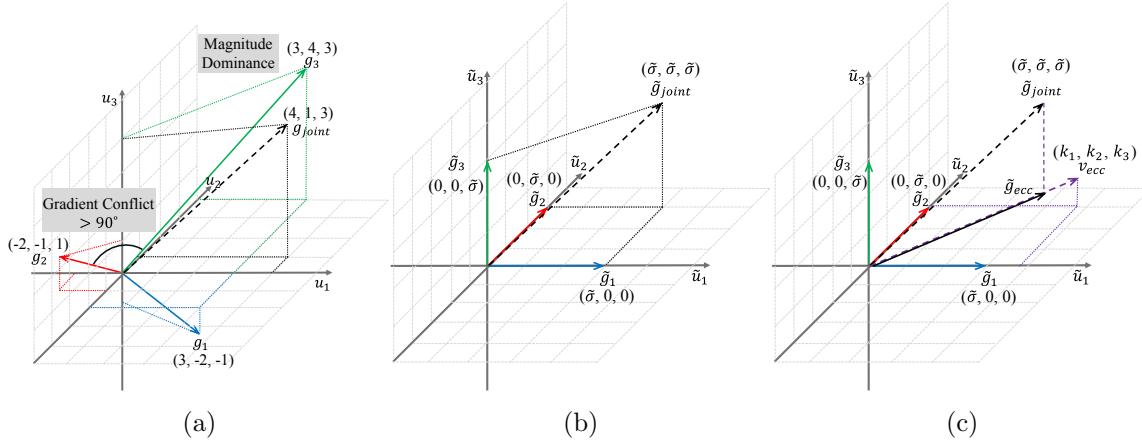


Figure 5.3: Illustration of EGA: (a) Original gradient space with gradient conflict and magnitude dominance; (b) The projection of the original gradient space into the orthogonal space with equal “learning rate”; (c) The implementation of eccentric gradient alignment to skew the joint gradient \tilde{g}_{joint} towards the hard task by introducing the eccentric vector v_{ecc} .

Preparations for EGA Optimization

As a gradient-based MTL optimization method with objective function in (5.3), EGA requires to access task-specific gradient in terms of the shared parameters $\boldsymbol{\theta}$, and the gradients can be obtained as $\mathbf{g}_i = \nabla_{\boldsymbol{\theta}} \mathcal{L}_i(\boldsymbol{\theta}), i \in [n]$, forming the original gradient matrix as $\mathbf{G} = \{\mathbf{g}_1, \dots, \mathbf{g}_n\} \in \mathbb{R}^{n \times m}$. Then, the joint gradient for optimizing the shared parameter space can be linearly combined as $\mathbf{g}_{joint} = \mathbf{G}^\top \mathbf{w}$, with $\mathbf{w} = [1, \dots, 1]^\top$ representing the weights for each \mathbf{g}_i . The original gradient matrix \mathbf{G} normally has gradient conflict and magnitude dominance issues, as shown in Figure 5.3(a).

Gradients Projection and Normalization

In order to solve the conflict inside the gradient matrix \mathbf{G} , the orthogonal projection problem can be formulated as finding a gradient matrix $\tilde{\mathbf{G}}$ with the new joint gradient $\tilde{\mathbf{g}}_{joint} = \tilde{\mathbf{G}}^\top \mathbf{w}$ close to the original \mathbf{g}_{joint} :

$$\min \| \mathbf{g}_{joint} - \tilde{\mathbf{g}}_{joint} \|_2^2 \quad \text{s.t.} \quad \tilde{\mathbf{G}} \tilde{\mathbf{G}}^\top = \mathbf{I} \quad (5.4)$$

Then, according to the derivation based on triangle inequality:

$$\| \mathbf{g}_{joint} - \tilde{\mathbf{g}}_{joint} \|_2^2 = \| \mathbf{G}^\top \mathbf{w} - \tilde{\mathbf{G}}^\top \mathbf{w} \|_2^2 \leq \| \mathbf{G}^\top - \tilde{\mathbf{G}}^\top \|_F^2 \| \mathbf{w} \|_2^2 \quad (5.5)$$

At last, the projection problem can be finally formulated as:

$$\min_{\tilde{\mathbf{G}}} \| \mathbf{G} - \tilde{\mathbf{G}} \|_F^2 \quad \text{s.t.} \quad \tilde{\mathbf{G}} \tilde{\mathbf{G}}^\top = \mathbf{I} \quad (5.6)$$

The solution to the problem in (5.6) has been given in the orthogonal Procrustes problem [113] by simply applying singular value decomposition (SVD) to \mathbf{G} as:

$$\mathbf{G} = \mathbf{U} \Sigma \mathbf{V}^\top \quad (5.7)$$

Then, the orthogonal gradient matrix $\tilde{\mathbf{G}}$ with unit singular values can be obtained as:

$$\tilde{\mathbf{G}} = \mathbf{U}\mathbf{V}^\top \quad (5.8)$$

In addition, the calculation can be simplified by applying the eigenvalue decomposition to the Gram matrices $\mathbf{G}\mathbf{G}^\top$ as:

$$\mathbf{G}\mathbf{G}^\top = \mathbf{U} (\Sigma \Sigma^\top) \mathbf{U}^\top \quad (5.9)$$

Then, the final solution in (5.8) can be rewritten by combining (5.7) and (5.9) as:

$$\tilde{\mathbf{G}} = \mathbf{U} \Sigma^{-1} \mathbf{U}^\top \mathbf{G} \quad (5.10)$$

The current $\tilde{\mathbf{G}}$ in (5.10) is orthogonal but with unit singular values, and the next step is to re-scale the task-specific gradients to avoid magnitude dominance. According to the literature [102], the original magnitude of task-specific gradients is proportional to the singular values of $\tilde{\mathbf{G}}$. Therefore, to ensure the convergence to the optima of all the tasks, the minimal singular value is selected to calculate the scaling factor instead of using the original singular values, and the re-scaled $\tilde{\mathbf{G}}$ can be obtained as:

$$\tilde{\mathbf{G}} = \tilde{\sigma} \mathbf{U} \Sigma^{-1} \mathbf{U}^\top \mathbf{G}, \quad \text{with } \tilde{\sigma} = \min(\sqrt{\text{eigenvalue}(\mathbf{G}\mathbf{G}^\top)}) \quad (5.11)$$

At last, the orthogonal gradient matrix with equal magnitude is shown in Figure 5.3(b), but all the tasks are currently compromised on the same learning rate, causing the stall of the optimization for certain hard tasks.

Eccentric Gradient Alignment

To estimate the intrinsic task difficulty, the first step is to assess the current learning rate lr_i based on the loss value \mathcal{L}_i of each task:

$$lr_i(t-1) = \frac{\mathcal{L}_i(t-1)}{\mathcal{L}_i(t_{\text{warm}})} \quad (5.12)$$

Algorithm 3 EGA Optimization Strategy for MTL

1: **Input:** Loss values for n tasks $[\mathcal{L}_1, \dots, \mathcal{L}_n], i \in [n]$, Shared parameters $\boldsymbol{\theta}$
and Step length η , T for softmax and t_{warm} for warmup epoch
 2: **Output:** Optimal parameters $\boldsymbol{\theta}^*$ for updating $\boldsymbol{\theta}$
OBJECTIVE:
 3: $\boldsymbol{\theta}^*$ such that $\boldsymbol{\theta}^* = \arg \min_{\boldsymbol{\theta} \in \mathbb{R}^m} \left\{ \mathcal{F}(\boldsymbol{\theta}) \triangleq \frac{1}{n} \sum_{i=1}^n \mathcal{L}_i(\boldsymbol{\theta}) \right\}$
 FOR THE INPUT BATCH IN CERTAIN EPOCH:
 4: - Initialize eccentric vector $\mathbf{v}_{ecc} = [1, \dots, 1]^\top \in \mathbb{R}^n$
 5: - Get the current epoch as t
 6: - Calculate task-specific gradient $\mathbf{g}_i = \nabla_{\boldsymbol{\theta}} \mathcal{L}_i(\boldsymbol{\theta}), i \in [n]$
 7: - Form gradient matrix $\mathbf{G} = \{\mathbf{g}_1, \dots, \mathbf{g}_n\} \in \mathbb{R}^{n \times m}$
 8: - Calculate eigenvalues/eigenvectors of Gram matrix as in (5.9):

$$\mathbf{G}\mathbf{G}^\top = \mathbf{U} (\Sigma \Sigma^\top) \mathbf{U}^\top \text{ with eigenvalues } \boldsymbol{\lambda}$$

 9: - Get scaling factor: $\tilde{\sigma} = \min(\sqrt{\boldsymbol{\lambda}})$
 10: - Calculate the orthogonal and normalized gradient matrix as in (5.11):

$$\tilde{\mathbf{G}} = \tilde{\sigma} \mathbf{U} \Sigma^{-1} \mathbf{U}^\top \mathbf{G}$$

 11: **if** $t = t_{warm}$ **then**
 12: - Record the loss values for all the tasks $\mathcal{L}_i(t_{warm})$
 13: **else if** $t > t_{warm}$ **then**
 14: - Calculate the intrinsic task difficulty as in (5.13):

$$k_i(t) = \text{softmax}(lr_i(t-1))$$

 15: - Form eccentric vector $\mathbf{v}_{ecc} = [k_1, \dots, k_n]^\top$
 16: **end if**
 17: - Calculate final joint gradient $\tilde{\mathbf{g}}_{ecc} = \tilde{\mathbf{G}}^\top \mathbf{v}_{ecc}$
 18: - Calculate optimal parameters $\boldsymbol{\theta}^* = \boldsymbol{\theta} - \eta \tilde{\mathbf{g}}_{ecc}$

with $\mathcal{L}_i(t-1)$ and $\mathcal{L}_i(t_{warm})$ representing the loss value for Task i at the previous epoch and the warmup epoch (e.g., $t_{warm} = 4$ in this paper), and the lr_i is inversely proportional to the learning rate (i.e., small lr_i for fast learning rate). Then, a softmax function is applied to mapping the lr_i to the intrinsic task difficulty k_i as:

$$k_i(t) = \text{softmax}(lr_i(t-1)) = \frac{n \exp(lr_i(t-1)/T)}{\sum_{j=1}^n \exp(lr_j(t-1)/T)} \quad (5.13)$$

with T controlling the discrepancy of the mapped task difficulties (i.e., small T enlarges the discrepancy between k_i), and the summation of the weights should be $\sum_{i=1}^n k_i = n$. In addition, the intrinsic task difficult k_i is positive without the negative transfer issue and can be formed as eccentric vector $\mathbf{v}_{ecc} = [k_1, \dots, k_i]^\top$ as in Figure 5.3(c) to guide the final joint gradient $\tilde{\mathbf{g}}_{ecc}$ for optimization as $\tilde{\mathbf{g}}_{ecc} = \tilde{\mathbf{G}}^\top \mathbf{v}_{ecc}$. At last, the optimal parameter set $\boldsymbol{\theta}^*$ for updating the shared parameter space can be obtained after providing a step length η based on the current parameter set $\boldsymbol{\theta}$ as $\boldsymbol{\theta}^* = \boldsymbol{\theta} - \eta \tilde{\mathbf{g}}_{ecc}$.

The entire EGA optimization strategy is summarized in Algorithm 3 to repeatedly update the shared parameter space (i.e., Backbone&Encoder in this work) based on all the batches in each epoch, and the optimization will be terminated until achieving a pre-defined epoch number.

5.4 Experimental Setting and Result Evaluation

5.4.1 Dataset and Implementation

Dataset for ECG Recovery

MMECG [1] is a dataset used for radar-based ECG recovery and is collected by TI AWR-1843 radar with 77GHz start frequency and 3.8GHz bandwidth with the scenario of data collection shown in Figure 4.7. A total of 91 trials for 11 subjects (8 males and 3 females) are included in the dataset, and each trial lasts for 3 minutes with synchronous ECG/radar signals sampled at 200Hz. Following the link budget analysis in [21, 53], the collected raw radar signals have a good SNR level of 37dB and are enough for later signal processing or model training.

All the subjects are healthy people without knowing diseases (e.g., premature ventricular contractions [36]) that may change the common ECG patterns and are asked to keep a quasi-static status to ensure good SNR with the least RBM noise. In addition, the ground ECG signal is collected by TI ADS1292 board with AC coupling and integrated right-leg drive amplifier to remove potential baseline drift or power-line noise.

Dataset for Evaluating EGA

NYUv2 [96] is a dataset for indoor scene understanding recorded using the RGB and Depth cameras and has been widely used as a unified task for validating MTL optimization strategies based on the performance of semantic segmentation, depth estimation, and surface normal prediction [102–109].

Implementation Details

The proposed radarODE-MTL along with the radarODE [2] and MMECG [1] are coded using PyTorch and trained on the NVIDIA RTX A4000 (16GB) for 200 epochs with SGD optimizer [93]. The hyperparameters used for training are empirically obtained as batch size 32, learning rate 5×10^{-3} , weight decay 5×10^{-4} and momentum 0.937. The dataset is split based on different subjects, with the trials from 1 fixed subject for testing and the other 10 subjects alternatively selected for training or validation (i.e., 11-fold cross-validation), ensuring to make use of all the possible trials while not involving the testing data in the training phase. At last, the Python package NeuriKit2 [80] is applied to all the evaluations regarding ECG signals, such as the identification of single cardiac cycles, PQRST peaks detection and heart rate estimation.

The deep learning framework used for scene understanding is implemented in [94] with many popular MTL optimization strategies embedded for comparison and optimal hyperparameters provided for training. The training is on the same GPU as before with 200 epochs, batch size 18, Adam optimizer [114], learning rate 10^{-4} and weight decay 10^{-5} .

5.4.2 Performance of EGA

Radar-based ECG Recovery

The performance of EGA is evaluated on three tasks in terms of different metrics: RMSE, PCC and coefficient of determination R^2 for the recovered single-cycle ECG pieces; absolute PPI Error for the cycle lengths estimation; and absolute Timing

Table 5.1: Comparison of different optimization strategies on radar-based ECG recovery

Methods	Tasks	ECG Shape Recovery			Cycle Length Estimation		ECG Anchor Estimation		$\Delta m\% \uparrow$	P Value ($\times 10^{-2}$)
		RMSE (mV) \downarrow	PCC \uparrow	$R^2 \uparrow$	PPI Error (ms) \downarrow	Timing Error (ms) \downarrow	MDR \downarrow			
Single-task baseline		0.106	86.6%	0.81	9.6	7.5	6.67%	0.00 \pm 1.43	-	
Loss Balancing Methods										
Equal Weight		0.125	79.7%	0.63	8.0	9.7	5.51%	-1.78 \pm 2.16	9.26	
UW [109]		0.066	88.5%	0.85	11.2	5.5	6.44%	4.04 \pm 3.79	2.43	
GLS [108]		0.087	87.3%	0.81	14.1	6.7	4.32%	-5.89 \pm 2.02	0.02	
DWA [107]		0.133	80.7%	0.79	8.3	6.4	5.33%	6.45 \pm 3.71	0.20	
STCH [106]		0.070	88.0%	<u>0.86</u>	13.9	5.5	3.28%	2.90 \pm 3.21	5.12	
Gradient Balancing Methods										
CAGrad [105]		0.107	84.2%	0.79	10.2	6.2	3.98%	6.84 \pm 2.12	0.01	
IMTL [104]		0.088	<u>89.4%</u>	<u>0.86</u>	9.3	6.0	6.22%	8.43 \pm 1.39	0.00	
MoCo [103]		0.179	61.0%	0.66	8.7	6.8	4.27%	-2.32 \pm 1.37	1.16	
Aligned-MTL [102]		0.092	87.9%	0.84	10.0	6.9	3.52%	10.14 \pm 1.11	0.00	
EGA ($T = 0.1$)		0.119	79.0%	0.72	10.6	6.8	3.34%	2.83 \pm 0.98	0.19	
EGA ($T = 0.5$)		<u>0.082</u>	89.6%	0.87	9.9	6.3	4.19%	<u>11.55</u> \pm 1.44	0.00	
EGA ($T = 1.0$)		0.085	87.4%	0.85	8.5	7.2	4.31%	13.37 \pm 1.36	0.00	
EGA ($T = 1.5$)		0.105	82.9%	0.78	<u>8.1</u>	6.3	5.13%	10.94 \pm 1.30	0.00	
EGA ($T = 2.0$)		0.091	86.3%	0.84	9.2	7.3	4.01%	10.43 \pm 0.95	0.00	

Bold and underline represent the best and the second best results, respectively.

