

Titre: Localisation des humains en présence d'artéfacts
Title: environnementaux à l'aide du radar FMCW

Auteur: Wissal Zarrami
Author:

Date: 2025

Type: Mémoire ou thèse / Dissertation or Thesis

Référence: Zarrami, W. (2025). Localisation des humains en présence d'artéfacts
Citation: environnementaux à l'aide du radar FMCW [Mémoire de maîtrise, Polytechnique
Montréal]. PolyPublie. <https://publications.polymtl.ca/66165/>

 **Document en libre accès dans PolyPublie**
Open Access document in PolyPublie

URL de PolyPublie: <https://publications.polymtl.ca/66165/>
PolyPublie URL:

**Directeurs de
recherche:** Guillaume-Alexandre Bilodeau
Advisors:

Programme: Génie informatique
Program:

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

**Localisation des humains en présence d'artéfacts environnementaux à l'aide du
radar FMCW**

WISSAL ZARRAMI

Département de génie informatique et génie logiciel

Mémoire présenté en vue de l'obtention du diplôme de *Maîtrise en sciences appliquées*
Génie informatique

Juin 2025

POLYTECHNIQUE MONTRÉAL

affiliée à l'Université de Montréal

Ce mémoire intitulé :

**Localisation des humains en présence d'artéfacts environnementaux à l'aide du
radar FMCW**

présenté par **Wissal ZARRAMI**

en vue de l'obtention du diplôme de *Maîtrise en sciences appliquées*
a été dûment accepté par le jury d'examen constitué de :

Lama SÉOUD, présidente

Guillaume-Alexandre BILODEAU, membre et directeur de recherche

François NOUGAROU, membre et codirecteur de recherche

Messaoud AHMED-OUAMEUR, membre externe

REMERCIEMENTS

Je tiens tout d’abord à exprimer ma profonde gratitude envers mon directeur de recherche, le Professeur Guillaume-Alexandre BILODEAU, pour ses précieux conseils scientifiques, ses éclairages rigoureux et son accompagnement bienveillant tout au long de ce projet. Son soutien constant, tant sur le plan académique que personnel, ainsi que son appui décisif dans l’obtention d’une bourse d’études, ont été des piliers essentiels à la réalisation de ce mémoire.

Je tiens également à remercier chaleureusement Professeur François NOUGAROU pour son soutien indéfectible et ses conseils avisés durant toutes les étapes de ce travail.

Je souhaite également remercier chaleureusement les membres du jury, Professeur Lama SÉOUD et Professeur Messaoud AHMED-OUAMEUR, pour avoir accepté d’examiner ce travail avec attention et rigueur. Leurs retours et réflexions enrichissants ont grandement contribué à l’aboutissement de cette recherche.

Un merci particulier aux membres du laboratoire LITIV, dont les échanges multidisciplinaires techniques, scientifiques et humains ont été d’une aide précieuse pour surmonter des défis complexes. Je remercie aussi Morphée+ .Inc et Mitacs pour leur soutien financier.

Je tiens à témoigner ma profonde reconnaissance à mon ami très proche, Abdellatif, et à ma famille, pour leur soutien inconditionnel durant cette aventure.

RÉSUMÉ

Les systèmes de localisation intérieure destinés à la surveillance des personnes âgées doivent allier précision, respect de la vie privée et robustesse face aux artefacts. Ils assurent une détection fiable des chutes et des situations d’urgence pour des individus souvent isolés. Les dispositifs portables, comme les montres connectées, préservent la confidentialité mais souffrent de faux positifs et d’une faible adoption par les utilisateurs, en particulier ceux présentant des troubles neuro-cognitifs qui limitent leur capacité à utiliser ces technologies de manière fiable. Les systèmes de surveillance par caméras portent atteinte à la vie privée et échouent en cas d’occultation. Le radar à onde continue modulée en fréquence (FMCW) constitue une alternative intéressante, car il préserve l’anonymat des personnes. Toutefois, les approches radar traditionnelles s’appuient sur un prétraitement manuel (transformées de Fourier, filtre des signaux, la formation de faisceaux) qui écarte les informations de phase subtiles et peine à distinguer les mouvements humains des interférences dues au mobilier ou aux appareils électroménagers dans des environnements intérieurs complexes.

Ce travail présente la première étape d’un projet de surveillance respectueuse de la vie privée pour les personnes âgées, en proposant un cadre d’apprentissage profond dédié à la localisation d’individus en intérieur. En traitant directement les signaux bruts du radar, notre méthode préserve à la fois l’amplitude et la phase : l’amplitude permet de repérer l’énergie renvoyée, la phase encode les changements de position. L’élément central de notre approche est un encodeur double branche Phase-Amplitude, qui sépare et exploite ces deux types d’informations pour capturer les signatures micro-Doppler des mouvements humains et les décalages de phase dus aux déplacements. Ces représentations sont ensuite affinées par un encodeur Transformer appliqué à chaque antenne, apprenant les dépendances spatio-temporelles pour surmonter les occlusions et les réflexions parasites, assurant une localisation robuste dans un environnement clos réaliste, sans compromettre l’anonymat des utilisateurs. En traitant la localisation comme une régression sur une carte de chaleur plutôt que comme une détection de pics par seuil, notre système assure une surveillance continue et confidentielle sans caméras ni dispositifs portables. Validé avec des données expérimentales dans deux locaux différents, l’approche proposée permet d’atteindre une erreur absolue moyenne de distance de 0,76 m et une erreur angulaire de 4,2 ° pour un individu dans son meilleur des cas, soit une amélioration de 45 % par rapport aux radars conventionnels.

ABSTRACT

Indoor localization systems for monitoring elderly people must balance accuracy, privacy, and robustness against artifacts. They provide reliable fall and emergency detection for individuals who often live alone. Wearable devices such as smartwatches protect privacy, but suffer from false positives and low user adoption, especially among those with neuro-cognitive impairments that limit their ability to use these technologies reliably.

