A Thesis for the Degree of Ph.D. in Engineering

Contactless Vital Signal Detection and Activity Recognition via Deep Learning with Radar and Doppler Ultrasound

September 2025

Graduate School of Science and Technology
Keio University

Shi Xintong

別表６
（３）

Keio University

Thesis Abstract

No.

Registration Number	□ "KOU" □ "OTSU" No. *Office use only	Name	Shi Xintong

Thesis Title

Contactless Vital Signal Detection and Activity Recognition via Deep Learning with Radar and Doppler Ultrasound

Thesis Summary

This thesis investigates robust, non-contact sensing methodologies for physiological signal estimation and human behavior understanding using Doppler ultrasound (DUS) and millimeter-wave (mmWave) radar technologies. In response to growing demand for non-contact, privacy-preserving monitoring in healthcare, smart home, and rehabilitation applications, this work aims to develop deep learning-based systems that can reliably interpret sparse, noisy, and environment-sensitive signals in contactless sensing systems.

Chapter 2 of this thesis addresses fetal heart rate (FHR) estimation from DUS signals. We propose a robust framework that integrates time-frequency domain information of DUS signals via short-time Fourier transform (STFT), dual autocorrelation-based peak tracking, and an unsupervised signal quality assessment (SQA) module based on variational autoencoders (VAEs) and self-organizing maps (SOMs). This design enables fetal heart rate estimation while adaptively excluding low-quality signal segments, achieving improved accuracy and robustness without requiring ground truth annotations.

Chapter 3 presents a rough-to-fine heart rate estimation framework using MIMO FMCW radar. This system integrates Curve-Length (CL)-based person localization and a deep Convolutional Neural Network (CNN) model that refines heart rate (HR) estimation using both filtered IQ signals and coarse HR estimates. The approach supports short-window inference, making it suitable for real-time health monitoring.

Chapter 4 focuses on human skeleton estimation using multi-frame mmWave radar point cloud sequences. A CNN and Bi-directional Long Short-Term Memory (BiLSTM) architecture is developed to capture both spatial structure and temporal dynamics, while an Long Short-Term Memory (LSTM)-based point cloud reliability assessment mechanism selectively suppresses unreliable input frames. The proposed model avoids voxelization and maintains real-time performance, significantly enhancing accuracy in low-quality data scenarios.

In Chapter 5, we extend the framework introduced in Chapter 4 and introduce *mmPoint-Attention*, a unified Transformer-based architecture for human pose estimation and activity recognition. This model

incorporates attention and Transformer encoder layers for both human pose estimation and activity recognition. Evaluations on MM-Fi and mRI benchmark datasets demonstrate that *mmPoint-Attention* outperforms prior voxelized or two-stage models while maintaining computational efficiency.

Collectively, the contributions of this thesis form a comprehensive foundation for robust and efficient contactless sensing systems. By advancing deep learning-based technologies, this work provides practical methodologies for future smart sensing platforms in healthcare intelligence.