Error and missed detection rate (MDR) for the anchors prediction, with the corresponding comparison across other MTL optimization strategies as shown in Table 5.1. In addition, all the experiments are repeated five times, and the last column $\Delta m\%$ in Table 5.1 shows a comprehensive assessment across n tasks with 95% confidence interval (CI) and is calculated as:

$$\Delta m\% = \frac{1}{n} \sum_{i=1}^n \frac{1}{n_i} \sum_{j=1}^{n_i} S_{i,j} \frac{M_{m,i,j} - M_{b,i,j}}{M_{b,i,j}} \times 100\% \quad (5.14)$$

where n_i is the number of metrics for task i , $M_{m,i,j}$ means the performance of a method m on the task i measured with the metric j , $M_{b,i,j}$ represents the performance for the single-task baseline, and $S_{i,j} = 1/0$ if lower/higher values are better for the current metric (indicated by \downarrow / \uparrow). Lastly, the T-test is adopted with the P -value calculated for all the experiments as shown in Table 5.1, and the statistical analysis will be given at the end of each subsection.

In general, the effect of unbalanced task-specific gradients is revealed by using

equal weight as shown in Table 5.1, and the performance of the hard task (ECG shape recovery) is much worse than baseline while the PPI error even achieves the best accuracy. After balancing the magnitudes and directions of task-specific gradients, the proposed EGA strategy meets the expectation by adjusting the value of T with the following evaluations:

- EGA with $T = 1.0$ achieves the largest improvement with $\Delta m\% = 13.37$ but none of the individual metrics gets the best or second-best result, and $T = 1.0$ can be viewed as a suitable estimation of intrinsic task difficulty to achieve unbiased improvements across all tasks.
- EGA with $T = 0.5$ obtains the second-best overall performance with $\Delta m\% = 11.55$ and becomes the best in learning ECG morphological features according to RMSE/PCC/ R^2 , indicating $T = 0.5$ slightly overrates the difficulty of Task 1.
- EGA with $T = 0.1$ cannot balance the task difficulties, hence getting a low score.
- EGA with large T values (1.5 and 2.0) tend to evenly distribute the task difficulty weights, and the performance should be similar to other orthogonality-based methods (e.g., Aligned-MTL).

In addition, it is also worth noticing that some methods achieve a significant improvement on a particular task, e.g., UW obtains RMSE = 0.066mV and PPI Error = 5.5ms, implying a potential improvement probably by enlarging the parameter space (scaling the model size) or designing a more efficient MTL architecture instead of using simple HPS [115]. However, the method with remarkable performance on the single task cannot achieve unified improvement on other tasks, e.g., UW and STCH both get good results in ECG anchor estimation (Timing Error = 5.5ms), but a huge degradation happens on the cycle length estimation (PPI Error > 10ms), revealing the effectiveness of EGA to avoid overvaluing one certain task.

Lastly, when comparing EGA to the method also based on orthogonality, Aligned-MTL stalls after Task 3 achieves convergence (low $MDR = 3.52\%$), while EGA ($T =$

Table 5.2: Comparison of different optimization strategies on indoor scene understanding

Methods	Tasks		Segmentation \uparrow		Depth Estimation \downarrow		Surface Normal Prediction					$\Delta m\% \uparrow$	P Value $(\times 10^{-2})$		
			mIoU	Pixel Acc.	Abs. Err.	Rel. Err.	Angle Distance \downarrow		Within $t^\circ \uparrow$						
							Mean	Median	11.25	22.5	30				
Single-task baseline	52.08	74.11	0.4147	0.1751	23.83	17.36	34.34	60.22	71.47	0.00±0.19	-				
Loss Balancing Methods															
Equal Weight	53.36	74.94	0.3953	0.1672	24.35	17.55	34.22	59.64	70.71	1.75±1.60	1.67				
UW [109]	<u>53.33</u>	75.43	<u>0.3878</u>	0.1639	24.03	17.24	34.80	60.33	71.31	2.92±2.05	0.44				
GLS [108]	53.04	74.68	0.3951	0.1600	24.03	17.30	34.78	60.17	71.28	2.69±1.52	0.12				
DWA [107]	53.12	<u>75.23</u>	0.3883	0.1615	24.26	17.60	34.25	59.51	70.62	2.55±1.91	0.62				
STCH [106]	52.87	74.78	0.3915	0.1615	23.27	<u>16.34</u>	36.61	<u>62.33</u>	72.98	3.99±0.61	0.00				
Gradient Balancing Methods															
CAGrad [105]	52.19	74.07	0.3976	0.1634	23.83	17.16	34.89	60.65	71.77	2.09±1.11	0.09				
IMTL [104]	52.34	74.35	0.3897	<u>0.1579</u>	23.76	17.00	35.28	60.92	71.89	3.24±0.78	0.00				
MoCo [103]	52.78	74.59	0.3858	0.1612	23.34	16.51	36.21	61.90	72.65	3.94±0.72	0.00				
Aligned-MTL [102]	52.19	74.17	0.3911	0.1605	23.44	16.73	35.45	61.74	72.70	3.24±1.08	0.00				
EGA ($T = 0.1$)	52.16	74.23	0.3944	0.1651	23.32	16.62	35.87	61.81	72.72	2.84±0.88	0.00				
EGA ($T = 0.5$)	51.82	73.98	0.3904	0.1614	23.41	16.66	35.87	61.65	72.51	3.11±0.51	0.00				
EGA ($T = 1.0$)	51.75	74.38	0.3913	0.1609	23.09	16.29	<u>36.54</u>	62.51	73.22	3.71±1.08	0.00				
EGA ($T = 1.5$)	52.37	74.65	0.3950	0.1571	23.15	16.46	36.07	62.22	73.07	<u>3.96±0.98</u>	0.00				
EGA ($T = 2.0$)	52.18	74.23	0.3922	0.1605	23.28	16.61	35.77	61.95	72.86	3.39±1.18	0.00				

Bold and underline represent the best and the second best results, respectively.

1.0) keeps improving Task 1 and 2 and gets a better result on RMSE = 0.085mV and PPI Error = 8.5ms with only a slight degradation on Task 3 (Timing Error = 7.2ms, MDR = 4.31%), showing the ability of EGA to focus on the hard task without distracted by the well-trained easy tasks.

Indoor Scene Understanding

The indoor scene understanding based on NYUv2 is a commonly adopted task by all the studies about MTL optimization strategies [96]. The metrics for each task are: mean intersection over union (mIoU) and pixel accuracy (Pixel Acc.) for segmentation, absolute/related error (Abs./Rel. Err.) for depth estimation and mean/median angle distance, and the percentage of surface normal within t° for surface normal prediction, as shown in the heads of Table 5.2.

According to the improvements $\Delta m\%$ in Table 5.2, EGA ($T = 1.5$ and 1.0) achieves a competitive result compared with other powerful methods, indicating that EGA can be applied to other MTL tasks with an appropriate selection of T . An interesting observation is that some methods with average or even poor performance

in Table 5.1 (i.e., STCH and MoCo) achieve remarkable results in scene understanding. A possible explanation is that the indoor scene understanding task may have a small discrepancy in task difficulties and fewer conflicts in gradient directions. This guess can also be verified by the fact that loss balancing methods achieve competitive performance compared with gradient balancing methods, and different T values have limited impacts on the final performance of EGA.

Statistical Analysis for EGA: To verify the significance of the EGA performance, the well-known T-test is performed with the null hypothesis as the performances of the compared methods are identical. The yielded P values for all the methods are shown in Table 5.1 and 5.2, and all the methods achieve $P < 0.05$ except for using the equal weights, indicating that the obtained mean values are reliable with significance. In addition, it is worth noticing that the CIs of the loss balancing methods are larger than that of the gradient balancing methods for both tasks as shown in Table 5.1 and 5.2, coinciding with the previous conclusion that the access of the gradients is beneficial to the stability of model training. Lastly, the proposed EGA not only achieves the best performance in ECG recovery, but the obtained results are also stable with small CIs, owing to the orthogonal projection that decreases the condition number of the gradient system [102].

To conclude the above evaluations in terms of different tasks, the proposed EGA could successfully alleviate the gradient conflicts and magnitude dominance in MTL optimization, while the intrinsic task difficulty can be successfully estimated to guide the optimization direction by introducing eccentric vector \mathbf{v}_{ecc} . Compared with other methods, EGA achieves an outstanding result for the tasks with disparate difficulties and is also competitive in the common tasks, but the hyperparameter T should be carefully selected. During practice, the T value can be adjusted until achieving optimum based on the fact that large T evenly treats all the tasks and small T enhances the hard task.

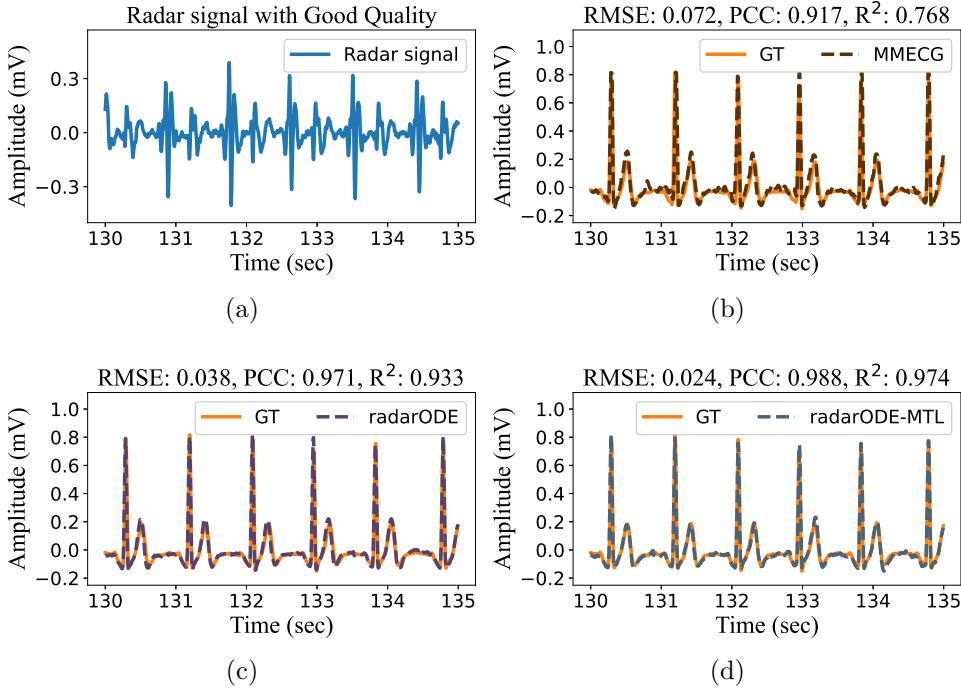


Figure 5.4: Illustration of the recovered ECG with ground truth (GT) for different frameworks under no noise:(a) - (d) Radar signal with good quality and the recovered ECG signals.

5.4.3 Evaluations on the Long-term Recovered ECG

General Visualization for ECG Reconstruction

The outputs from Task 1 – 3 can form the long-term ECG signal as depicted in Figure 5.4. All three frameworks successfully reconstruct the ECG signals from high-quality radar signals as shown in Figure 5.4(b), 5.4(c) and 5.4(d), only with certain fluctuations in MMECG result and causing low RMSE/PCC/ R^2 .

In the presence of constant or abrupt noise, the signal SNR will decrease with the subtle features (e.g., v_2) ruined as shown in Figure 5.5(a) and 5.6(a). MMECG shows the least noise robustness and cannot resist abrupt noise as also reported in the benchmark paper [1] as shown in Figure 5.5(b) and 5.6(b), while radarODE achieves robust ECG recovery within each single cardiac cycle but shows obvious misalignment

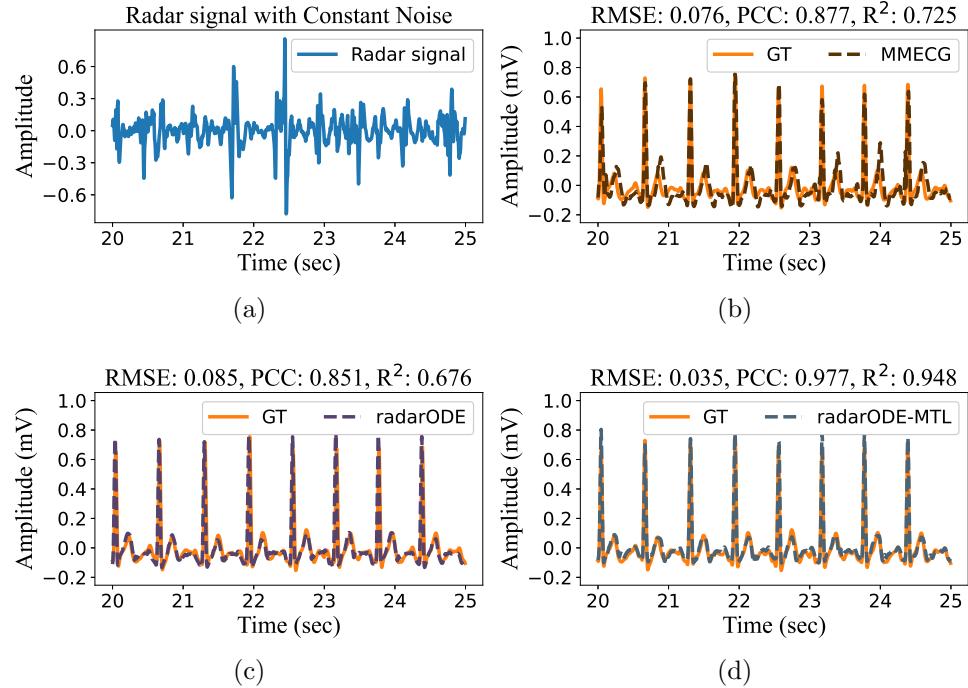


Figure 5.5: Illustration of the recovered ECG with ground truth (GT) for different frameworks under constant noise:(a) - (d) Radar signal with constant noise and the recovered ECG signals.

due to the PPI estimation error as shown in Figure 5.5(c) and 5.6(c). Lastly, the proposed radarODE-MTL realizes the ECG reconstruction in an end-to-end manner without reintroducing the noises, and the recovered ECG is less corrupted by the noises as shown in Figure 5.5(d) and 5.6(d).

Corrupt ECG Reconstruction

The successful reconstruction in Figure 5.6(d) owes to the design of radarODE-MTL with deconstructed tasks for ECG recovery. Different from other frameworks with equal length of input and output, radarODE-MTL adopts a 4-sec segment to reconstruct the ECG piece for one cardiac cycle, and the radar signal from adjacent cardiac cycles (e.g., the zoomed part in Figure 5.6(a)) also contributes to the recovery of the

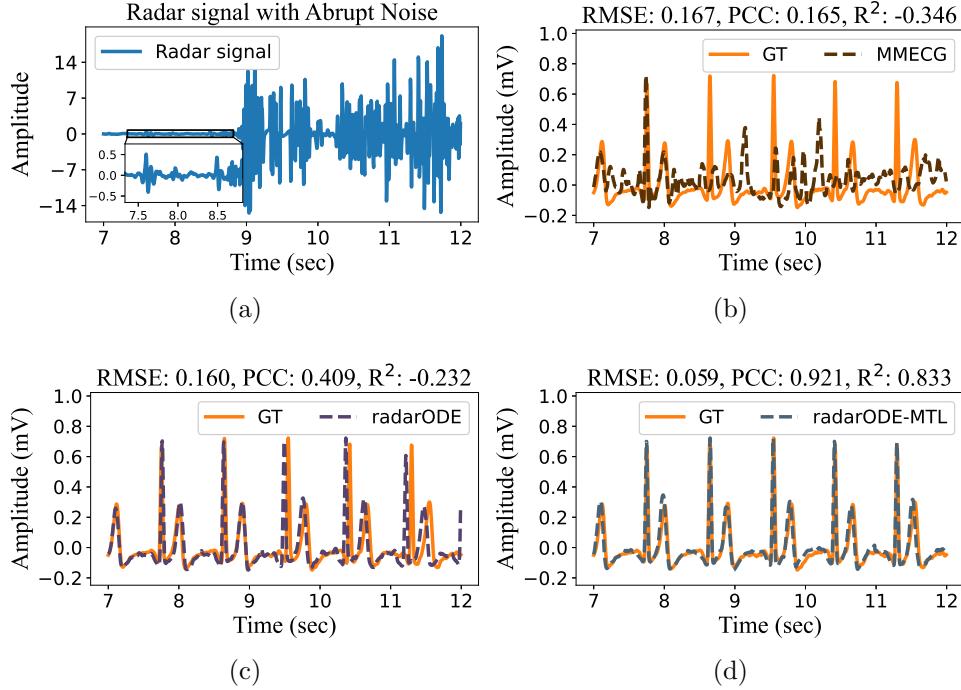


Figure 5.6: Illustration of the recovered ECG with ground truth (GT) for different frameworks under abrupt noise: (a) - (d) Radar signal with abrupt noise and the recovered ECG signals.

current ECG piece. In addition, if the input radar signal is fully destroyed by noise, radarODE-MTL may fail to extract any information, and the failures can be revealed by the MDR to statistically evaluate the corruptions in recovered ECG signals due to noise distortion.