Camera-based monitoring systems invade privacy and fail when sightlines are blocked. Frequency-Modulated Continuous-Wave (FMCW) radar offers an attractive alternative because it preserves user anonymity. However, traditional radar methods rely on manual preprocessing, including Fourier transforms, signal filtering, and beamforming, which discards subtle phase information and struggles to distinguish human motion from clutter caused by furniture or appliances in complex indoor environments.

This work marks the first step toward a privacy-preserving monitoring solution for elderly care by introducing a deep-learning framework for indoor localization. By operating directly on raw radar signals, our method retains both amplitude, which captures returned signal strength, and phase, which encodes positional changes. At its core lies a dual-branch Phase-Amplitude encoder that separates and leverages these two information streams to detect micro-Doppler signatures of human movement and phase shifts linked to displacement. A per-antenna Transformer encoder then refines these representations, learning spatio-temporal dependencies to overcome occlusions and multipath reflections, ensuring robust localization in realistic, enclosed settings without compromising anonymity.

By formulating localization as a heatmap regression task rather than threshold-based peak detection, our system delivers continuous, confidential monitoring without cameras or wearables. Validated with experimental data in two different rooms, the proposed approach achieves a best-case mean absolute distance error of 0.76 m and an angular error of 4.2 °, representing a 45 % improvement over conventional radar methods.

TABLE DES MATIÈRES

REMERCIEMENTS	iii
RÉSUMÉ	iv
ABSTRACT	v
TABLE DES MATIÈRES	vi
LISTE DES TABLEAUX	ix
LISTE DES FIGURES	x
LISTE DES SIGLES ET ABRÉVIATIONS	xv
CHAPITRE 1 INTRODUCTION	1
1.1 Localisation radar des humains : contexte et principes	1
1.2 Éléments de la problématique	3
1.3 Objectifs de recherche et contributions	5
1.4 Plan du mémoire	5
CHAPITRE 2 REVUE DE LITTÉRATURE	7
2.1 Définitions et concepts de base	7
2.1.1 Principe de base du radar FMCW	7
2.1.2 Représentation I/Q et exploitation des signaux FMCW	10
2.1.3 Bases méthodologiques du traitement spatio-temporel dans les radars FMCW	11
2.1.4 Définition et présentation des cartes Range-Doppler, Range-Azimuth et Range-Angle	13
2.2 Localisation des cibles sur la carte Range-Doppler : Approches classiques et méthodes basées sur l'apprentissage profond	15
2.2.1 Méthodes classiques et mathématiques	15
2.2.2 Exploitation des séquences range-Doppler par apprentissage profond	19
2.3 Localisation des cibles sur la carte Range-Angle/Azimuth : Approches classiques et méthodes basées sur l'apprentissage profond	23
2.3.1 Méthodes de traitement du signal RAM	23

2.3.2	Localisation par apprentissage profond et génération de cartes de confiance	26
2.4	Estimation de la localisation à partir des données brutes	29
2.5	Conclusion	33
CHAPITRE 3	DÉMARCHE GÉNÉRALE	34
3.1	Acquisition de données radar FMCW pour des scénarios variés	34
3.2	Localisation des personnes basée sur les cartes de chaleurs et DBSCAN	36
3.3	Architecture du modèle	36
CHAPITRE 4	ARTICLE 1 : MULTI-BRANCH FEATURE FUSION NETWORK FOR RADAR-BASED HUMAN LOCALIZATION	39
4.1	Introduction	40
4.2	FMCW Radar Background	42
4.2.1	Operating Principle of FMCW Radar	42
4.2.2	Beat Frequency and IF Signal	42
4.2.3	Raw Data Acquisition and Signal Processing	43
4.3	Related Works	44
4.4	Methodology	45
4.4.1	Magnitude-Phase vs. Real-Imaginary Representations	46
4.4.2	Phase-Magnitude Encoder (PME)	47
4.4.3	Antenna-wise Transformer Encoder	49
4.4.4	Channel Feature Fusion	50
4.4.5	Heatmap and Target Localization	50
4.4.6	Polar Heatmap Ground Truth Generation	51
4.5	Experimental Results	52
4.5.1	Data Acquisition	52
4.5.2	Dataset Partitioning	53
4.5.3	Experimental Setup	54
4.5.4	Evaluation Metrics	55
4.5.5	Performance for the various test scenarios	56
4.5.6	Comparison with SOTA methods	60
4.5.7	Discussion	62
4.6	Conclusion	62
CHAPITRE 5	DISCUSSION GÉNÉRALE	64
5.1	Retour sur les résultats présentés dans l'article	64
5.1.1	Performances globales et métriques clés	64

5.2	Résultats additionnels	65
5.2.1	Variabilité inter-individuelle et biais d'entraînement	65
5.3	Perspectives futures : vers une détection robuste des chutes par radar FMCW	66
CHAPITRE 6	CONCLUSION	68
RÉFÉRENCES	69

LISTE DES TABLEAUX

Tableau 2.1	Tableau de synthèse des approches de localisation sur cartes Range–Doppler (Section 2.2).	22
Tableau 3.1	Paramètres clés du radar FMCW utilisé	35
Tableau 4.1	FMCW Radar Parameters for Data Acquisition	53
Tableau 4.2	MAE and Standard Deviation for Distance (m) and Angle (degrees) with person-based training on persons 1 (P1), 2 (P2), and 3 (P3), and separate testing per person.	56
Tableau 4.3	MAE and Std for Distance (m) and Angle (°). Model trained on persons 1 (P1), 2 (P2), and 3 (P3) (including both P1 and P2 simultaneously (P1+2)). Testing involves pairwise predictions : errors are reported for one person while the other is present.	58
Tableau 4.4	MAE and Std for Distance (m) and Angle (°) under different train-test splits among Persons 1, 2, and 3 (P1,P2,P3).	58
Tableau 4.5	MAE and Standard Deviation for Distance and Angle with Cross-Lab Training and Testing	59
Tableau 5.1	MAE et écart-type pour la distance (m) et l’angle (°) lorsque le modèle est entraîné sur une seule personne (P1, P2 ou P3) et testé sur les deux autres.	66