The result of MDR is shown as the cumulative distribution function (CDF) in Figure 5.7(a) with the median MDR as 1.7%, 0.13% and 0.13% for MMECG, radarODE and radarODE-MTL respectively, and $\Delta m\%$ across 91 trials are both 14%. The reason for the similar performance of two ODE-based methods is that the misaligned ECG pieces with small deviations ($< 150\text{ms}$) in radarODE will not be identified as ‘missed detected’, and hence the CDFs of MDR share a similar pattern and trend in Figure 5.7(a).

Coarse Cardiac Feature Reconstruction

All three frameworks evaluated in this chapter are designed for fine-grained cardiac features reconstruction and should perform well on the coarse cardiac feature (i.e., HR monitoring). The result in Figure 5.7(b) coincides with the expectation with median HR error as 0.6, 0.3 and 0.3 beats/min respectively, and $\Delta m\%$ for the ODE-based methods are 54% and 59%. It is notable in Figure 5.7(b) that the performances of ODE-based methods are very similar at the beginning, while the radarODE tends to get more errors when the noise in the raw radar signal affects the R peaks recovery, because the calculation of HR is based on the R peak positions.

Fine-Grained Morphological Feature Reconstruction

The morphological feature is an essential fine-grained feature to describe the general similarity between the recovered and ground truth ECG signals, and the morphological accuracy can be evaluated by RMSE, PCC and R^2 , with RMSE sensitive to the peak deviation, PCC focusing on the similarity of the general shape and R^2 shows the interpretability of the well-trained neural network. The results are shown in Figure 5.7(c), 5.7(d) and 5.7(e) as the CDF of RMSE/PCC/ R^2 across 91 trials in the dataset, and three frameworks get the median RMSE/PCC as 0.125mV/82.1%/0.74, 0.098mV/90.1%/0.81 and 0.083mV/92.7%/0.85 respectively.

As indicated by $\Delta m\%$, the improvements of RMSE (28%, 33%) are larger than PCC (18%, 21%) for radarODE and radarODE-MTL respectively, because the ODE model embedded in the decoder preserves the main features of ECG even under noises and contributes more on the peaks than on the shapes. In addition, radarODE-MTL further improves the results by aligning the ECG pieces with the predicted anchors, avoiding the misalignment issue in radarODE. Lastly, the resultant improvements in R^2 (17% and 19% in Figure 5.7(e)) indicate that radarODE-MTL could capture more dependency in the domain transformation of cardiac activities due to the induction of ODE model as prior knowledge, but the improvements are less than the other two metrics for morphological assessment, because R^2 is not sensitive to the outliers and could objectively evaluate the model ability.

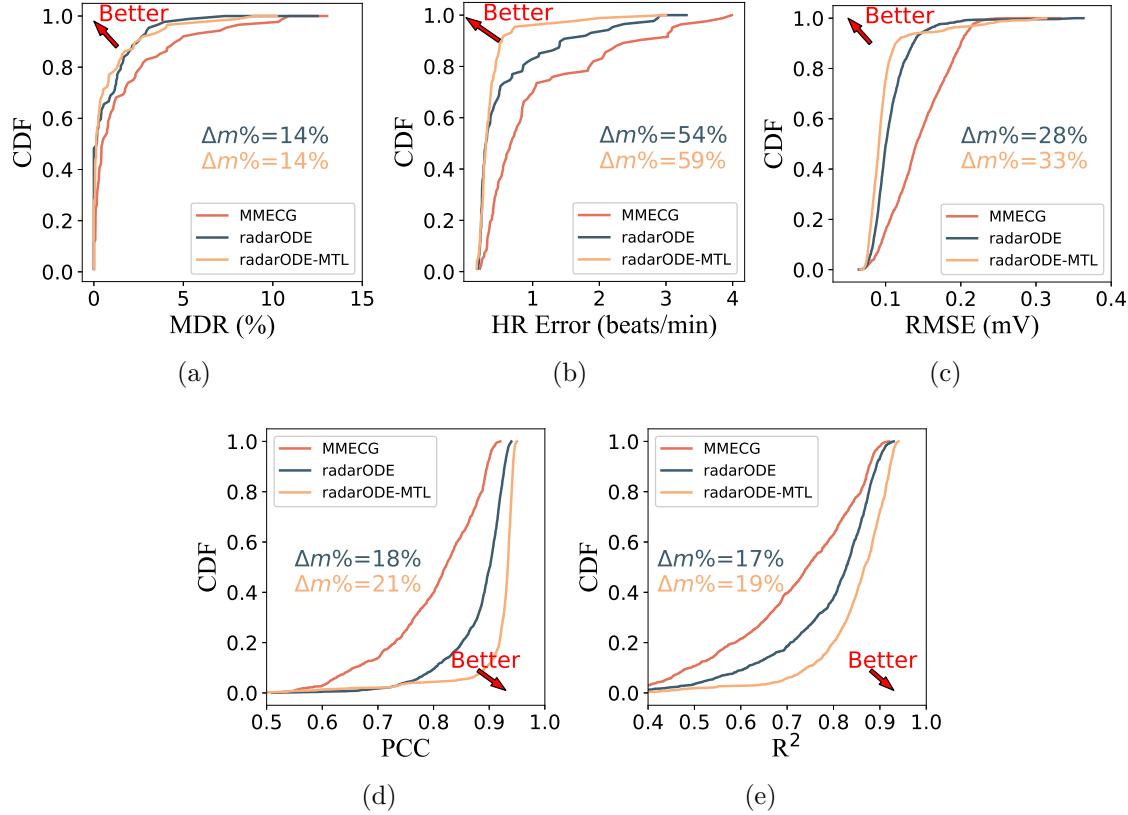


Figure 5.7: Evaluations for long-term ECG recovery: (a) - (e) CDF plots of MDR, HR Error, RMSE, PCC and R^2 , with corresponding improvements.

Fine-Grained ECG Peaks Reconstruction

In the evaluations of timing errors of the ECG peaks it is common only to analyze QRST peaks because the inconspicuous P peaks can be miss-detected even in some ground truth signals [1, 2]. The CDF plots for the absolute timing errors of QRST peaks are shown in Figure 5.8 with the following observations:

- Both ODE-based methods reveal better performance than the benchmark, but the radarODE-MTL only achieves equivalent performance as radarODE with similar $\Delta m\%$ around 31%, 35% and 24% as shown in Figure 5.8(a), 5.8(c) and 5.8(d). The possible reason is that radarODE-MTL only aligns the ECG

pieces with R peaks, but the impacts on the QST peaks are random. In other words, the alignment of the R peak may degrade the accuracy of other peaks, and hence the overall performance of radarODE and radarODE-MTL on the QST peaks are similar.

- It is worth noticing that $\Delta m\%$ of the radarODE-MTL (33%) on the R peak is obviously larger than that of the radarODE (25%), with the median timing error as 14, 10 and 6ms for three frameworks as shown in Figure 5.8(b). Therefore, radarODE-MTL is a better way to generate long-term ECG signals by aligning the ECG pieces with predicted R peaks, instead of reintroducing the noisy time-domain radar signal as in radarODE.

5.4.4 Noise Robustness Test

In this work, 10 trials (No. 75–84) are selected for the noise robustness test by adding different types of synthesized noises with certain decibel (dB) only in the test stage, while the training and validation stage will use the original data because adding noises into the training dataset is a data augmentation technique to improve the model performance, causing an unfair comparison in the noise robustness test [88]. In addition, adding noises into the validation stage is equivalent to selecting appropriate models for the scenarios with different SNR levels and cannot prove the noise robustness of the proposed radarODE-MTL.

Constant Noise

The constant noise normally affects the SNR of the signal and could be caused by thermal noise from electronic components or long-range detection [27, 46], e.g., the SNR for the current data collection scenario will decrease to 0dB by increasing the monitoring distance to 5m. In the literature, low SNR scenarios can be simulated by adding Gaussian noise with different intensities as implemented in [24, 27, 37, 116]. The baseline results for three frameworks are firstly obtained in terms of the RMSE, PCC, R^2 , R-peak error and MDR as shown in Table 5.3, and $\Delta m\%$ is calculated

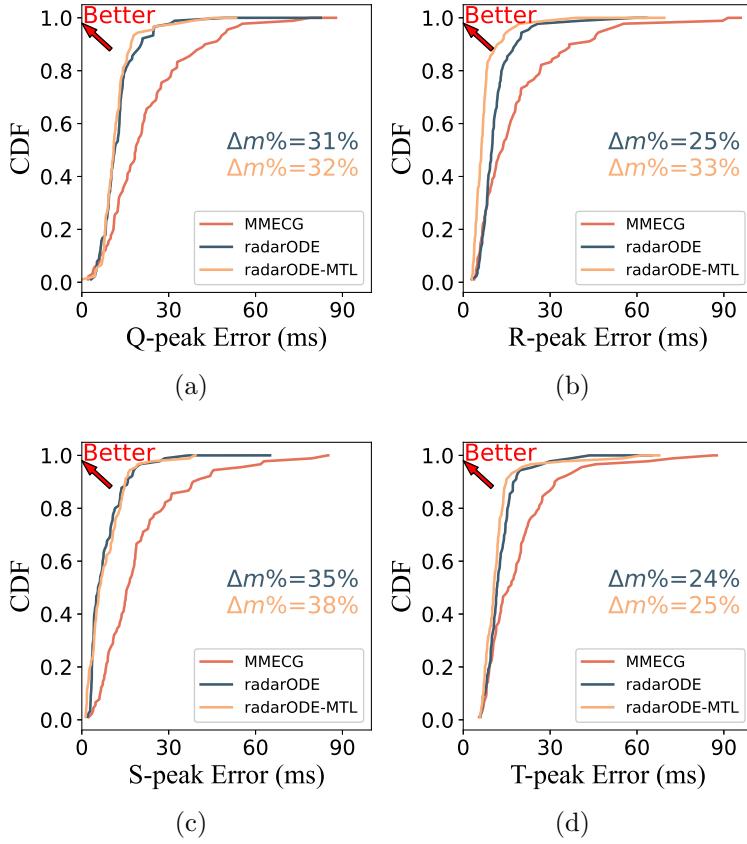


Figure 5.8: Evaluations for fine-grained ECG peaks recovery: (a) - (d) CDF plots of the timing error for QRST peaks, with corresponding improvements.

as 0%, 7.47% and 10.64% as indicated by the initial points in Figure 5.9. Then, the Gaussian noises with 6 to -3dB are added into the raw radar signal without retraining the deep-learning framework, and the results are shown in Table 5.3 with the trends of performance degradation shown in Figure 5.9.

A general observation of Table 5.3 is that all the frameworks perform well before 0dB with a similar degradation rate as in Figure 5.9(a). Then, radarODE-MTL could still provide reasonable results with mild degradation after 0dB because the MTL paradigm split the ECG reconstruction task into several sub-tasks, and each task can either be constrained by prior knowledge or leverage the information from context

Table 5.3: Comparison of the frameworks under different SNR

SNR	RMSE (mV) ↓	PCC ↑	R^2 ↑	Peak Error (ms) ↓	MDR ↓	$\Delta m\%^1$ ↑	P Value ($\times 10^{-2}$)
MMECG [1]							
Baseline	0.107	83.75%	0.77	9.45	4.52%	0.00 ± 0.55	-
6 dB	0.107	82.60%	0.76	9.76	4.37%	-0.28 ± 1.68	73.85
3 dB	0.108	82.64%	0.76	9.85	4.84%	-4.17 ± 1.91	0.04
0 dB	0.109	80.00%	0.74	11.80	4.92%	-12.38 ± 3.18	0.00
-1 dB	0.114	78.55%	0.69	12.20	5.32%	-18.17 ± 4.10	0.00
-2 dB	0.120	74.32%	0.65	14.64	5.59%	-30.53 ± 3.78	0.00
-3 dB	0.127	62.45%	0.54	21.28	6.40%	-63.81 ± 2.15	0.00
radarODE [2]							
Baseline	0.091	83.53%	0.79	9.08	4.03%	0.00 ± 0.39	-
6 dB	0.093	83.30%	0.78	9.12	4.36%	-3.29 ± 1.71	0.11
3 dB	0.095	83.01%	0.76	9.01	4.70%	-6.26 ± 1.43	0.00
0 dB	0.101	82.21%	0.69	9.89	5.86%	-20.91 ± 2.09	0.00
-1 dB	0.116	79.66%	0.63	11.90	5.36%	-27.17 ± 2.92	0.00
-2 dB	0.157	70.87%	0.58	13.95	6.19%	-48.44 ± 3.53	0.00
-3 dB	-	-	-	-	-	Failed ²	-
radarODE-MTL							
Baseline	0.089	85.03%	0.81	8.22	4.08%	0.00 ± 1.24	-
6 dB	0.088	85.31%	0.82	8.18	4.20%	-0.52 ± 0.72	19.81
3 dB	0.089	84.29%	0.80	8.31	4.27%	-2.15 ± 1.55	0.78
0 dB	0.091	83.77%	0.79	8.03	4.76%	-5.47 ± 2.59	0.05
-1 dB	0.093	84.01%	0.78	8.10	5.10%	-8.89 ± 1.21	0.00
-2 dB	0.093	84.51%	0.79	8.02	5.45%	-11.22 ± 1.43	0.00
-3 dB	0.094	84.96%	0.78	8.19	6.02%	-16.77 ± 1.60	0.00

1. $\Delta m\%$ is calculated for each framework based on each baseline.

2. The ECG recovery fails if $PCC < 60\%$, according to the empirical observation of the morphological ECG features.

data with less pollution. In contrast, radarODE could generate high-fidelity ECG pieces as claimed in [2] and gets the second best baseline result in Table 5.3, but the design of PPI estimation stage does not consider the noise robustness. Therefore, the performance is heavily dropped to the worst in Figure 5.9(a) because of the bad results of Peak Error as shown in Table 5.3. Lastly, the MMECG considers the ECG recovery as an arbitrary domain transformation problem without any constraints in the network design, and the performance also heavily degrades in Figure 5.9(a) because only meaningless results will be generated as shown previously in Figure 5.6(b).

Abrupt Noise

In this part, the Gaussian noises with different intensities (0 and -9dB) are used to simulate mild body movement (e.g., during talking or writing) and extensive body

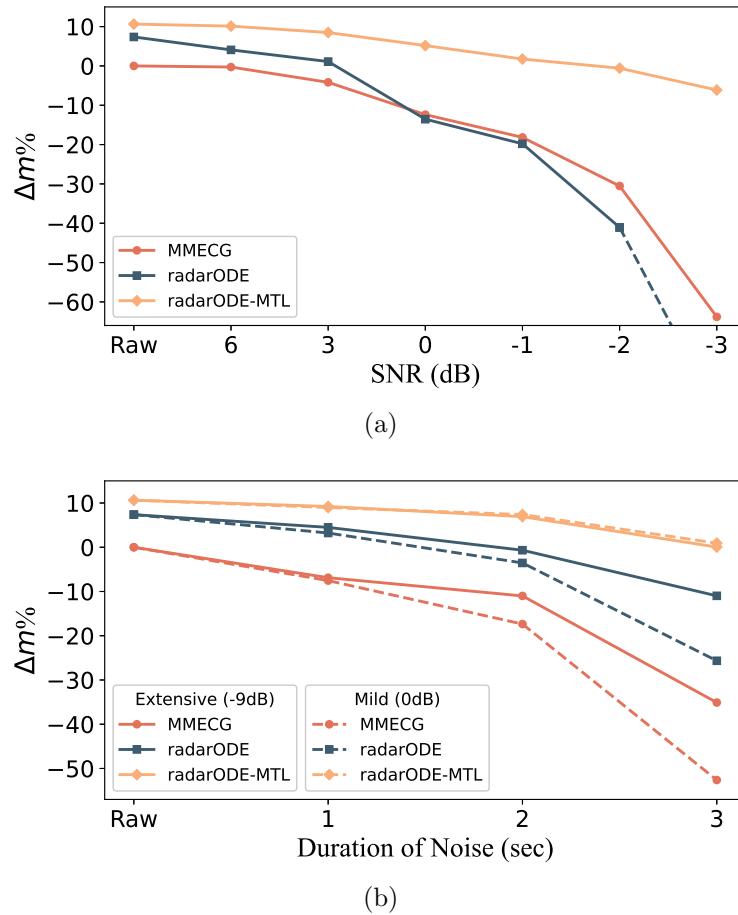


Figure 5.9: Noise robustness test: (a) Impact of constant noises with different intensities, (b) Impact of abrupt noises with different intensities and durations.

movement (e.g., during torso movement) as suggested in the literature [33]. In practice, the body movements have orders of magnitude larger than cardiac activities to ruin the cardiac activities, and the ability of radarODE-MTL to recover ECG signal during RBM comes from the contextual information provided by previous cardiac cycles without RBM noise. Only 20% of the segments randomly selected from one trial are doped, and the duration of noise varies from 1 to 3 sec.