LISTE DES FIGURES

Figure 1.1	Principe de fonctionnement du radar FMCW : émission et réception des signaux pour la détection de la présence humaine Figure reproduite avec permission issue de [1]. © 2022 Sensors, MDPI.	2
Figure 2.1	Schéma de principe du fonctionnement d'un radar FMCW. Le signal chirp est généré et émis par l'antenne TX, se propage dans l'environnement, puis est réfléchi par les cibles (véhicule, piéton). Le signal réfléchi est capté par les antennes RX, amplifié (LNA), mélangé avec le signal émis (Mixer), filtré (LPF), puis converti en numérique (ADC). Les données sont ensuite traitées dans un cube de données pour extraire les informations de distance (R), vitesse (V) et angle (θ). Figure utilisée avec permission depuis [2].	8
Figure 2.2	Illustration du principe d'un radar FMCW montrant l'évolution temporelle de la fréquence du signal émis (Transmitted signal) et du signal réfléchi (Reflected signal). Le décalage temporel Δt est proportionnel à la distance R entre le radar et la cible, tandis que le décalage fréquentiel Δf permet l'estimation de la vitesse relative. Source : Renesas Electronics	9
Figure 2.3	Représentation des composantes I/Q et schéma de leur exploitation dans un radar FMCW.	11
Figure 2.4	Illustration du principe mathématique de base du radar FMCW montrant la relation entre le signal chirp émis, l'écho réfléchi et la fréquence de battement, proportionnelle à la distance cible. Source : wirelesspi.com.	12
Figure 2.5	Schéma mathématique montrant l'application de la FFT spatiale sur les signaux issus de multiples antennes pour déterminer précisément l'angle d'arrivée grâce à l'évolution spatiale de la phase. Source : wirelesspi.com.	12
Figure 2.6	Illustration de la mesure de vitesse par FFT Doppler sur une série temporelle de chirps FMCW, permettant l'extraction des variations fréquentielles induites par le déplacement relatif des cibles. Source : wirelesspi.com.	13

Figure 2.7	Obtention des cartes Range-Doppler (RDM), Range-Azimuth (RAM) et Range-Elevation (REM) à partir d'un cube de données radar mm-Wave FMCW. Illustration tirée de l'article <i>Multi-Feature Transformer-Based Learning for Continuous Human Motion Recognition with High Similarity Using mmWave FMCW Radar</i> [3], publié dans <i>Sensors</i> (2022), vol. 22, n°21, p. 8409 (doi:10.3390/s22218409). Figure utilisée avec permission. Sous licence Creative Commons Attribution (CC BY 4.0). © 2022 MDPI.	14
Figure 2.8	Chaîne de traitement FFT multidimensionnelle pour l'extraction de la portée, du Doppler et de l'azimut (adapté de Saponara & Neri, 2017).	16
Figure 2.9	Visualisation de la cellule test (CUT), de la fenêtre de garde et de la fenêtre de référence dans les détecteurs CFAR.	17
Figure 2.10	Chaînes de traitement pour la localisation avec un radar FMCW : (a) détection, clustering, association de données et suivi sur carte distance-azimut ; (b) détection, clustering, association de données et suivi sur carte distance-Doppler.	18
Figure 2.11	(a) Diagramme du workflow de prétraitement du signal montrant la FFT 2D, le filtrage des échos statiques, l'OS-CFAR, le beamforming MVDR et le mapping de position pour générer un nuage de points.	19
Figure 2.12	Pipeline de traitement radar pour la génération des cartes Range-Doppler (RDM) et des tenseurs Range-Angle-Doppler (RAD). La méthode DA-ROD utilise les données à la sortie du spectre Range-Doppler, réduisant ainsi la complexité computationnelle en omettant l'estimation angulaire.	20
Figure 2.13	Architecture du réseau ST-ConvLSTM pour la reconnaissance des mouvements humains à partir des cartes Range-Doppler dynamiques. (a) Flux de traitement du ST-ConvLSTM. (b) Module d'attention spatiale pour l'amélioration des caractéristiques spatiales. (c) Structure de la cellule ConvLSTM combinant convolutions et mémoire à long terme. (d) Module d'attention temporelle pondérant les trames les plus informatives.	21
Figure 2.14	Architecture du modèle HOOD pour la reconnaissance d'activités à partir des cartes Range-Doppler. Le traitement est séparé en deux flux : un pour les mouvements globaux via le suivi des pics de réflexion, et un autre pour les micro-mouvements à partir de l'analyse fine des spectres Doppler. Les deux branches sont fusionnées lors de la classification.	23