For mild body movement, the experimental results are shown in Table 5.4 with the changes of $\Delta m\%$ shown in Figure 5.9(b). Firstly, it is evident that the impact of

Table 5.4: Comparison of the frameworks under abrupt noises

Duration	RMSE (mV) ↓	PCC ↑	R^2 ↑	Peak Error ↓ (ms)	MDR ↓	$\Delta m\%^1 \uparrow$	RMSE (mV) ↓	PCC ↑	R^2 ↑	Peak Error ↓ (ms)	MDR ↓	$\Delta m\% \uparrow$
MMECG [1]:												
Baseline	0.107	83.75%	0.77	9.45	4.52%	0.00±0.55	0.107	83.75%	0.77	9.45	4.52%	0.00±0.55
1 sec	0.107	85.53%	0.78	10.84	4.82%	-6.88±2.16	0.107	84.05%	0.76	10.93	4.82%	-7.54±1.42
2 sec	0.110	82.64%	0.75	11.31	5.02%	-11.00±1.68	0.108	79.01%	0.68	12.31	5.23%	-17.36±2.06
3 sec	0.114	76.87%	0.66	15.56	5.92%	-35.10±2.91	0.116	75.09%	0.59	12.50	9.56%	-52.61±2.15
radarODE [2]:												
Baseline	0.091	83.53%	0.79	9.08	4.03%	0.00±0.39	0.091	83.53%	0.79	9.08	4.03%	0.00±0.39
1 sec	0.091	83.49%	0.79	9.12	4.36%	-2.88±1.81	0.095	82.96%	0.73	9.15	4.33%	-4.15±0.99
2 sec	0.092	83.39%	0.78	9.82	4.64%	-8.04±1.50	0.098	82.16%	0.70	9.31	4.97%	-10.92±1.09
3 sec	0.095	83.01%	0.75	10.01	5.70%	-18.35±2.19	0.102	81.87%	0.68	9.66	7.39%	-33.03±2.05
radarODE-MTL:												
Baseline	0.089	85.03%	0.81	8.22	4.08%	0.00±1.24	0.089	85.03%	0.81	8.22	4.08%	0.00±1.24
1 sec	0.090	84.62%	0.80	7.87	4.42%	-1.67±1.36	0.090	84.31%	0.80	8.28	4.18%	-1.42±0.94
2 sec	0.090	84.78%	0.82	8.29	4.44%	-3.25±1.10	0.091	84.21%	0.79	8.32	4.41%	-3.73±1.21
3 sec	0.091	84.44%	0.78	8.34	5.12%	-9.72±1.02	0.095	84.17%	0.77	8.43	5.10%	-10.60±1.85

1. $\Delta m\%$ is calculated for each framework based on the corresponding baseline.

1-sec abrupt noise is limited for all the frameworks, and the results for ODE-based methods are almost equivalent to the baselines. Secondly, 2-sec noise starts to have a noticeable impact on MMECG, while the ODE-based methods could preserve the performance on the morphological features (RMSE/PCC/ R^2) with small degradation on the Peak Error and MDR. Lastly, 3-sec noise has distorted 3/4 of the input radar segment, and the performances of MMECG and radarODE drop obviously as shown in Figure 5.9(b), while radarODE-MTL only loses some points on MDR = 5.12% as shown in Table 5.4.

In comparison, the extensive body movements with 1 and 2 sec have similar impacts with mild ones on ODE-based methods, because the ODE decoder could preserve the ECG shape even under strong noises, whereas the segments affected by noise cannot contribute to the recovery for MMECG as evident by the significant drop of PCC (from 84.05% to 79.01%) as shown in Table 5.4. In addition, the 3-sec noise destroys the ECG recovery for MMECG and radarODE with a significant degradation as shown in Figure 5.9(b), whereas the radarODE-MTL only sacrifices certain RMSE and peak accuracy with the overall degradation dropping slightly from -9.72% to -10.60% as shown in Table 5.4.

Statistical Analysis for Noise Robustness Test: The same T-test is also

implemented for the noise robustness test as shown in Table 5.3 and 5.4. For constant noise, 6dB noise does not have a significant impact on the performance of MMECG and radarODE-MTL with $P > 0.05$, while the performance of radarODE degrades because of the PPI error accumulation [2]. In addition, all the P values for the experiments of abrupt noise are less than 0.05 and are not listed in Table 5.4. Lastly, the impact of the noise level or duration is statistically significant, and the proposed radarODE-MTL could significantly improve the noise robustness as shown in Table 5.3 and 5.4, because the CIs of all $\Delta m\%$ have no overlapping with MMECG or radarODE under the same noise.

In summary, the noise-robustness tests indicate that it is necessary to consider the noise robustness when designing the deep-learning model, because both MMECG and radarODE reveal a severe degradation in the performance, especially for the low SNR scenarios. In addition, the deconstruction of the ECG recovery task in radarODE-MTL could effectively resist the noises, because the ODE decoder protects the morphological feature, and the peak accuracy can be compensated from the adjacent cardiac cycles with less noise distortion.

5.5 Conclusions

This chapter investigates the radar-based ECG monitoring technique and proposes a deep-learning framework radarODE-MTL to provide accurate ECG monitoring under noises. The radarODE-MTL adopts the MTL paradigm to realize the ECG reconstruction through 3 sub-tasks, and a novel optimization strategy called EGA is also proposed to simultaneously optimize all the tasks without stall or negative transfer issues. The performance of EGA has been evaluated on various MTL tasks, and the experimental results evidence that EGA is competitive with other state-of-the-art optimization strategies on the unified task and achieves outstanding results on radar-based EGA recovery with unbalanced task difficulties. In addition, the well-trained radarODE-MTL could provide long-term ECG reconstructions with high fidelity in terms of MDR, morphological similarity and peak accuracy. Lastly, this is the first study that conducts noise-robustness tests for deep-learning frameworks, and

the proposed radarODE-MTL could also achieve reasonable ECG recovery with mild degradation under constant and abrupt noises.

Chapter 6

Alleviate Data Scarcity for Radar-Based ECG Recovery

This chapter focuses on the data scarcity in the training of deep neural network (DNN) for radar-based ECG recovery. In the previous research, most studies either develop advanced signal processing methods to improve SNR [4, 24] or leverage various deep learning frameworks to realize accurate ECG recovery [37, 40]. However, the method for reducing dependence on data quantity is rarely investigated for radar-based ECG recovery, and all the deep-learning-based ECG recovery models are trained in a supervised manner [1, 30, 37]. In this chapter, a transfer learning framework is proposed following self-supervised learning (SSL) paradigm to extract cardio-related information from the radar signal without ECG ground truth based on the intrinsic sparsity of cardiac features, and only a few synchronous radar-ECG pairs are required to fine-tune the pre-trained model for the ECG recovery. In addition, a data augmentation method is designed to increase the diversity of input samples while the augmented spectrogram is still faithful to the original ground truth vital sign.

6.1 Introduction

The trials on the radar-based ECG recovery can be categorized into two paradigms. The first paradigm only performs the extraction of high-resolution cardiac mechanical activities to produce quasi-ECG signals, omitting the morphological ECG features while maintaining certain fine-grained features. For example, the mostly adopted quasi-ECG signal only preserves R and T peaks and can be realized by signal decomposition [27] or state estimation [13, 73]. In contrast, the second paradigm aims to reconstruct the ECG waveform as measured by clinical apparatus, because the doctor and ECG analysis toolbox all rely on the shape of ECG to make diagnosis [80]. However, decoupling the ECG signal from the measured radar signal requires establishing an extremely complex model from the perspective of electrophysiology (i.e., excitation-contraction coupling [23]), and the existing research can only leverage deep learning methods to learning such domain transformation from the dataset containing numerous radar/ECG pairs [1, 2, 35, 36].

Similar to all the research fields involved with deep learning, radar-based ECG recovery also asks for numerous radar signals to train the deep learning model with synchronous ECG ground truths [30, 40, 41]. However, data scarcity is an inevitable issue, because radar system requires expertise in configuration setting and signal pre-processing and the ECG collection requires cumbersome procedures, causing difficulties for the deployment in new scenarios with limited data.

In the literature, data augmentation and SSL are two popular methods to alleviate data scarcity, but there is no existing work that applies them in radar-based ECG recovery. Firstly, most augmentation techniques are designed for classification problems, and the inputs (signal or image) can be safely masked/recombined/flipped/stretched without changing the ground truth (labeled class). However, the aforementioned techniques are not applicable to the regression-based vital sign reconstruction (e.g. radar-based ECG recovery), because the changes on the input cannot be directly applied on the ground truth ECG, e.g., the real ECG peaks have a similar width, therefore randomly stretching the ECG signals is unfaithful to the biological facts [23]. Secondly, SSL, as another popular technique for reducing dependence on data quantity, is rarely

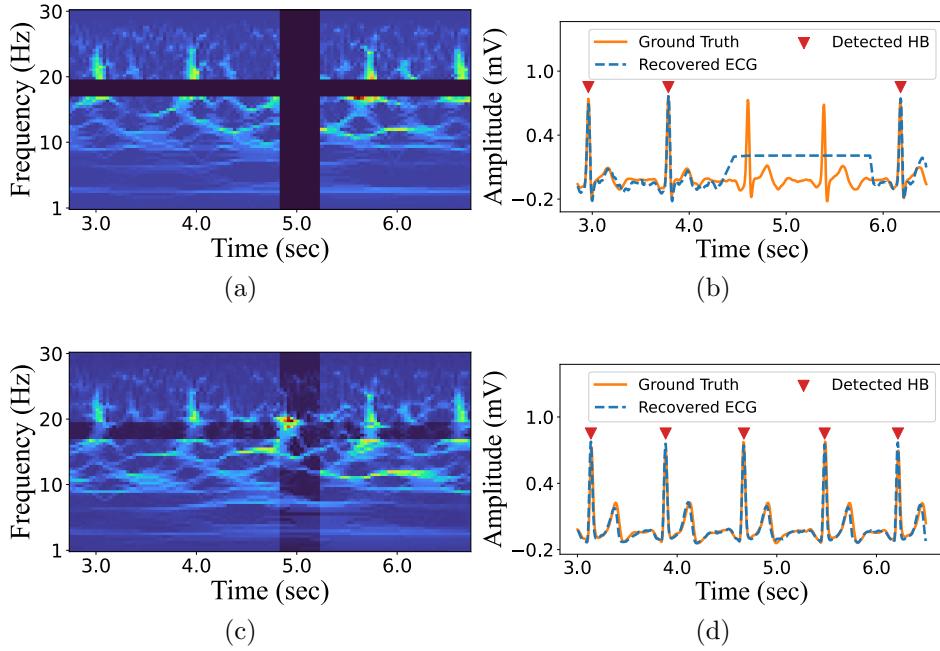


Figure 6.1: Illustration of Horcrux: (a) Traditional augmentation with zero mask; (b) Missed detected heartbeat (HB) and distorted ECG recovery; (c) Horcrux with time consistency preserved; (d) Ideal cardiac monitoring with high fidelity.

investigated for radar-based ECG recovery, and all the deep-learning-based ECG recovery models are trained in a supervised manner with the large dataset containing 3 – 32 hours of synchronous radar-ECG pairs [1, 30, 37].

Based on the discussion above, the contributions of this study can be listed as:

- A data augmentation method called Horcrux is designed for extracting cardiac features from radar by preserving the intrinsic time consistency hidden in the input radar spectrogram to restrict the potential distribution shift as illustrated in Figure 6.1, expanding the diversity of the limited training dataset without distorting the key features.
- A transfer learning framework RFcardi is proposed following an SSL paradigm to effectively learn the latent representations from radar signals by leveraging an appropriate pre-text task. Accordingly, this work further designates the sparse

signal reconstruction/recovery (SSR) as the pre-text task, assisting the RFcardi to learn essential representations for the later ECG recovery.

- The experimental results illustrate that the proposed Horcrux outperforms the existing augmentation methods in both classification and regression tasks (i.e., heartbeat detection and ECG recovery). In addition, the pre-trained RFcardi framework can be easily adapted to realize radar-based ECG recovery with a small amount of synchronous radar-ECG measurements for fine-tuning.

The rest of the chapter is organized as follows. Section 6.2 and 6.3 elaborates the proposed Horcrux and RFcardi framework, with the experimental results shown in Section 6.4 and 6.5. The final conclusion is shown in Section 6.6.

6.2 Data Augmentation for Radar Spectrogram

6.2.1 Overview

The overview of the proposed Horcrux is based on the signal model in (5.1) and is depicted in Figure 6.2 with two parallel modules:

- The harmonic and percussive (H&P) decomposition module decomposes the original radar signal into the harmonic and percussive components, with the former referring to the stable component along certain frequency bands on a spectrogram and the latter representing the transient component with a pulsatile nature occurring at certain time indices, as shown in the red boxes in Figure 6.2.
- The dynamic template matching (DTM) module aims to figure out the most characteristic vibration based on the signal template in (5.1) and output the corresponding time indices T_1, T_2 and central frequencies f_1, f_2 (omit k for simplicity), with the comparison between the raw and synthetic radar signal shown in Figure 6.2.

The traditional augmentation methods either cannot be directly applied to the ground truth data or could potentially erase the crucial part for prediction (especially for

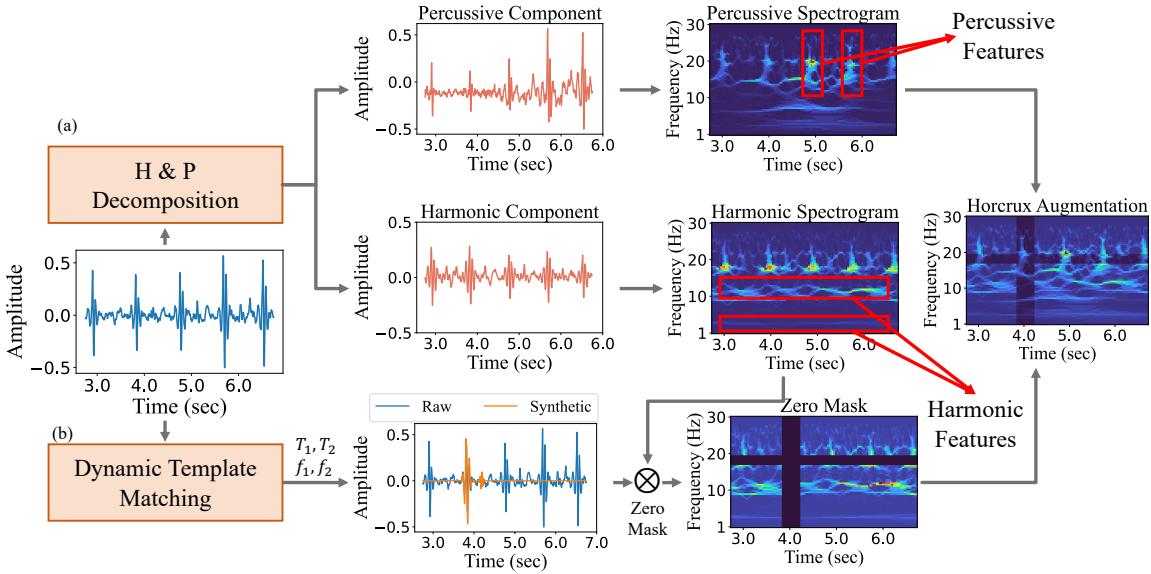


Figure 6.2: Pipeline of Horcrux with two branches: (a) Harmonic and percussive (H&P) components decomposition; (b) Dynamic template matching.

heartbeat detection), as shown in Figure 6.1(a). Therefore, Horcrux only masks the spectrogram of harmonic component and preserves the percussive feature, as shown in Figure 6.1(c). In addition, by deliberately masking the related part based on T_1, T_2, f_1, f_2 , it is expected that the well-trained deep learning model will concentrate more on the regions that help cardiac feature extraction instead of being distracted by the noises spread on the spectrogram.

6.2.2 Harmonic and Percussive Components Decomposition

Signal decomposition is widely used in signal processing to identify components with different features and help the downstream algorithm extract target information e.g., independent component analysis (ICA) decomposes the signal from different sources based on statistical independence, and empirical mode decomposition (EMD) decomposes the signal based on the oscillatory modes with different time scales [1]. Considering the nature of the cardiac activities, the radar signal reflected from the chest region reveals two characteristics on the spectrogram: (a) the stable components

along certain frequency bands that represent either the heart rate frequency (near 1Hz) or the center frequency of the prominent vibrations (10 – 25 Hz) [40]; (b) the transient components simply indicates the pulsatile heartbeats with large energy. In this research, these two components are named harmonic and percussive components (features) as highlighted in the red boxes in Figure 6.2.

The decomposition process is based on the spectrogram $\mathcal{Y}(m, n)$ obtained using any time-frequency representation tools such as synchrosqueezed wavelet transform [40], with m and n indicating the time frame index and frequency bin index, respectively. Then, the median filter $\mathcal{F}(\cdot)$ is applied along the time axis of $\mathcal{Y}(m, n)$ to get the harmonic-enhanced spectrogram as:

$$Y_h(m, n) = \mathcal{F}(\mathcal{Y}(m, n), k_h) \quad (6.1)$$

where the median filter replaces the middle value within a window length k_h by the median value in this window and is widely used in image and signal processing.

Similarly, the percussion-enhanced spectrogram can be obtained using a median filter with window length k_p along the frequency axis as:

$$Y_p(m, n) = \mathcal{F}(\mathcal{Y}(m, n), k_p) \quad (6.2)$$

Then, a soft mask based on the Wiener filter is suggested in [117] to adaptively suppress the unwanted components, and the Wiener masks M_h and M_p for isolating harmonic and percussive component from the original spectrogram are calculated as:

$$\begin{aligned} M_h(m, n) &= \frac{Y_h(m, n)}{Y_h(m, n) + Y_p(m, n)} \\ M_p(m, n) &= \frac{Y_p(m, n)}{Y_h(m, n) + Y_p(m, n)} \end{aligned} \quad (6.3)$$

Lastly, the harmonic spectrogram $\mathcal{Y}_h(m, n)$ and percussive spectrogram $\mathcal{Y}_p(m, n)$ can be obtained as:

$$\begin{aligned} \mathcal{Y}_h(m, n) &= M_h(m, n) \otimes \mathcal{Y}(m, n) \\ \mathcal{Y}_p(m, n) &= M_p(m, n) \otimes \mathcal{Y}(m, n) \end{aligned} \quad (6.4)$$

with \otimes representing element-wise multiplication, and the resultant spectrograms are shown in Figure 6.2 with the corresponding features enhanced. In addition, it can be observed that certain percussive features are still revealed on the harmonic component and vice versa, because the soft Wiener mask can not totally eliminate the other component with much larger energy on the spectrogram. In addition, the Horcrux augmentation does not require a complete separation of H&P component, H&P decomposition module is to preserve the percussive features, therefore it is acceptable

6.2.3 Dynamic Template Matching (DTM)

The DTM module is designed to figure out the regions in the time- and frequency-domain that are closely related to the prominent vibrations, encouraging the deep learning model to focus on these regions for capturing crucial information. According to the literature [40], the time indices T_1, T_2 of the prominent vibrations can be identified by the deep learning model based on their central frequencies f_1, f_2 , and Horcrux proposes to dynamically obtain these four parameters for each input segment as a template matching problem to find the optimal parameter set:

$$\theta^* = \arg \min_{\theta=\{T_1, T_2, f_1, f_2\}} \|y(t) - \tilde{d}(t, \theta)\|_2 \quad (6.5)$$

Subject to: $T_2 - T_1 < \tau$

where $y(t)$ is the segment from dataset and $\tilde{d}(t, \theta)$ represents the signal synthesized from (5.1) with the parameter set θ . The constraint represents that T_2 always follows T_1 within a distance τ , because the identified vibrations should belong to the same cardiac cycle and the time interval between aortic valve opening (T_1) and closure (T_2) should be less than τ sec [23].

To simplify the optimization, only one cardiac cycle is considered for one segment, and only the key parameters (i.e., $\theta = \{T_1, T_2, f_1, f_2\}$) are left to be determined with initial values shown in Table 6.1, while the values of other parameters are empirically assigned as shown in Table 6.1. In addition, the optimization problem in (6.5) is solved by sequential least squares programming (SLSQP) to yield optimal parameter set θ^*

Table 6.1: Parameters for Signal Model

Par.	Value	Par.	Value	Par.	Value	Par.	Value
T_1	0.4	T_2	0.85	f_1	10	f_2	23
a_1	0.5	a_2	0.1	b_1	0.05	b_2	0.03

that indicates the time indices and central frequencies of the prominent vibrations as shown as the synthetic signal in Figure 6.2.

6.2.4 Augmented Spectrogram Generation

The last step of Horcrux is to randomly select one vibration from v_1, v_2 and mask the corresponding regions on the harmonic spectrogram based on θ^* with the width of w_t or w_f for time or frequency domain. The applied zero mask alters the data distribution of the original spectrogram to enhance the diversity of the limited dataset, and the deep learning model will concentrate more on the masked regions selected by the DTM module with informative features for cardiac activities, as shown in Figure 6.2.

Lastly, the harmonic spectrogram with zero mask is added with the percussive spectrogram yielded from H&P decomposition module to provide necessary information that preserves time consistency in the final Horcrux augmentation spectrogram as shown in Figure 6.2. In addition, the added percussive spectrogram restricts the shift of data distribution brought by zero masks, and the augmented spectrogram still shares a similar representation with the normal spectrogram that is used in the evaluation or test stage of deep neural work training.

6.3 Transfer Learning Framework Design

6.3.1 Deep Learning Model Design

In the previously designed data deep learning model [40], the radar signals with high-SNR should be convert to spectrograms, providing extra frequency-domain informa-

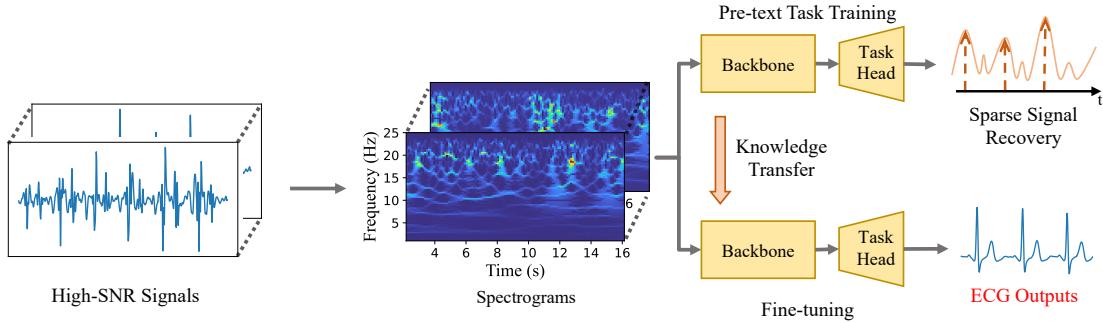


Figure 6.3: Illustration of RFcardi with pre-text training and fine-tuning stages.

tion to assist model training. The deep learning model adopts the popular backbone-decoder structure as designed in [40]:

- The backbone leverages ResNet [118] framework with deformable 2D convolution layer [89] to efficiently extract cardiac features from image-like spectrogram inputs.
- The decoder is based on 1D convolutional neural network (CNN) to generate corresponding signals either for the pre-text task or ECG signal recovery.

6.3.2 Pre-text Task Training and Fine-tuning

Inspired by other signal-based research [31,88], transfer learning is a promising paradigm to learn the latent representation from unlabeled radar signal to capture basic cardiac features in an SSL manner, reducing the requirement of cumbersome ECG collection. Then, only a small amount of synchronous radar-ECG pairs is required to fine-tune the pre-trained model to realize the ECG recovery for new scenarios. The proposed RFcardi framework adopts high-SNR radar signals as inputs to pre-train the backbone with SSR as the pre-text task. Then, the same backbone will be used for the fine-tuning stage so that the latent representations learned in the pre-trained model can be seemingly transferred for the ECG recovery task, as shown in Figure 6.3.

The efficient SSL requires an appropriate design of the pre-text task to help the deep learning model capture essential features that assist ECG recovery, and the pre-

text task used for SSL should reveal certain inherent features in the radar-monitored vital sign. Two major features used for traditional heart rate estimation are periodicity and sparsity [4]. In this work, the duration of each segment is 4 sec and may not reveal strong periodicity. Therefore, SSR will be used as the pre-text task in RFcardi and is defined as:

$$h = \Phi x + n \quad (6.6)$$

where h is the high-SNR radar signal, Φ means the observation matrix, x is the sparse representation for heartbeats and n represents the residual noise. The traditional SSR task can be seemly converted to a system identification problem by viewing Φ as a multi-channel adaptive filter, and the estimation of the filter coefficient is the same as training a CNN-based neural network (i.e., training CNN kernels) [17].

In this case, the SSR task is realized by using the aforementioned CNN-based backbone-decoder structure with the loss function:

$$\mathcal{L} = \|x - x'\|_2 + \underbrace{\lambda_s \frac{\|x\|_1 / \|x\|_2 - 1}{\sqrt{m} - 1}}_{\text{sparse penalty}} \quad (6.7)$$

where m is the length of the signal, x is the output the from deep learning model, x' is the sparse ground truth with values for the radar peaks maintained (1st vibration in Figure 3.4(a)) while other values are set to 0. The sparse penalty has a range of $[0, \lambda_s]$ with a smaller value indicating better sparsity [119].

After pre-training based on SSR, the parameters of backbone will be retained with a new decoder connected (same structure as for pre-text task training), and a few radar-ECG pairs are used for fine-tuning the pre-trained RFcardi model using MSE as the loss function.

6.4 Experimental Results for Horcrux

Evaluation Metrics

For the evaluation of the proposed data augmentation method Horcrux, the performance of the ECG recovery and heartbeat detection are evaluated from the following perspectives:

- **Root mean square error (RMSE):** RMSE describes the morphological fidelity of the recovered ECG signal and is sensitive to the peak deviation.
- **Pearson-correlation coefficient (PCC):** PCC also reveals morphological fidelity but is sensitive to the general shape of the recovered ECG signal.
- **Heartbeat error (H. E.):** Heartbeat error simply shows the absolute timing error of the detected heartbeat with the ground truth.
- **Missed detection rate (MDR):** MDR is a metric to assess the performance of deep learning model against noises by calculating the percentage of the missed detected heartbeats, because the strong noise (e.g., body movement) or low SNR scenarios may drown the subtle cardiac features and cause missed detection.

In addition, to comprehensively evaluate the overall performance of each method, $\Delta m\%$ is adopted as in many multi-task learning studies to assess the performance based on multiple metrics as:

$$\Delta m\% = \frac{1}{n} \sum_{i=1}^n S_{i,j} \frac{M_{m,i} - M_{b,i}}{M_{b,i}} \times 100\% \quad (6.8)$$

where n is the number of metrics, $M_{m,i}$ means the performance of a method m measured by metric j , $M_{b,i}$ means the performance of the baseline, and $S_{i,j} = 1$ or 0 if lower or higher values are better for the current metric (indicated by \downarrow or \uparrow).

6.4.1 Evaluation for DTM

The DTM module aims to figure out the most prominent vibrations for a certain cardiac cycle within a 4-sec input segment, and the illustration of the performance under different radar signal qualities (SNR) is shown in Figure 6.4. It is worth noticing that the proposed DTM module not only identifies the prominent vibrations under good signal quality (e.g., Figure 6.4(a) and 6.4(c)), but could tolerate the low-SNR scenarios with certain constant noise drowning the characteristic peaks (e.g., Figure 6.4(e) and 6.4(f)).

In addition, the DTM module could successfully identify 87% and 53% of the first and the second prominent vibrations (v_1 and v_2) within the absolute tolerance of 0.15s [1], ensuring a precise mask applied on the spectrogram in Horcrux. The specific impacts of the DTM module on the final ECG recovery quality will be evaluated in the later ablation study section, and the wrong detection in DTM will not degrade the overall performance of Horcrux because applying a random mask also contributes to the radar-based ECG recovery.

6.4.2 Evaluation for Horcrux

Comparison with State-of-the-art Methods

The performance of the proposed Horcrux is shown in Table 6.2 in terms of RMSE, PCC, H. E. and MDR, with the comparison among the state-of-the-art augmentation methods for regression tasks by mixing up the input and ground truth within homogeneous samples [120–122]. In addition, the Horcrux is implemented on the dataset with different proportions to evaluate the impact of masks on the change of data distributions and model performance as shown in Table 6.2.

From Table 6.2, Horcrux achieves the best result when applying on 20% of the dataset with an overall improvement of 16.20% and outperforms all the existing augmentation methods, because the mix-up-based methods enhance the diversity of the dataset in sacrificing the time consistency inside one cardiac cycle. Therefore, the mix-up-based methods show less improvement or even degradation in RMSE and PCC,

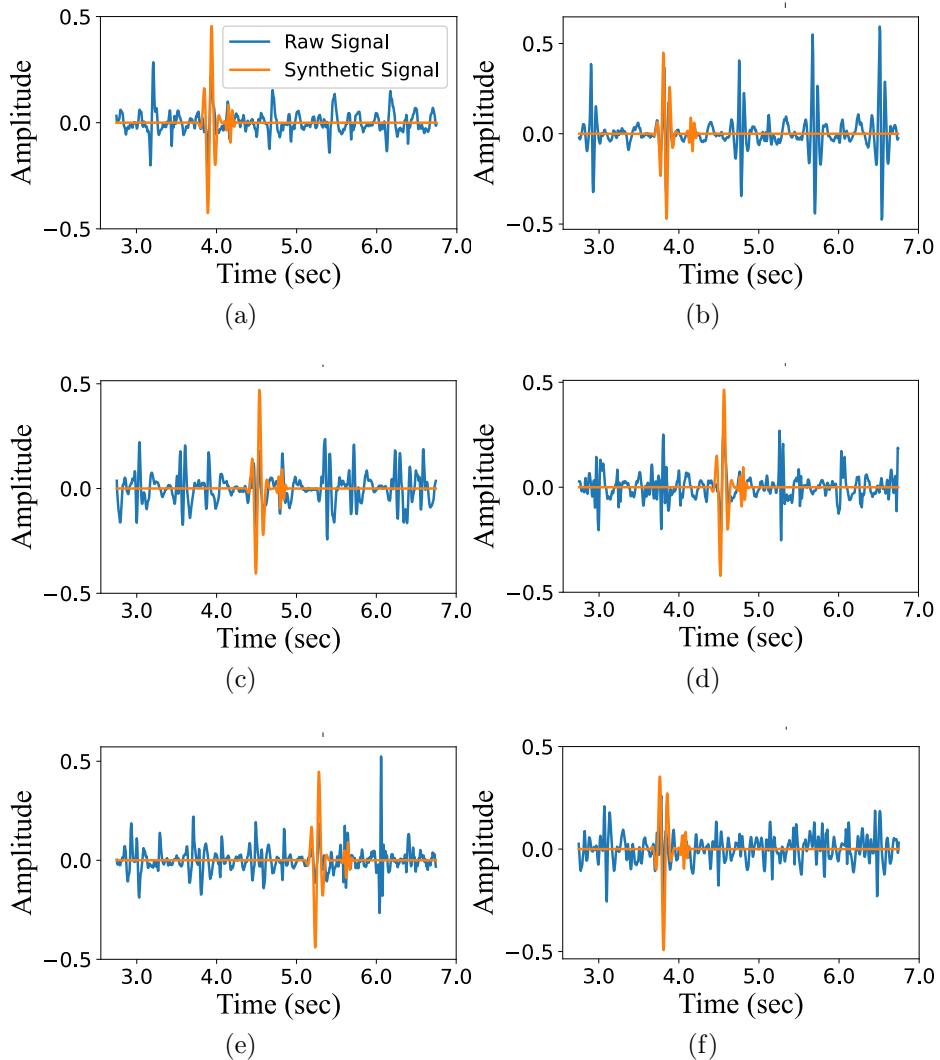


Figure 6.4: Identified vibrations using DTM for radar signals with various qualities (illustrated in normalized amplitude).

Table 6.2: Comparison of Different Augmentation Methods

Methods	RMSE (mV) ↓	PCC ↑	H. E. (ms) ↓	MDR ↓	$\Delta m\%$
Baseline	0.096	82.65%	8.82	6.73%	0.0%
C-Mixup [122]	<u>0.088</u>	82.25%	7.55	6.14%	7.80%
ADA [121]	0.097	81.28%	7.20	6.69%	4.12%
RC-Mixup [120]	0.089	83.36%	7.72	5.79%	8.69%
Horcrux (10%)	0.096	82.24%	6.89	6.67%	5.63%
Horcrux (15%)	0.087	<u>84.75%</u>	6.96	5.19%	<u>14.02%</u>
Horcrux (20%)	0.086	85.41%	<u>6.21</u>	<u>5.30%</u>	16.20%
Horcrux (25%)	0.092	81.99%	6.16	6.36%	9.81%
Horcrux (30%)	0.094	80.18%	6.55	6.59%	6.78%

The best/second best results are indicated by **Bold**/underline.

while Horcrux preserves the percussive component to maintain the time consistency inside the spectrogram. However, the masks applied on the spectrogram still change the data representation, and the overuse of Horcrux may cause the drift in the learned data representation or distribution, causing the degradation starting from the ECG recovery, as indicated by the bad RMSE and PCC in Table 6.2 for Horcrux (25% and 30%).