Figure 2.15	Illustration de la formation de faisceau pour la construction d’une carte range–angle : application d’un vecteur de pilotage (steering vector) sur une plage angulaire de $\pm 90^\circ$ pour localiser le torse humain.	25
Figure 2.16	Architecture de DeepMUSIC : le réseau profond (Deep Networks) traite la matrice de covariance R_y pour produire un spectre MUSIC amélioré (Deep Spectra), suivi d’une étape de détection de pics (Peak Finding) pour extraire les angles DoA estimés.	26
Figure 2.17	Architecture du CNN proposé par Meiresone et al. pour la segmentation des cartes Range–Azimuth : après beamforming de Capon, la carte RA est traitée par des blocs convolutifs contractants (Conv2D, Batch-Norm, LeakyReLU, MaxPool), puis deux couches entièrement connectées (fc1 , fc2) prédisent, pour chaque case (bin en anglais) azimuthal, la distance normalisée au premier obstacle.	27
Figure 2.18	Pipeline de RODNet : un extrait de RA-Maps est traité par le réseau RODNet pour produire des cartes de confiance (ConfMaps), qui sont ensuite post-traitées pour générer les résultats finaux de détection. . .	28
Figure 2.19	Architecture de GANomaly pour la détection et la correction d’anomalies dans les ConfMaps. Le Generative Net (orange) assure la reconstruction et le transfert latent, tandis que le Discriminative Net (vert) différencie les cartes réelles des reconstructions	28
Figure 2.20	Schéma de l’architecture de TransRadar : Adaptive-Directional Attention Block pour la segmentation sémantique multi-vues radar FMCW.	29
Figure 2.21	Architecture d’ADCNet : le module de traitement du signal apprenable (Learnable Signal Processing) transforme les données ADC brutes en une représentation distance–Doppler latente, qui est ensuite affinée par un backbone profond. Deux têtes de sortie assurent respectivement la détection (heatmap + régression) et la segmentation du cube distance–azimut–Doppler reconstruit.	30
Figure 2.22	Architecture du DCNN proposé par Seo et al. [4] pour l’imagerie radar : après la rétroprojection (BP), la normalisation L2 et la prise de valeur absolue (Abs), le réseau applique des blocs convolutionnels (CL) avec normalisation par lot (BN) et fonctions d’activation (AF), suivis d’un aplatissement (Flatten) et de couches entièrement connectées (FC). La sortie est optimisée par une perte d’entropie croisée binaire (BCE). Figure utilisée avec permission	31

Figure 3.1	Configuration d'acquisition des données en laboratoire : le radar FMCW est monté sur un trépied et les positions cibles sont à différentes distances dans l'intervalle de (0,5–5 m) et pour les angles de (0° – 136°).	35
Figure 3.2	Architecture proposée, fondée sur les transformeurs, pour la localisation par radar. Les signaux complexes bruts sont décomposés en amplitude et en phase, puis traités par le PM-Encodeur avant d'être affinés par un module à base de transformeur. Les caractéristiques issues de plusieurs antennes sont fusionnées pour générer une carte de chaleur représentant les probabilités de localisation des personnes.	37
Figure 4.1	(Left) Transmitted (TX) and received (RX) chirps : The beat frequency f_b arises from the delay τ_1 between TX and RX signals. (Right) Data cube structure : Dimensions include antennas (spatial), slow time (Doppler), and fast time (range).	43
Figure 4.2	Proposed Transformer-based architecture for radar-based localization. Raw complex signals are decomposed into magnitude and phase, processed through the PM-Encoder, and refined using a transformer-based module. Features from multiple antennas are fused to generate a heatmap of prediction probabilities for human localization.	46
Figure 4.3	Comparison of spectrogram representations after applying 1D convolution. (Top-Left) Magnitude spectrogram, (Top-Right) Phase spectrogram, (Bottom-Left) Real part spectrogram, (Bottom-Right) Imaginary part spectrogram.	48
Figure 4.4	Antenna Feature Fusion : Features from multiple antennas are concatenated and processed through a $1 \times 1 \times 1$ convolution with batch normalization and ReLU.	51
Figure 4.5	Heatmap generation for a person in the position range=3 and angle=45 degrees in LAB1	52
Figure 4.6	Data acquisition environments : LAB1 (right), LAB2 (left). LAB1 has an higher clutter levels, with additional interference caused by furniture, walls, and electromagnetic reflections. The second environment (LAB2) exhibited lower clutter levels. Ventilation systems, and other active equipment contributed additional noise.	54
Figure 4.7	Ground truth vs. predicted location : The left image shows an image of the scene, while the middle and right heatmaps compare the actual and predicted positions.	57

Figure 4.8	Polar Plot of Ground Truth vs. Predicted Points for Unseen angles or distances	59
Figure 4.9	SOTA comparison : Std and MAE Analysis of the distance Estimation	61
Figure 4.10	SOTA comparison : MAE and Std Analysis of the Angle Estimation .	61

LISTE DES SIGLES ET ABRÉVIATIONS

FMCW	Frequency Modulated Continuous Wave
FFT	Fast Fourier Transform
RDM	Range–Doppler Map
RAM	Range–Azimuth/Angle Map
REM	Range–Elevation Map
I/Q	In-Phase/Quadrature
LNA	Low Noise Amplifier
ADC	Analog-to-Digital Converter
PLL	(Phase-Locked Loop)
MIMO	Multiple-Input Multiple-Output
CFAR	Constant False Alarm Rate
MVDR	Minimum Variance Distortionless Response
CVNN	Complex-Valued Neural Network
RODNET	Radar Object Detection Network
DCNN	Deep Convolutional Neural Network
ST-ConvLSTM	Spatio-Temporal Convolutional LSTM
HOOD	Human presence and Out-Of-Distribution detection
DAROD	Deep Automotive Radar Object Detector
RDM	Range–Doppler Map
RAD	Range–Azimuth–Doppler cube
BP	Back-Projection
SNR	Signal-to-Noise Ratio
CFAR	Constant False-Alarm-Rate detector
OS-CFAR	Ordered-Statistics CFAR
GO-CFAR	Greatest-Of CFAR
CA-CFAR	Cell-Averaging CFAR
MAE	Mean Absolute Error
Std	Standard Deviation

CHAPITRE 1 INTRODUCTION

La localisation précise des humains dans des environnements complexes est un enjeu, notamment dans des contextes sensibles tels que la surveillance, l’assistance aux personnes ou l’interaction homme-machine. La localisation est particulièrement essentielle lorsqu’il s’agit de protéger les personnes âgées vivant seules, souvent exposées au risque de chutes graves pouvant entraîner la mort en l’absence d’intervention rapide.