Ablation Study

The improved performance achieved by Horcrux comes from two aspects, i.e., H&P decomposition and DTM module, and an ablation study is performed to specify the contributions as shown in Table 6.3. Firstly, the original spectrogram is used as inputs with masks applied in the time or frequency domain randomly or based on the DTM module. It is worth noticing that simply applying random masks on the original set could achieve an improvement of 8.42%, and masking the frequency domain (7.04%) contributes more than masking the time domain (4.28%), because zero mask will cover all the useful information and affects heartbeat detection. This phenomenon can also be verified by precisely masking the prominent vibrations based on DTM, causing the

Table 6.3: Ablation Study for Horcrux

Input	Mask Type		Selected Domain		Evaluation Metrics				$\Delta m\%$
	Random	DTM	Time	Frequency	RMSE (mV) ↓	PCC ↑	H. E. (ms) ↓	MDR ↓	
Original Spectrogram (without Horcrux)	-	-	-	-	0.096	82.65%	8.82	6.73%	0.0%
	✓	-	✓	-	0.089	82.86%	7.14	7.38%	4.28%
	✓	-	-	✓	0.088	83.53%	7.80	6.27%	7.04%
	✓	-	✓	✓	0.087	83.03%	7.26	6.25%	8.42%
	-	✓	✓	-	0.089	83.71%	9.53	6.34%	-1.13%
	✓	✓	-	✓	0.099	80.00%	8.03	7.01%	-0.24%
	✓	✓	✓	✓	0.098	81.19%	7.88	7.95%	-2.41%
H&P Spectrogram (with Horcrux)	-	-	-	-	0.091	83.73%	7.45	5.92%	8.57%
	✓	-	✓	-	0.084	85.40%	6.90	6.71%	9.48%
	✓	-	-	✓	0.086	84.03%	8.26	5.35%	9.86%
	✓	-	✓	✓	0.085	84.89%	7.22	5.74%	11.80%
	-	✓	✓	-	0.083	84.94%	7.04	5.72%	13.18%
	✓	✓	-	✓	0.085	87.20%	7.96	5.66%	10.69%
	✓	✓	✓	✓	0.086	85.41%	6.21	5.30%	16.20%

The best/second best results are indicated by **Bold**/underline.

heavy degradation with negative $\Delta m\%$ as shown in Table 6.3.

In addition, the H&P decomposition itself also improves the cardiac features extraction with an improvement of 8.57%, because the percussive and harmonic features are both enhanced after adding them up. Afterward, the random mask could further improve the performance to 9.86%, while the DTM-based mask achieves the largest improvement of 16.20%. It is also worth noticing that only masking the time domain (13.18%) achieves a better result than covering the frequency domain (10.69%), because the H&P decomposition preserves the percussive information, and the heartbeat is still detectable even after masking the prominent vibrations precisely.

Performance under Different Dataset Scale

As a data augmentation technique, Horcrux is tested with different scales of dataset with the overall performance $\Delta m\%$ shown in Figure 6.5. It is notable that the degradation of the performance is not in a linear relationship with the input data scale, showing an obvious tipping point after axing the input data to 80% without augmentation. After applying Horcrux, the performance is boosted using the same scale of input data as already proved by the previous ablation study. In addition, the deep

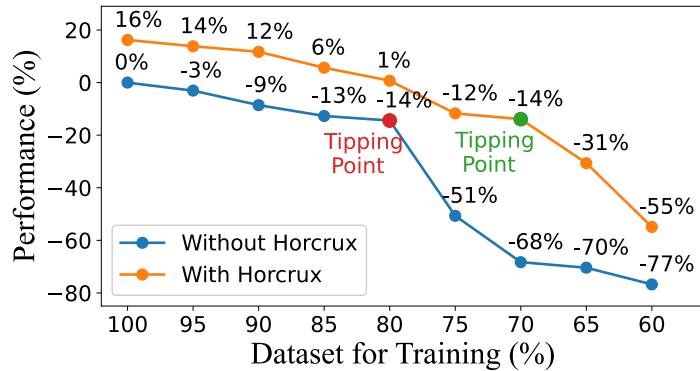


Figure 6.5: Performance of using different scales of the dataset.

learning model with Horcrux reveals a mild degradation with reduced input data, and the occurrence of the tipping point is also postponed to 70% as shown in Figure 6.5, because the faithful alteration of the data representation could improve the generalization of deep learning model and alleviate the risk of overfitting.

6.5 Experimental Results for RFcardi

For the evaluation of the proposed transfer learning framework RFcardi, the experimental results are illustrated for the SSL stage and fine-tuning stage, respectively.

6.5.1 Evaluations and Ablation Studies of SSL

The SSR task is crucial in the proposed transfer learning framework to provide latent representation that assists the further ECG pattern recovery, reducing the demand for radar-ECG pairs in the fine-tuning stage. The results of SSL are shown in Table 6.4 in terms of MSE and sparsity to illustrate the former and latter part (without λ_s) in the loss function (6.7) for SSL training.

The experiment is repeated for different dataset scales with the ablation study on the use of sparse penalty, and the results indicate that both MSE and sparsity decrease with the reducing training data as shown in Table 6.4. Training with 100% or 80% dataset could achieve convergence and realize a successful SSR with similar

Table 6.4: Performance of SSL with ablation study

Methods	MSE ($\times 10^2$) ↓	Sparsity ↓	MSE ($\times 10^2$) ↓	Sparsity ↓
	100% Dataset		80% Dataset	
SSL w/o sp*	0.91	0.36	0.96	0.41
SSL with sp	0.82	0.20	0.85	0.22
	60% Dataset		40% Dataset	
	SSL w/o sp	1.43	0.44	Failed
SSL with sp	0.92	0.26	0.98	0.31

*sp for sparse penalty

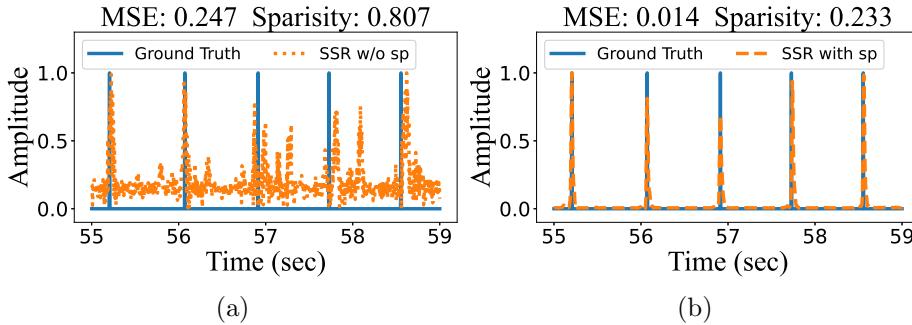


Figure 6.6: Results of SSR: (a) Failed SSR due to lack of data and sparse penalty; (b) Ideal SSR result with good MSE and sparsity.

MSE (0.0091, 0.0082) or (0.0096, 0.0085). However, the performance of SSR degrades heavily when further decreasing the training data without the constraint of sparse penalty, because the SSR results might fluctuate, as shown in Figure 6.6(a). In contrast, introducing the sparse penalty could suppress the fluctuation and force the deep learning model to focus only on the dominant peaks of the input radar signals, as shown in Figure 6.6(b). Therefore, the training with sparse penalty loss still achieves good results with an MSE of 0.0092 and 0.0098 by using 60% and 40% of the dataset, while the model cannot be well-trained without sparse penalty by using 40% dataset as shown in Table 6.4 and Figure 6.6(a).

Table 6.5: Performance of ECG Recovery using different percentages of labeled data

Methods	MSE ($\times 10^{-2}$) ↓	PCC ↑	Peak Error (ms) ↓	MDR ↓	Overall ↑	MSE ($\times 10^{-2}$) ↓	PCC ↑	Peak Error (ms) ↓	MDR ↓	Overall ↑
	100% Labeled					80% Labeled				
Supervised	0.80	85.47%	7.61	6.85%	0.00%	0.84	84.60%	8.90	8.04%	-10.09%
TF w/o sp*	0.81	85.35%	8.46	5.51%	1.75%	0.82	86.36%	8.35	7.02%	-3.42%
TF with sp	0.80	85.51%	8.40	5.14%	3.66%	0.81	84.29%	8.31	7.14%	-4.02%
60% Labeled					40% Labeled					
Supervised	0.93	79.91%	10.65	8.93%	-23.27%	0.98	75.89%	11.15	12.15%	-39.4%
TF w/o sp	0.85	83.74%	8.84	8.65%	-12.68%	0.97	76.56%	10.87	10.99%	-34.05%
TF with sp	0.86	84.92%	8.58	7.72%	-8.71%	0.93	78.72%	8.70	9.02%	-17.54%

*TF for transfer learning and sp for sparse penalty

6.5.2 Evaluations of Fine-tuning Results

The fine-tuning is based on the model pre-trained by 100% dataset with or without sparse penalty, and the experiment is repeated for different percentages of labeled data (i.e., radar signal with ECG ground truth). In addition, the same deep-learning model will be trained in a supervised manner by using the same amount of labeled data as the reference for transfer-learning ECG recovery. At last, an overall improvement will be provided by calculating the percentage of improvement across all four metrics to provide straightforward evaluations for different methods, as shown in Table 6.5.

The fine-tuning with 100% labeled data provides very similar performance in morphological accuracy with the MSE and PCC around 0.0080 and 85.47%. It is worth noticing that the peak error and MDR are slightly improved, because the pre-text task SSR for pre-training is equivalent to identifying the peak position of the radar signal, and the learned representations can be seemly transferred to improve the accuracy of the recovered ECG R peaks, contributing to the overall improvement for transfer learning (3.66% and 1.75%).

Reducing 20% of labeled data causes a 10% overall degradation as shown in Table 6.5, and the decline of peak error and MDR is more than MSE and PCC. The reason is that ECG morphological patterns for different cardiac cycles are similar and can be well-learned from 80% labeled data with good MSE and PCC (0.0084 and 84.60%), while the location of each ECG piece is random and can be distorted by noises, requiring more training data for convergence.

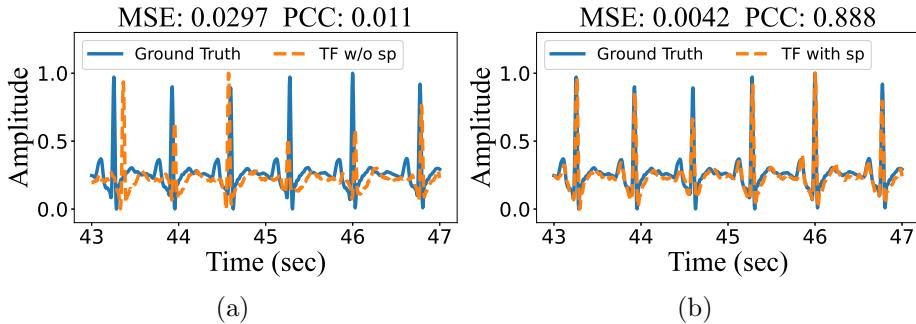


Figure 6.7: Results of transfer learning using limited labeled data: (a) Poor ECG recovery without proper morphological feature and peak location; (b) Good ECG recovery owing to the pre-trained model.

The supervised training with 60% labeled data cannot ensure a good morphological and peak accuracy and the overall degradation is 23.37%, with the PCC drop below 80% as shown in Table 6.5. In contrast, the pre-trained model still provided good results with mild degradations of 8.71% and 12.68%. It is noticed that the effectiveness of sparse penalty in the SSL stage also affects the fine-tuning stage, because both peak error and MDR for transfer learning with sparse penalty are better than those without sparse penalty, causing a large gap in the overall improvement compared with the previous training with 100% and 80% dataset.

Lastly, the deep learning model can barely learn from 40% labeled dataset and yield a bad morphological and peak accuracy for supervised learning. In addition, examples of the recovered ECG for transfer learning with or without sparse penalty are shown in Figure 6.7, and it is clear that the deep learning model struggles to learn both morphological and peak features from limited data if the pre-text task is not well-trained without sparse penalty, while Figure 6.7(b) shows the good recovery because the pre-trained model transfer the learned representations from radar inputs to ECG recovery.

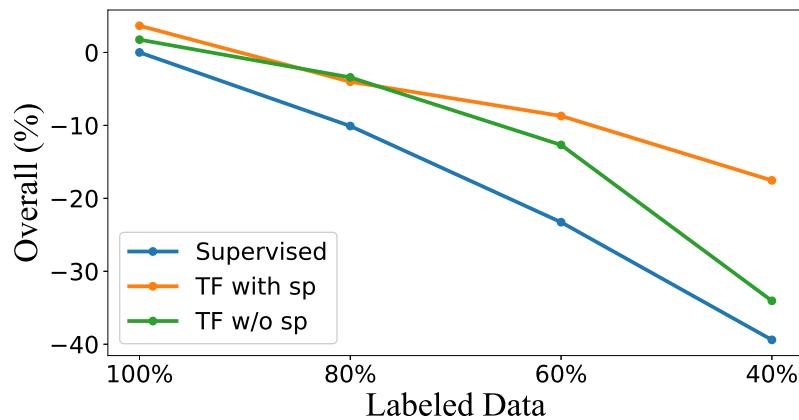


Figure 6.8: Overall performance of the radar-based ECG recovery.

6.5.3 Summary of Transfer-learning-based ECG Recovery

Previous evaluations in terms of SSL and fine-tuning stages have illustrated the ability of the proposed RFcardi to learn from unlabeled data and transfer the knowledge to the ECG recovery task using limited radar-ECG pairs. The overall performance in Table 6.5 are plotted in Figure 6.8 for a straightforward comparison:

- The performance of supervised learning drops heavily and cannot ensure high-quality ECG recovery after reducing 40% labeled dataset.
- Transfer learning could enhance the performance of ECG recovery for the cases with ample labeled training data, and the quality of the pre-trained model has a minor effect on the final result because the deep learning model could learn from numerous radar-ECG pairs.
- For the cases with limited labeled data (40%, 60%), the proposed RFcardi shows outstanding performance owing to the representations learned from unlabeled data. Furthermore, the quality of the pre-trained model does matter to alleviate the burden of deep learning model to learn both morphological ECG patterns and peak locations, as indicated by the increasing gap between orange and green lines in Figure 6.8.

6.6 Conclusion

This study proposes the data augmentation technique, Horcrux, for radar-based cardiac feature extraction with deep learning frameworks. Different from the traditional augmentation methods that may block the crucial parts or destroy the time consistency between input and ground truth, Horcrux preserves the percussive features of the input signal and only masks the cardiac-feature-related region on the harmonic spectrogram, enhancing the diversity of the limited dataset and also encouraging the deep learning model to only focus on the domain-preserved features. Compared with the existing methods, Horcrux reveals outstanding performance in heartbeat detection and ECG recovery and could effectively alleviate the degradation caused by the limited data. In addition, a transfer learning framework RFcardi is designed with SSR as a pre-text task for pre-training to reduce the dependency on cumbersome ECG ground truth collection. The experiments performed in different scenarios prove the feasibility of the CFT-RFcardi framework in radar signal extraction and ECG recovery with limited labeled data, enabling a convenient deployment in new scenarios with limited data for future contactless wellness monitoring.

Chapter 7

Conclusions and Future Work

7.1 Conclusions

This thesis investigates an emerging research area to reconstruct ECG signals from the measured radar signals, and multiple contributions are made to improve different stages in radar-based ECG recovery:

- Chapter 3 explores methods for efficiently collecting high-SNR radar signals that contain rich cardiac features, aiming to support deep-learning-based ECG recovery in later chapters. A novel **cardio-focusing and -tracking (CFT)** algorithm is designed to dynamically identify points with optimal SNR and track the cardiac location as subjects change postures.
- Chapter 4 aims to bridge the gap to realize a robust transformation from the mechanical domain to the electrical domain by proposing the signal model with fine-grained features considered. Furthermore, a deep learning framework **radarODE** is designed with morphological prior embedding as ODEs to provide faithful single-cycle ECG recoveries even under strong noises.
- Chapter 5 is based on the robust single-cycle ECG generator designed in the previous chapter and further investigates the long-term ECG recovery under abrupt or constant noises. The realization of ECG recovery is appropriately

deconstructed into three sub-tasks with an MTL framework called **radarODE-MTL** designed to generate long-term ECG signal. Additionally, a novel optimization strategy named EGA is presented to optimize all tasks simultaneously, effectively avoiding issues such as stalling or negative transfer.

- Chapter 6 tries to alleviate the data scarcity for DNN training, because radar-based ECG reconstruction is highly reliant on data-driven approaches. Firstly, a data augmentation method called **Horcrux** is proposed to expand the diversity of the limited training dataset without distorting the key features. Secondly, a transfer learning framework called **RFcardi** is designed by leveraging an appropriate pre-text task (i.e., sparse signal reconstruction/recovery (SSR)), enabling an effective learning of the latent representations from radar signals to assist the final ECG recovery task.

7.2 Future Work

The proposed **CFT**, **radarODE**, **radarODE-MTL**, **Horcrux** and **RFcardi** have shown outstanding performance compared with the previous work to generate faithful ECG signals under noisy scenarios using limited data, while the potential limitation will be discussed in this part to encourage future improvements in radar-based ECG recovery for real-life situations and applications.

- **Long-range ECG Monitoring:** The direct impact brought by long-range monitoring is to reduce the SNR of the received radar signal according to the link budget analysis [2]. In addition, the cardiac location requires to be pre-identified to perform the accurate beamforming, because the current dataset assumes the majority of range-bins contain cardiac-related signals [1].
- **Theoretical Model for Excitation-contraction Coupling:** The signal model (4.3) and (5.1) proposed in this thesis partially explain the recovering process from radar signal to ECG signal, while a key phase (i.e., excitation-contraction coupling [23,34]) still needs a compact model because the current decoupling still

purely relies on deep learning model, restring the reliable application especially for the implementation in clinical diagnosis.

- **Complex Monitoring Scenarios:** Various new noises might be introduced and need to be eliminated, such as radar self-vibration introduced from car vibrations or hand-held radar [66], mutual-radar interference for the future smart home with multiple electromagnetic devices [123] and signal attenuation caused by human tissues for the monitoring of people with random body orientations [24].
- **Evaluation on the Dataset for Patients:** An important future application of radar-based ECG recovery is for clinical monitoring and diagnosis, while the ECG waveform for patients (e.g., arrhythmia) might be quite different, requiring massive new data for training. Some recent research has shown the feasibility of recovering abnormal ECG from radar signal [37], but more studies are required to investigate transfer learning or data augmentation techniques due to the scarcity of patient data. In addition, it is hard to preserve the noise-robustness for the ECG monitoring of patients, because the ODE model in this work is not designed for abnormal ECG patterns.

In summary, the robust radar-based ECG recovery still needs improvements from both theoretical and practical perspectives. The aforementioned research directions are still waiting for a thorough investigation to enable a realistic radar-based ECG measurement in our daily lives.

Bibliography

- [1] J. Chen, D. Zhang, Z. Wu, F. Zhou, Q. Sun, and Y. Chen, “Contactless electrocardiogram monitoring with millimeter wave radar,” *IEEE Transactions on Mobile Computing*, Dec. 2022.
- [2] Y. Zhang, R. Guan, L. Li, R. Yang, Y. Yue, and E. G. Lim, “radarODE: An ODE-embedded deep learning model for contactless ECG reconstruction from millimeter-wave radar,” *IEEE Transactions on Mobile Computing*, Apr. 2025.
- [3] World Health Organization, “Ageing and health,” 2024, <https://www.who.int/news-room/fact-sheets/detail/ageing-and-health>, Accessed: May. 7, 2025.
- [4] Y. Zhang, R. Yang, Y. Yue, E. G. Lim, and Z. Wang, “An overview of algorithms for contactless cardiac feature extraction from radar signals: Advances and challenges,” *IEEE Transactions on Instrumentation and Measurement*, Aug. 2023.
- [5] R. Guan, S. Yao, X. Zhu, K. L. Man, E. G. Lim, J. Smith, Y. Yue, and Y. Yue, “Achelous: A fast unified water-surface panoptic perception framework based on fusion of monocular camera and 4D mmWave radar,” in *2023 IEEE 26th International Conference on Intelligent Transportation Systems*. IEEE, Sep. 2023, pp. 182–188.
- [6] F. Wang, F. Zhang, C. Wu, B. Wang, and K. R. Liu, “Vimo: Multiperson vital sign monitoring using commodity millimeter-wave radio,” *IEEE Internet of Things Journal*, vol. 8, no. 3, pp. 1294–1307, 2020.

- [7] X. Chen, J. Cheng, R. Song, Y. Liu, R. Ward, and Z. J. Wang, “Video-based heart rate measurement: Recent advances and future prospects,” *IEEE Transactions on Instrumentation and Measurement*, vol. 68, no. 10, pp. 3600–3615, Nov. 2018.
- [8] D.-M. Chian, C.-K. Wen, C.-J. Wang, M.-H. Hsu, and F.-K. Wang, “Vital signs identification system with Doppler radars and thermal camera,” *IEEE Transactions on Biomedical Circuits and Systems*, vol. 16, no. 1, pp. 153–167, Feb. 2022.
- [9] X. Xu, J. Yu, and Y. Chen, “Leveraging acoustic signals for fine-grained breathing monitoring in driving environments,” *IEEE Transactions on Mobile Computing*, vol. 21, no. 3, pp. 1018–1033, Mar. 2020.
- [10] W. Jia, H. Peng, N. Ruan, Z. Tang, and W. Zhao, “WiFind: Driver fatigue detection with fine-grained Wi-Fi signal features,” *IEEE Transactions on Big Data*, vol. 6, no. 2, pp. 269–282, Jun. 2020.
- [11] C. Feng, X. Jiang, M.-G. Jeong, H. Hong, C.-H. Fu, X. Yang, E. Wang, X. Zhu, and X. Liu, “Multitarget vital signs measurement with chest motion imaging based on MIMO radar,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 69, no. 11, pp. 4735–4747, Nov. 2021.
- [12] I. Nirmal, A. Khamis, M. Hassan, W. Hu, and X. Zhu, “Deep learning for radio-based human sensing: Recent advances and future directions,” *IEEE Communications Surveys & Tutorials*, vol. 23, no. 2, pp. 995–1019, 2nd Quart. 2021.
- [13] W. Xia, Y. Li, and S. Dong, “Radar-based high-accuracy cardiac activity sensing,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–13, Jan. 2021.
- [14] Y. Wang, W. Wang, M. Zhou, A. Ren, and Z. Tian, “Remote monitoring of human vital signs based on 77-GHz mm-wave FMCW radar,” *Sensors*, vol. 20, no. 10, p. 2999, May 2020.

- [15] J. C. Lin, "Noninvasive microwave measurement of respiration," *Proceedings of the IEEE*, vol. 63, no. 10, pp. 1530–1530, Oct. 1975.
- [16] S. M. Islam, O. Boric-Lubecke, V. M. Lubecke, A.-K. Moadi, and A. E. Fathy, "Contactless radar-based sensors: Recent advances in vital-signs monitoring of multiple subjects," *IEEE Microwave Magazine*, vol. 23, no. 7, pp. 47–60, Jul. 2022.
- [17] U. Ha, S. Assana, and F. Adib, "Contactless seismocardiography via deep learning radars," in *Proceedings of the 26th Annual International Conference on Mobile Computing and Networking (MobiCom)*, Apr. 2020, pp. 1–14.
- [18] A. B. Obadi, P. J. Soh, O. Aldayel, M. H. Al-Door, M. Mercuri, and D. Schreurs, "A survey on vital signs detection using radar techniques and processing with FPGA implementation," *IEEE Circuits and Systems Magazine*, vol. 21, no. 1, pp. 41–74, Feb. 2021.
- [19] Q. Lv, L. Chen, K. An, J. Wang, H. Li, D. Ye, J. Huangfu, C. Li, and L. Ran, "Doppler vital signs detection in the presence of large-scale random body movements," *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 9, pp. 4261–4270, Sep. 2018.
- [20] M. Nosrati and N. Tavassolian, "Accurate Doppler radar-based cardiopulmonary sensing using chest-wall acceleration," *IEEE Journal of Electromagnetics, RF and Microwaves in Medicine and Biology*, vol. 3, no. 1, pp. 41–47, Mar. 2018.
- [21] M. Mercuri, I. R. Lorato, Y.-H. Liu, F. Wieringa, C. V. Hoof, and T. Torfs, "Vital-sign monitoring and spatial tracking of multiple people using a contactless radar-based sensor," *Nature Electronics*, vol. 2, no. 6, pp. 252–262, Jun. 2019.
- [22] F. Cocconcelli, N. Mora, G. Matrella, and P. Ciampolini, "High-accuracy, unsupervised annotation of seismocardiogram traces for heart rate monitoring,"

- IEEE Transactions on Instrumentation and Measurement*, vol. 69, no. 9, pp. 6372–6380, Jan. 2020.
- [23] L. M. Swift, M. W. Kay, C. M. Ripplinger, and N. G. Posnack, “Stop the beat to see the rhythm: excitation-contraction uncoupling in cardiac research,” *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 321, no. 6, pp. H1005–H1013, Dec. 2021.
- [24] J. Liu, J. Wang, Q. Gao, X. Li, M. Pan, and Y. Fang, “Diversity-enhanced robust device-free vital signs monitoring using mmWave signals,” *IEEE Transactions on Mobile Computing*, Jun. 2024.
- [25] Y. Li, C. Gu, and J. Mao, “A robust and accurate FMCW MIMO radar vital sign monitoring framework with 4-D cardiac beamformer and heart-rate trace carving technique,” *IEEE transactions on microwave theory and techniques*, Oct. 2024.
- [26] J. Xiong, H. Hong, L. Xiao, E. Wang, and X. Zhu, “Vital signs detection with difference beamforming and orthogonal projection filter based on SIMO-FMCW radar,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 71, no. 1, pp. 83–92, 2022.
- [27] S. Dong, Y. Li, C. Gu, and J. Mao, “Robust cardiac timing detection technique with vectors analytic demodulation in Doppler cardiogram sensing,” *IEEE Transactions on Microwave Theory and Techniques*, Jan. 2024.
- [28] C. Ni, J. Pan, D. Du, X. Yang, C. Shi, S. Chen, D. Yang, and S. Liu, “Accurate heart rate measurement across various body postures using FMCW radar,” *IEEE Transactions on Instrumentation and Measurement*, vol. 73, pp. 1–13, Mar. 2024.
- [29] D. Zhang, X. Zhang, Y. Xie, F. Zhang, H. Yang, and D. Zhang, “From single-point to multi-point reflection modeling: Robust vital signs monitoring via mmWave sensing,” *IEEE Transactions on Mobile Computing*, Dec. 2024.

- [30] B. Li, W. Li, Y. He, W. Zhang, and H. Fu, “Radarnet: Non-contact ecg signal measurement based on fmcw radar,” *IEEE Transactions on Instrumentation and Measurement*, Oct. 2024.
- [31] H. Zhang, D. Zhang, R. Song, Z. Wu, J. Chen, L. Fang, Z. Lu, Y. Hu, H. Lin, and Y. Chen, “Umimo: Universal unsupervised learning for mmWave radar sensing with MIMO array synthesis,” *IEEE Transactions on Mobile Computing*, Feb, 2025.
- [32] J. Chen, D. Zhang, G. Zhang, H. Wang, Q. Sun, and Y. Chen, “Co-sense: Exploiting cooperative dark pixels in radio sensing for non-stationary target,” *IEEE Transactions on Mobile Computing*, Nov. 2024.
- [33] Z. Chen, T. Zheng, C. Cai, and J. Luo, “MoVi-Fi: Motion-robust vital signs waveform recovery via deep interpreted RF sensing,” in *Proceedings of the 27th Annual International Conference on Mobile Computing and Networking (MobiCom)*, Feb. 2021, pp. 392–405.
- [34] R. Orkand and R. Niedergerke, “Heart action potential: dependence on external calcium and sodium ions,” *Science*, vol. 146, no. 3648, pp. 1176–1177, Nov. 1964.
- [35] Y. Wu, H. Ni, C. Mao, and J. Han, “Contactless reconstruction of ECG and respiration signals with mmWave Radar based on RSSRnet,” *IEEE Sensors Journal*, Nov. 2023.
- [36] Z. Wang, B. Jin, S. Li, F. Zhang, and W. Zhang, “ECG-grained cardiac monitoring using UWB signals,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 6, no. 4, pp. 1–25, Dec. 2023.
- [37] L. Zhao, R. Lyu, H. Lei, Q. Lin, A. Zhou, H. Ma, J. Wang, X. Meng, C. Shao, Y. Tang *et al.*, “AirECG: Contactless electrocardiogram for cardiac disease monitoring via mmWave sensing and cross-domain diffusion model,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 3, pp. 1–27, Sep. 2024.

- [38] Y. Wang, Z. Wang, J. A. Zhang, H. Zhang, and M. Xu, “Vital sign monitoring in dynamic environment via mmWave radar and camera fusion,” *IEEE Transactions on Mobile Computing*, Jun. 2023.
- [39] P. Wang, X. Ma, R. Zheng, L. Chen, X. Zhang, D. Zeghlache, and D. Zhang, “SlpRoF: Improving the temporal coverage and robustness of RF-based vital sign monitoring during sleep,” *IEEE Transactions on Mobile Computing*, Dec. 2023.
- [40] Y. Zhang, R. Yang, Y. Yue, and E. G. Lim, “radarODE-MTL: A multi-task learning framework with eccentric gradient alignment for robust radar-based ECG reconstruction,” *IEEE Transactions on Instrumentation and Measurement*, Apr. 2025.
- [41] Z. Chu, R. Yan, and S. Wang, “Vessel turnaround time prediction: A machine learning approach,” *Ocean & Coastal Management*, vol. 249, p. 107021, Mar. 2024.
- [42] Y. Zhang, S. Xiong, R. Yang, E. G. Lim, and Y. Yue, “Recover from horcrux: A spectrogram augmentation method for cardiac feature monitoring from radar signal components,” in *2025 47th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Jul. 2025.
- [43] C. Li, V. M. Lubecke, O. Boric-Lubecke, and J. Lin, “A review on recent advances in Doppler radar sensors for noncontact healthcare monitoring,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 61, no. 5, pp. 2046–2060, May 2013.
- [44] A. Singh, S. U. Rehman, S. Yongchareon, and P. H. J. Chong, “Multi-resident non-contact vital sign monitoring using radar: A review,” *IEEE Sensors Journal*, vol. 21, no. 4, pp. 4061–4084, Feb. 2021.
- [45] W. Hirt, “Ultra-wideband radio technology: Overview and future research,” *Computer Communications*, vol. 26, no. 1, pp. 46–52, Jan. 2003.

- [46] H. Shen, C. Xu, Y. Yang, L. Sun, Z. Cai, L. Bai, E. Clancy, and X. Huang, “Respiration and heartbeat rates measurement based on autocorrelation using IR-UWB radar,” *IEEE Transactions on Circuits and Systems II: Express Briefs*, vol. 65, no. 10, pp. 1470–1474, Oct. 2018.
- [47] D. Wang, S. Yoo, and S. H. Cho, “Experimental comparison of IR-UWB radar and FMCW radar for vital signs,” *Sensors*, vol. 20, no. 22, p. 6695, Nov. 2020.
- [48] Y. Huang, *Antennas: From theory to practice*. New York, NY, USA: John Wiley & Sons, 2021.
- [49] K. Ramasubramanian and K. Ramaiah, “Moving from legacy 24 GHz to state-of-the-art 77-GHz radar,” *ATZelektronik Worldwide*, vol. 13, no. 3, pp. 46–49, Jun. 2018.
- [50] D. Obeid, G. Issa, S. Sadek, G. Zaharia, and G. El Zein, “Low power microwave systems for heartbeat rate detection at 2.4, 5.8, 10 and 16 GHz,” in *2008 First International Symposium on Applied Sciences on Biomedical and Communication Technologies*, Oct. 2008, pp. 1–5.
- [51] Y. Zhou, L. Liu, H. Zhao, M. López-Benítez, L. Yu, and Y. Yue, “Towards deep radar perception for autonomous driving: Datasets, methods, and challenges,” *Sensors*, vol. 22, no. 11, p. 4208, May 2022.
- [52] A. D. Droitcour, O. Boric-Lubecke, V. M. Lubecke, J. Lin, and G. T. Kovacs, “Range correlation and I/Q performance benefits in single-chip silicon Doppler radars for noncontact cardiopulmonary monitoring,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 52, no. 3, pp. 838–848, Mar. 2004.
- [53] Y.-H. Lin, J.-H. Cheng, L.-C. Chang, W.-J. Lin, J.-H. Tsai, and T.-W. Huang, “A broadband MFCW agile radar concept for vital-sign detection under various thoracic movements,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, no. 8, pp. 4056–4070, 2022.