Dans cette optique, les radars à ondes continues modulées en fréquence (FMCW – Frequency-Modulated Continuous Wave) offrent une solution prometteuse. Ils permettent de détecter efficacement la position des individus même en présence d’obstacles ou d’artefacts environnementaux variés tels que des équipements métalliques (armoires, tables), des appareils électroniques générant du bruit électromagnétique (ordinateurs), des objets mécaniques en mouvement (ventilateurs) susceptibles d’imiter les signatures Doppler humaines, des interférences électromagnétiques issues d’appareils sans fil (Wi-Fi, Bluetooth), des surfaces absorbantes (mousses acoustiques).

Cependant, l’interprétation des signaux radar dans ces environnements réels demeure complexe. La diversité des artefacts et la variabilité des matériaux peuvent affecter considérablement la précision des techniques actuelles. Ainsi, il est essentiel de développer des méthodes robustes capables de traiter efficacement les signaux bruts et d’extraire des informations pertinentes avec fiabilité.

L’objectif de ce mémoire est donc d’étudier et d’optimiser les techniques de localisation radar FMCW (distance et angle) des êtres humains dans un espace en intérieur réaliste. Une telle approche pourra, entre autres, servir à la détection précoce des chutes chez les personnes âgées vivant seules. En effet, la précision et la continuité de la localisation sont indispensables pour garantir la fiabilité de toute application ultérieure.

1.1 Localisation radar des humains : contexte et principes

En général, la localisation radar des humains repose sur l’analyse des signaux électromagnétiques réfléchis par les individus présents dans un environnement donné. L’objectif principal est de localiser avec précision ces personnes tout en distinguant clairement leur présence des artefacts physiques générés dans des environnements réels. Cela soulève plusieurs questions fondamentales : Comment détecter efficacement la présence humaine face à des artefacts réels ? Comment localiser précisément les individus sans se laisser perturber par des signaux

ambigus provenant de l'environnement réel ? Et surtout, comment exploiter les données brutes issues du radar, souvent complexes et bruitées, sans recourir à des étapes de prétraitement traditionnelles, tout en assurant une localisation robuste et fiable ?

Le radar FMCW est particulièrement adapté à cette tâche. Ce type de radar émet une onde continue dont la fréquence varie progressivement dans le temps. Cette technologie est avantageuse [5], car elle offre une mesure précise de la distance (haute résolution en distance) et de l'angle des cibles, tout en garantissant une grande portée de détection, même en présence d'obstacles. Contrairement à d'autres capteurs, les radars FMCW maintiennent une précision millimétrique à plusieurs mètres, un critère essentiel pour localiser fidèlement les individus avant d'analyser leur activité (ex. chutes), même dans des environnements complexes et bruités, tout en étant peu sensible aux variations d'éclairage et aux obstructions visuelles.

La figure 1.1 illustre les principales étapes du traitement des signaux radar FMCW pour la détection de la présence humaine. Le système radar est composé d'un émetteur (TX) et de plusieurs récepteurs (RX_1 , RX_2). Le signal RF émis par l'antenne TX est réfléchi lorsqu'il rencontre une cible (dans ce cas, une personne). Ce signal réfléchi est capté par les antennes RX, amplifié par un amplificateur faible bruit (LNA), converti en signal intermédiaire (IF) via un mélangeur (Mixer), puis amplifié par un amplificateur à gain variable (VGA). Le signal est ensuite converti en numérique par un convertisseur analogique-numérique (ADC), qui est traité par un ordinateur pour détecter la présence humaine. Le système est également alimenté par un générateur de fréquence (PLL + Ramp Generator) qui fournit une onde locale (LO) à l'émission et à la réception.

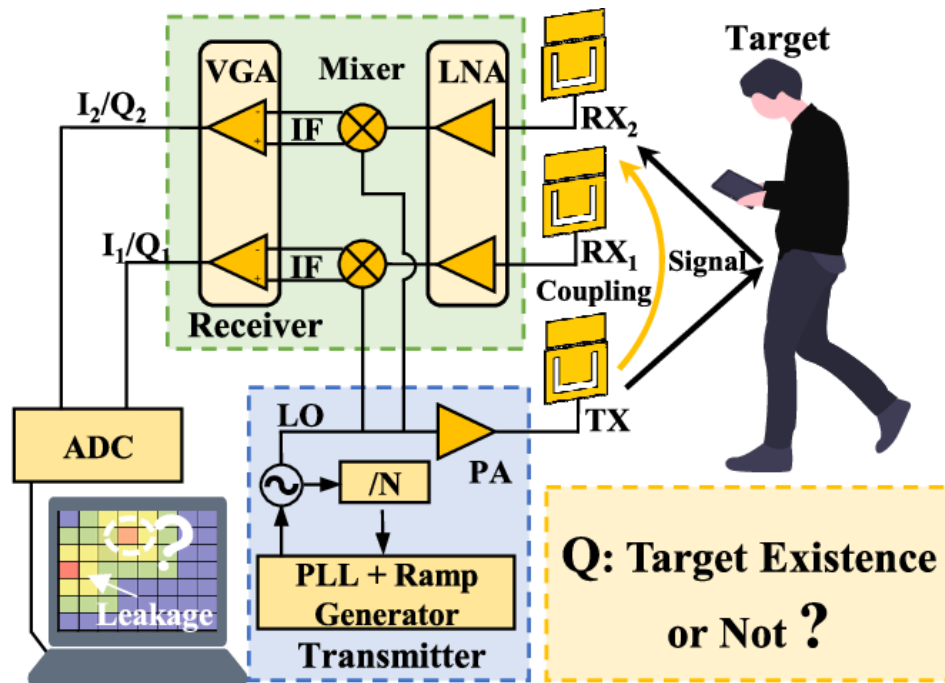


FIGURE 1.1 Principe de fonctionnement du radar FMCW : émission et réception des signaux pour la détection de la présence humaine Figure reproduite avec permission issue de [1]. © 2022 Sensors, MDPI.