- [54] J. Zhang, Y. Wu, Y. Chen, and T. Chen, “Health-radio: Towards contactless myocardial infarction detection using radio signals,” *IEEE Transactions on Mobile Computing*, vol. 21, no. 2, pp. 585–597, Feb. 2022.
- [55] M. Mercuri, Y. Lu, S. Polito, F. Wieringa, Y.-H. Liu, A.-J. van der Veen, C. Van Hoof, and T. Torfs, “Enabling robust radar-based localization and vital signs monitoring in multipath propagation environments,” *IEEE Transactions on Biomedical Engineering*, vol. 68, no. 11, pp. 3228–3240, Nov. 2021.
- [56] M. Tang, P. Teckchandani, J. He, H. Guo, and E. Soltanaghai, “BSENSE: In-vehicle child detection and vital sign monitoring with a single mmWave radar and synthetic reflectors,” in *Proceedings of the 22nd ACM Conference on Embedded Networked Sensor Systems*, Nov. 2024, pp. 478–492.
- [57] Texas Instruments, “AWR1843 - Single-chip 76-GHz to 81-GHz automotive radar sensor integrating DSP, MCU and radar accelerator,” 2018, <https://www.ti.com/product/AWR1843>, Accessed: Mar. 8, 2025.
- [58] E. Mogi and T. Ohtsuki, “Heartbeat detection with Doppler radar based on spectrogram,” in *2017 IEEE International Conference on Communications (ICC)*, May 2017, pp. 1–6.
- [59] W. Hu, H. Zhang, Z. Zhao, Y. Wang, and X. Wang, “Real-time remote vital sign detection using a portable Doppler sensor system,” in *2014 IEEE Sensors Applications Symposium (SAS)*, Feb. 2014, pp. 89–93.
- [60] J. Park, J.-W. Ham, S. Park, D.-H. Kim, S.-J. Park, H. Kang, and S.-O. Park, “Polyphase-basis discrete cosine transform for real-time measurement of heart rate with CW Doppler radar,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 66, no. 3, pp. 1644–1659, Mar. 2018.
- [61] K. Shi, T. Steigleder, S. Schellenberger, F. Michler, A. Malessa, F. Lurz, N. Rohleder, C. Ostgatthe, R. Weigel, and A. Koelpin, “Contactless analysis of heart rate variability during cold pressor test using radar interferometry and

- bidirectional LSTM networks,” *Scientific Reports*, vol. 11, no. 1, pp. 1–13, Feb. 2021.
- [62] C. Ye, K. Toyoda, and T. Ohtsuki, “A stochastic gradient approach for robust heartbeat detection with Doppler radar using time-window-variation technique,” *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 6, pp. 1730–1741, Jun. 2018.
- [63] K. Yamamoto, K. Toyoda, and T. Ohtsuki, “Spectrogram-based non-contact RRI estimation by accurate peak detection algorithm,” *IEEE Access*, vol. 6, pp. 60 369–60 379, Oct. 2018.
- [64] Y. Rong and D. W. Bliss, “Remote sensing for vital information based on spectral-domain harmonic signatures,” *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, no. 6, pp. 3454–3465, May 2019.
- [65] M. Le and B. Van Nguyen, “Multivariate correlation of higher harmonics for heart rate remote measurement using UWB impulse radar,” *IEEE Sensors Journal*, vol. 20, no. 4, pp. 1859–1866, Oct. 2019.
- [66] S. D. Da Cruz, H.-P. Beise, U. Schröder, and U. Karahasanovic, “A theoretical investigation of the detection of vital signs in presence of car vibrations and radar-based passenger classification,” *IEEE Transactions on Vehicular Technology*, vol. 68, no. 4, pp. 3374–3385, Apr. 2019.
- [67] C. Will, K. Shi, R. Weigel, and A. Koelpin, “Advanced template matching algorithm for instantaneous heartbeat detection using continuous wave radar systems,” in *2017 First IEEE MTT-S International Microwave Bio Conference (IMBIOC)*, May 2017, pp. 1–4.
- [68] T. Sakamoto, R. Imasaka, H. Taki, T. Sato, M. Yoshioka, K. Inoue, T. Fukuda, and H. Sakai, “Feature-based correlation and topological similarity for interbeat interval estimation using ultrawideband radar,” *IEEE Transactions on Biomedical Engineering*, vol. 63, no. 4, pp. 747–757, Apr. 2015.

- [69] J. Gong, X. Zhang, K. Lin, J. Ren, Y. Zhang, and W. Qiu, “RF Vital Sign Sensing Under Free Body Movement,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, vol. 5, no. 3, pp. 1–22, Sep. 2021.
- [70] J.-F. Cardoso, “Blind signal separation: statistical principles,” *Proceedings of the IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.
- [71] F. Weishaupt, I. Walterscheid, O. Biallawons, and J. Klare, “Vital sign localization and measurement using an LFMCW MIMO radar,” in *2018 19th International Radar Symposium (IRS)*, Jun. 2018, pp. 1–8.
- [72] K. Shyu, L. Chiu, P. Lee, T. Tung, and S. Yang, “Detection of breathing and heart rates in UWB radar sensor data using FVPIEF-based two-layer EEMD,” *IEEE Sensors Journal*, vol. 19, no. 2, pp. 774–784, Jan. 2019.
- [73] S. Ji, Z. Zhang, Z. Xia, H. Wen, J. Zhu, and K. Zhao, “RBHHM: A novel remote cardiac cycle detection model based on heartbeat harmonics,” *Biomedical Signal Processing and Control*, vol. 78, p. 103936, Sep. 2022.
- [74] S. Schellenberger, K. Shi, T. Steigleider, A. Malessa, F. Michler, L. Hameyer, N. Neumann, F. Lurz, R. Weigel, C. Ostgathe, and A. Koelpin, “A dataset of clinically recorded radar vital signs with synchronised reference sensor signals,” *Scientific Data*, vol. 7, no. 1, pp. 1–11, Sep. 2020.
- [75] S. Yao, R. Guan, Z. Wu, Y. Ni, Z. Zhang, Z. Huang, X. Zhu, Y. Yue, E. G. Lim, H. Seo *et al.*, “Waterscenes: A multi-task 4D radar-camera fusion dataset and benchmark for autonomous driving on water surfaces,” *IEEE Transactions on Intelligent Transportation Systems*, Jul. 2023.
- [76] Y. Song, K. Zhang, T. Yan, T. Jin, Y. Dai, and T. Wang, “Dual-modal human pose reconstruction using 3D UWB radar images and point clouds,” *IEEE Transactions on Instrumentation and Measurement*, Mar. 2025.

- [77] C. Chen, J. Shi, L. Feng, H. Yi, C. Wang, and H. Chen, “A two-stage fault diagnosis method with rough and fine classifiers for phased array radar transceivers,” *IEEE Transactions on Instrumentation and Measurement*, Oct. 2024.
- [78] L. Zhao, R. Lyu, Q. Lin, A. Zhou, H. Zhang, H. Ma, J. Wang, C. Shao, and Y. Tang, “mmArrhythmia: Contactless arrhythmia detection via mmWave sensing,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, vol. 8, no. 1, pp. 1–25, Mar. 2024.
- [79] J. Larson, M. Menickelly, and S. M. Wild, “Derivative-free optimization methods,” *Acta Numerica*, vol. 28, pp. 287–404, Jun. 2019.
- [80] D. Makowski, T. Pham, Z. J. Lau, J. C. Brammer, F. Lespinasse, H. Pham, C. Schölzel, and S. A. Chen, “NeuroKit2: A Python toolbox for neurophysiological signal processing,” *Behavior Research Methods*, pp. 1–8, Feb. 2021.
- [81] P. Khairy and A. J. Marelli, “Clinical use of electrocardiography in adults with congenital heart disease,” *Circulation*, vol. 116, no. 23, pp. 2734–2746, Dec. 2007.
- [82] C. Xu, H. Li, Z. Li, H. Zhang, A. S. Rathore, X. Chen, K. Wang, M.-c. Huang, and W. Xu, “CardiacWave: A mmWave-based scheme of non-contact and high-definition heart activity computing,” *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies (IMWUT)*, vol. 5, no. 3, pp. 1–26, Sep. 2021.
- [83] L. Li, Y. Shuang, Q. Ma, H. Li, H. Zhao, M. Wei, C. Liu, C. Hao, C.-W. Qiu, and T. J. Cui, “Intelligent metasurface imager and recognizer,” *Light: Science & Applications*, vol. 8, no. 1, pp. 1–9, Oct. 2019.
- [84] I. Daubechies, J. Lu, and H.-T. Wu, “Synchrosqueezed wavelet transforms: An empirical mode decomposition-like tool,” *Applied and Computational Harmonic Analysis*, vol. 30, no. 2, pp. 243–261, Aug. 2011.

- [85] Y. Xue, R. Yang, X. Chen, Z. Tian, and Z. Wang, “A novel local binary temporal convolutional neural network for bearing fault diagnosis,” *IEEE Transactions on Instrumentation and Measurement*, Jul. 2023.
- [86] G. R. Terrell and D. W. Scott, “Variable kernel density estimation,” *The Annals of Statistics*, pp. 1236–1265, Sep. 1992.
- [87] L. Li, Y. Zhang, and S. Wang, “The Euclidean space is evil: hyperbolic attribute editing for few-shot image generation,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Oct. 2023, pp. 22714–22724.
- [88] X. Chen, R. Yang, Y. Xue, B. Song, and Z. Wang, “TFPred: Learning discriminative representations from unlabeled data for few-label rotating machinery fault diagnosis,” *Control Engineering Practice*, vol. 146, p. 105900, Feb. 2024.
- [89] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, and Y. Wei, “Deformable convolutional networks,” in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 764–773.
- [90] D. Yang, J. Lu, H. Dong, and Z. Hu, “Pipeline signal feature extraction method based on multi-feature entropy fusion and local linear embedding,” *Systems Science & Control Engineering*, vol. 10, no. 1, pp. 407–416, Apr. 2022.
- [91] P. E. McSharry, G. D. Clifford, L. Tarassenko, and L. A. Smith, “A dynamical model for generating synthetic electrocardiogram signals,” *IEEE Transactions on Biomedical Engineering*, vol. 50, no. 3, pp. 289–294, Mar. 2003.
- [92] Y. Yao, L. Rosasco, and A. Caponnetto, “On early stopping in gradient descent learning,” *Constructive Approximation*, vol. 26, no. 2, pp. 289–315, Apr. 2007.
- [93] I. Loshchilov and F. Hutter, “SGDR: Stochastic gradient descent with warm restarts,” *arXiv preprint arXiv:1608.03983*, Aug. 2016.
- [94] B. Lin and Y. Zhang, “LibMTL: A Python library for multi-task learning,” *Journal of Machine Learning Research*, vol. 24, no. 209, pp. 1–7, Jul. 2023.

- [95] R. Guan, R. Zhang, N. Ouyang, J. Liu, K. L. Man, X. Cai, M. Xu, J. Smith, E. G. Lim, Y. Yue *et al.*, “Talk2radar: Bridging natural language with 4D mmwave radar for 3D referring expression comprehension,” *arXiv preprint arXiv:2405.12821*, Jul. 2024.
- [96] N. Silberman, D. Hoiem, P. Kohli, and R. Fergus, “Indoor segmentation and support inference from RGBD images,” in *Proceedings of the European Conference on Computer Vision*. Springer, Oct. 2012, pp. 746–760.
- [97] C. Yeshwanth, Y.-C. Liu, M. Nießner, and A. Dai, “Scannet++: A high-fidelity dataset of 3D indoor scenes,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, Oct. 2023, pp. 12–22.
- [98] L.-Y. Liu, W.-Z. Liu, and L. Feng, “A primary task driven adaptive loss function for multi-task speech emotion recognition,” *Engineering Applications of Artificial Intelligence*, vol. 127, p. 107286, Jan 2024.
- [99] M. Guo, A. Haque, D.-A. Huang, S. Yeung, and L. Fei-Fei, “Dynamic task prioritization for multitask learning,” in *Proceedings of the European Conference on Computer Vision*, Sep. 2018, pp. 270–287.
- [100] M. Lin, R. Chi, and N. Sheng, “Data driven latent variable adaptive control for nonlinear multivariable processes,” *International Journal of Systems Science*, pp. 1–18, Aug. 2024.
- [101] R. Caruana, “Multitask learning: A knowledge-based source of inductive bias,” in *Proceedings of the Tenth International Conference on Machine Learning*. Citeseer, 1993, pp. 41–48.
- [102] D. Senushkin, N. Patakin, A. Kuznetsov, and A. Konushin, “Independent component alignment for multi-task learning,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Sep. 2023, pp. 20 083–20 093.

- [103] H. Fernando, H. Shen, M. Liu, S. Chaudhury, K. Murugesan, and T. Chen, “Mitigating gradient bias in multi-objective learning: A provably convergent approach,” in *International Conference on Learning Representations*, May 2023.
- [104] L. Liu, Y. Li, Z. Kuang, J.-H. Xue, Y. Chen, W. Yang, Q. Liao, and W. Zhang, “Towards impartial multi-task learning,” in *International Conference on Learning Representations*, Oct. 2021.
- [105] B. Liu, X. Liu, X. Jin, P. Stone, and Q. Liu, “Conflict-averse gradient descent for multi-task learning,” *Advances in Neural Information Processing Systems*, vol. 34, pp. 18 878–18 890, Dec. 2021.
- [106] X. Lin, X. Zhang, Z. Yang, F. Liu, Z. Wang, and Q. Zhang, “Smooth Tchebycheff scalarization for multi-objective optimization,” in *International Conference on Machine Learning*, Jul. 2024.
- [107] S. Liu, E. Johns, and A. J. Davison, “End-to-end multi-task learning with attention,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2019, pp. 1871–1880.
- [108] S. Chennupati, G. Sistu, S. Yogamani, and S. A Rawashdeh, “Multinet++: Multi-stream feature aggregation and geometric loss strategy for multi-task learning,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Nov. 2019, pp. 0–0.
- [109] A. Kendall, Y. Gal, and R. Cipolla, “Multi-task learning using uncertainty to weigh losses for scene geometry and semantics,” in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 7482–7491.
- [110] Z. Chen, V. Badrinarayanan, C.-Y. Lee, and A. Rabinovich, “Gradnorm: Gradient normalization for adaptive loss balancing in deep multitask networks,” in *International Conference on Machine Learning*. PMLR, Jul. 2018, pp. 794–803.

- [111] J. Wang, Y. Zhuang, and Y. Liu, “FSS-Net: A fast search structure for 3D point clouds in deep learning,” *International Journal of Network Dynamics and Intelligence*, pp. 100 005–100 005, Jun. 2023.
- [112] X. Dong, R. Wu, C. Xiong, H. Li, L. Cheng, Y. He, S. Qian, J. Cao, and L. Mo, “Gdod: Effective gradient descent using orthogonal decomposition for multi-task learning,” in *Proceedings of the 31st ACM International Conference on Information & Knowledge Management*, Oct. 2022, pp. 386–395.
- [113] P. H. Schönemann, “A generalized solution of the orthogonal Procrustes problem,” *Psychometrika*, vol. 31, no. 1, pp. 1–10, Mar. 1966.
- [114] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” *arXiv preprint arXiv:1412.6980*, 2014.
- [115] W. Jeong and K.-J. Yoon, “Quantifying task priority for multi-task optimization,” in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Jun. 2024, pp. 363–372.
- [116] X. Qian and B. Cui, “A mobile sensing approach to distributed consensus filtering of 2d stochastic nonlinear parabolic systems with disturbances,” *Systems Science & Control Engineering*, vol. 11, no. 1, p. 2167885, Jan. 2023.
- [117] D. Fitzgerald, “Harmonic/percussive separation using median filtering,” in *2010 13th International Conference on Digital Audio Effects (DAFx-10)*, Sep. 2010.
- [118] F. He, T. Liu, and D. Tao, “Why Resnet works? residuals generalize,” *IEEE Transactions on Neural Networks and Learning Systems*, vol. 31, no. 12, pp. 5349–5362, Feb. 2020.
- [119] P. O. Hoyer, “Non-negative matrix factorization with sparseness constraints,” *Journal of machine learning research*, vol. 5, no. Nov, pp. 1457–1469, Apr. 2004.
- [120] S.-H. Hwang, M. Kim, and S. E. Whang, “Rc-mixup: A data augmentation strategy against noisy data for regression tasks,” in *Proceedings of the 30th*

- ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, Aug. 2024, pp. 1155–1165.
- [121] N. Schneider, S. Goshtasbpour, and F. Perez-Cruz, “Anchor data augmentation,” *Advances in Neural Information Processing Systems*, vol. 36, Nov. 2024.
 - [122] H. Yao, Y. Wang, L. Zhang, J. Y. Zou, and C. Finn, “C-mixup: Improving generalization in regression,” *Advances in Neural Information Processing Systems*, vol. 35, pp. 3361–3376, Nov. 2022.
 - [123] S. Yang, D. Zhang, Y. Li, Y. Hu, Q. Sun, and Y. Chen, “iSense: Enabling radar sensing under mutual device interference,” *IEEE Transactions on Mobile Computing*, Mar. 2024.