Les signaux reçus (I/Q) sont ensuite numérisés par un convertisseur analogique-numérique (ADC) pour être traités sur un ordinateur. Ce traitement permet notamment de compenser les fuites du signal direct (coupling TX–RX) et d’analyser la présence ou l’absence d’une cible grâce à des algorithmes adaptés. Le système s’appuie sur un générateur de rampe avec PLL (Boucle à Verrouillage de Phase ou Phase-Locked Loop en anglais) pour contrôler la modulation en fréquence et générer l’onde chirp typique du radar FMCW.

L’objectif final est de répondre à la question suivante : "La cible existe-t-elle ou non ? Si oui, quelle est la distance et son angle par rapport au radar ?", ce qui constitue la première étape dans des applications plus complexes de localisation ou d’analyse de mouvement.

La localisation radar des humains implique ensuite la résolution de plusieurs aspects clés :

1. **Détection de présence** : Identifier rapidement si une personne est présente ou non dans une zone observée par le radar, indépendamment des détails précis sur sa position.
2. **Localisation** : Estimer avec la plus grande exactitude possible la position des individus détectés en termes de distance et d’angle par rapport au radar.
3. **Gestion des perturbations environnementales** : Identifier et réduire l’impact des divers artefacts pouvant perturber les signaux radar. Ces artefacts incluent les réflexions parasites générées par des structures fixes (murs, étagères métalliques, tables de travail), les interférences électromagnétiques liées à la densité d’équipements électroniques (ventilateurs, alimentations, réseaux Wi-Fi/Bluetooth). Ces artefacts, intrinsèques aux espaces clos et équipés, brouillent la distinction entre les cibles humaines et le bruit de fond, compromettant la précision de la localisation préalable à la détection de chutes [6].

Pour résoudre ces aspects, des méthodes avancées de traitement du signal radar sont utilisées. Traditionnellement, ces techniques reposent sur des approches déterministes, telles que la transformation de Fourier [7, 8]. Toutefois, les approches modernes basées sur l’intelligence artificielle, notamment les réseaux de neurones profonds [9, 10], permettent d’améliorer considérablement la précision et la robustesse des résultats.

1.2 Éléments de la problématique

Les systèmes de localisation humaine basés sur un radar FMCW font face à plusieurs défis techniques et environnementaux qui affectent la précision et la fiabilité des mesures. Le radar FMCW utilise l’analyse des signaux électromagnétiques réfléchis pour estimer la position des individus dans une scène. Cependant, ces signaux sont sensibles aux conditions de l’environnement et aux propriétés des objets présents. Les principaux défis rencontrés sont :

- **Extraction des informations spatiales (distance et angle) :** L'estimation précise de la distance nécessite une mesure du temps de vol du signal réfléchi, tandis que la détermination de l'angle de la cible repose sur une analyse des différences de phase perçues par un réseau d'antennes, ce qui représente un défi technique majeur en raison du bruit, des interférences et de la complexité des environnements réels. [11, 12].
- **Réflexions parasites provenant de structures fixes :** Les échos dominants générés par des équipements métalliques (étagères, tables), des murs, des rideaux, des ventilateurs et divers objets statiques peuvent masquer les cibles humaines, particulièrement dans des configurations spatiales complexes.
- **Variabilité des signatures radar humaines :** La diversité des postures (debout, assis, mouvement des membres) et des tailles corporelles modifie la réflectivité radar, introduisant des incertitudes dans l'estimation de la distance et de l'angle.
- **Instabilité du signal en environnement dynamique :** Les variations de l'environnement électromagnétique (extinction d'équipements, déplacement d'objets métalliques) perturbent la cohérence des mesures radar, générant des erreurs de localisation (dérive en distance/angle) lors de changements de configuration spatiale.

Pour surmonter ces défis, il est essentiel d'avoir des données radar suffisamment variées et représentatives des différents scénarios réels. Cependant, la disponibilité limitée des jeux de données radar publics constitue un obstacle majeur aux travaux de recherche, limitant les possibilités d'entraînement efficace des algorithmes par apprentissage. Cela empêche souvent d'étudier en profondeur les différents types de signaux radar associés à diverses postures humaines et conditions environnementales.

Un autre aspect important concerne l'absence actuelle de solutions exploitant directement les données radar brutes, sans recourir systématiquement à la transformation de Fourier rapide (FFT). Très peu de travaux se sont penchés sur l'utilisation directe des signaux bruts pour détecter et localiser les individus en environnement intérieur, tout en traitant simultanément la distance et de l'angle sans prétraitements intensifs ou filtrages complexes. Cette approche directe pourrait potentiellement améliorer la précision et préserver davantage les informations intrinsèques contenues dans les signaux radar.

Enfin, l'intégration d'approches basées sur l'intelligence artificielle et l'apprentissage automatique représente un potentiel important, notamment pour gérer efficacement la variabilité des données radar et améliorer la robustesse face aux artefacts environnementales.

1.3 Objectifs de recherche et contributions

À la lumière des défis existants de la localisation des individus par radar FMCW dans des environnements intérieurs complexes, ce projet vise à développer une méthode permettant d'améliorer la précision et la robustesse des systèmes de localisation en présence d'artefacts environnementaux. Plus précisément, l'objectif général est de concevoir une méthode basée sur l'apprentissage automatique pour le traitement avancé du signal afin d'améliorer la précision et la robustesse de la localisation d'individus par radar FMCW en environnements intérieurs réalistes, en discriminant efficacement les échos humains des artefacts environnementaux, avec comme sous-objectif d'évaluer ses performances comparativement aux approches traditionnelles basées sur des modèles mathématiques.

Les contributions de ce projet ont été formalisées dans un article actuellement soumis à un journal scientifique à comité de lecture. Elles se déclinent comme suit :

1. **Création d'une base de données originale** : collecte de signaux radar FMCW bruts dans deux environnements intérieurs complexes présentant des conditions d'encombrement variées, avec plusieurs individus de morphologies et dynamiques différentes. Cette base vise à faciliter le développement et l'évaluation de méthodes de localisation robustes.
2. **Développement d'une méthode innovante de localisation humaine** : proposition d'une approche de bout en bout fondée sur l'apprentissage automatique, traitant directement les signaux radar bruts pour améliorer la précision tout en préservant la vie privée.
3. **Validation expérimentale approfondie** : évaluation des performances de la méthode proposée à partir des bases de données créées, démontrant une précision sub-métrique et une faible erreur angulaire, même dans des scénarios impliquant plusieurs individus. Cette validation est faite en comparaison avec des méthodes conventionnelles.

1.4 Plan du mémoire

Ce mémoire est structuré en six chapitres, organisés afin de présenter progressivement l'introduction, les travaux réalisés dans la littérature et les résultats obtenus. Le deuxième chapitre propose les définitions et les concepts de base pour le projet ainsi qu'une revue de littérature détaillée, positionnant notre approche par rapport aux méthodes existantes de localisation humaine par radar, en mettant en lumière les limites rencontrées, notamment dans des environnements intérieurs réalistes. Le troisième chapitre est dédié à la méthodologie adoptée

pour concevoir, implémenter et évaluer la solution proposée. Ce chapitre décrit la stratégie expérimentale, le protocole de collecte de données, les outils algorithmiques et les critères de performance retenus. Le quatrième chapitre contient l'article scientifique issu de ce travail, actuellement soumis à une revue à comité de lecture. Ici est présenté de manière structurée la méthode développée, les expérimentations menées, ainsi que les résultats obtenus. Le cinquième chapitre propose une discussion générale sur les apports de la méthode, les limitations identifiées et les conditions de reproductibilité des expériences. Enfin, le sixième chapitre conclut ce mémoire en synthétisant les contributions majeures et propose des pistes de recherche futures pour renforcer la robustesse, l'adaptabilité et le déploiement opérationnel des solutions développées.

CHAPITRE 2 REVUE DE LITTÉRATURE

Dans ce chapitre, nous proposons une revue détaillée des travaux existants sur la localisation et l'analyse des signaux radar FMCW, en nous appuyant sur les différentes représentations spatiales et temporelles offertes par les cartes Range–Doppler, Range–Azimuth et Range–Angle. Après avoir rappelé les concepts de base et les principes de traitement spatio-temporel (section 2.1), nous examinerons d'abord les méthodes classiques et les évolutions récentes fondées sur l'apprentissage profond pour la localisation à partir des cartes Range–Doppler (section 2.2). Nous aborderons ensuite les techniques de formation de faisceau et les architectures hybrides ou de bout en bout (end-to-end) exploitant les cartes Range–Azimuth (section 2.3), avant de présenter les approches les plus innovantes qui traitent directement les données brutes (IF ou ADC) sans prétraitement intermédiaire (section 2.4). Cette revue de littérature mettra en lumière les atouts et les limites de chaque famille de méthodes, et préparera le terrain pour la démarche intégrée que nous détaillerons au chapitre suivant.

2.1 Définitions et concepts de base

2.1.1 Principe de base du radar FMCW

Le radar à onde continue à modulation de fréquence (FMCW - *Frequency-Modulated Continuous Wave*) est largement utilisé pour l'estimation précise de la distance, de la vitesse et de l'angle des cibles. Son principe repose sur l'émission continue d'un signal électromagnétique dont la fréquence augmente linéairement dans le temps, selon une modulation appelée *chirp*.

Le processus global d'acquisition et de traitement du radar FMCW, regroupant l'ensemble de ces mesures et leur traitement numérique, est synthétisé dans la Figure 2.1. Celle-ci illustre les différentes étapes depuis l'émission du signal jusqu'à l'obtention des paramètres clés des cibles (distance, vitesse, angle).

Le fonctionnement du radar FMCW commence par l'émission d'un balayage de fréquence linéaire (signal à fréquence modulée linéairement, chirp en anglais) représenté mathématiquement par le signal $s_t(t)$:

$$s_t(t) = A_t \cos \left(2\pi \left(f_0 t + \frac{K}{2} t^2 \right) \right) \quad (2.1)$$

où A_t désigne l'amplitude du signal émis, f_0 est la fréquence initiale du chirp et K la pente

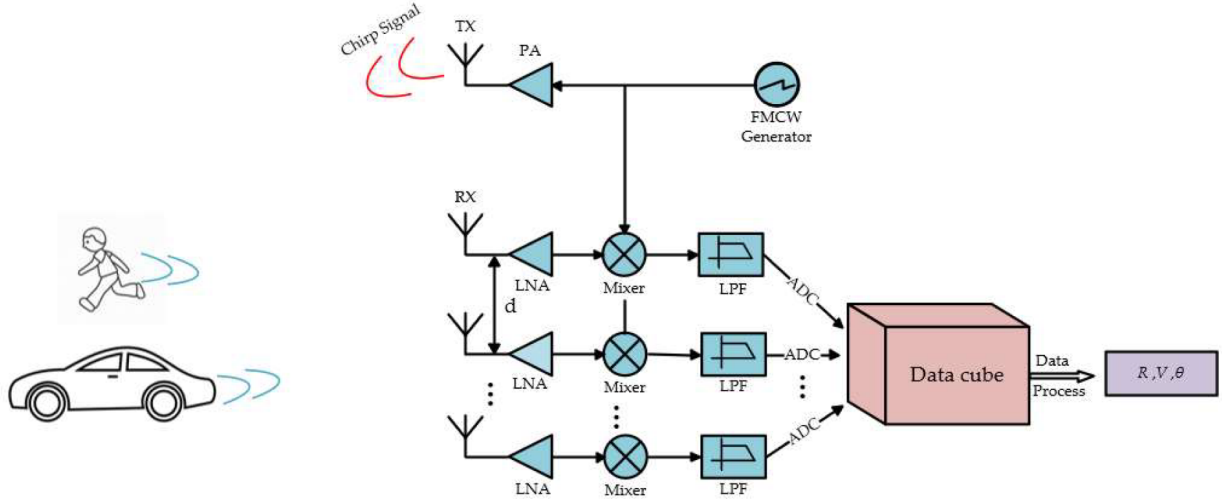


FIGURE 2.1 Schéma de principe du fonctionnement d'un radar FMCW. Le signal chirp est généré et émis par l'antenne TX, se propage dans l'environnement, puis est réfléchi par les cibles (véhicule, piéton). Le signal réfléchi est capté par les antennes RX, amplifié (LNA), mélangé avec le signal émis (Mixer), filtré (LPF), puis converti en numérique (ADC). Les données sont ensuite traitées dans un cube de données pour extraire les informations de distance (R), vitesse (V) et angle (θ). Figure utilisée avec permission depuis [2].

Copyright © Renesas Electronics.

de modulation (taux de variation fréquentielle en Hz/s). Cette modulation linéaire permet au radar FMCW de distinguer précisément les cibles situées à des distances variées, en associant chaque distance à une fréquence particulière du signal réfléchi.

Lorsqu'une cible réfléchit ce signal, il revient vers le radar après un délai temporel τ , proportionnel à la distance R séparant la cible du radar. Ce signal réfléchi peut alors être décrit par :

$$s_r(t) = A_r \cos \left(2\pi \left(f_0(t - \tau) + \frac{K}{2}(t - \tau)^2 \right) \right) \quad (2.2)$$

où A_r est l'amplitude du signal reçu. Cette équation introduit explicitement le délai temporel τ , essentiel à la mesure précise de la distance à la cible.

Pour extraire l'information de distance, le radar FMCW mélange le signal émis avec le signal réfléchi. La Figure 2.2 illustre le signal résultant de ce mélange qui présente une fréquence appelée fréquence de battement (f_b), directement proportionnelle au délai τ , et donc à la distance R :

$$f_b = K\tau = \frac{2KR}{c} \quad (2.3)$$

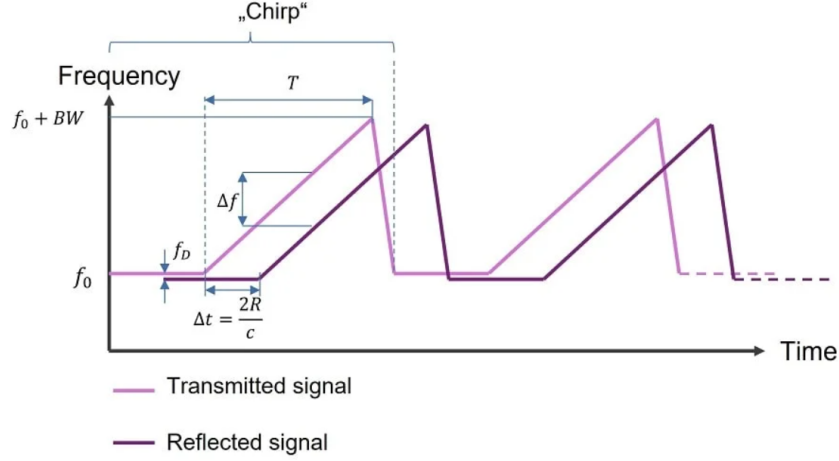


FIGURE 2.2 Illustration du principe d'un radar FMCW montrant l'évolution temporelle de la fréquence du signal émis (Transmitted signal) et du signal réfléchi (Reflected signal). Le décalage temporel Δt est proportionnel à la distance R entre le radar et la cible, tandis que le décalage fréquentiel Δf permet l'estimation de la vitesse relative. Source : Renesas Electronics

Figure utilisée avec permission. Copyright © Renesas Electronics.

Cette relation permet une estimation directe de la distance en mesurant simplement la fréquence du signal résultant après mélange. À partir de la fréquence de battement, la distance R s'exprime comme :

$$R = \frac{cf_b}{2K}. \quad (2.4)$$

La résolution en distance du radar FMCW, notée ΔR , dépend directement de la bande passante B du chirp. Plus la bande passante est élevée, meilleure est la résolution en distance, ce qui permet de distinguer précisément des cibles très proches les unes des autres :

$$\Delta R = \frac{c}{2B}. \quad (2.5)$$

L'importance de cette équation réside dans le choix judicieux de la bande passante du radar, condition essentielle à l'amélioration de la finesse de détection et de localisation.

En outre, lorsque la cible est en mouvement relatif par rapport au radar, le signal réfléchi subit un décalage fréquentiel supplémentaire, appelé effet Doppler. Ce décalage, noté $f_{Doppler}$,

permet au radar FMCW de mesurer directement la vitesse relative v de la cible par rapport au radar :

$$v = \frac{f_D \cdot c}{2f_0} \quad (2.6)$$

Cette équation est particulièrement utile pour les applications où la mesure de vitesse est nécessaire, telles que le suivi d'objets en mouvement rapide, les applications automobiles ou aéronautiques.

Enfin, les radars FMCW modernes intègrent souvent plusieurs antennes en configuration MIMO afin d'estimer l'angle d'arrivée du signal réfléchi. L'estimation angulaire repose sur la mesure de la différence de phase $\Delta\phi$ du signal entre deux antennes espacées d'une distance d :

$$\Delta\phi = \frac{2\pi d \sin(\theta)}{\lambda} \quad (2.7)$$

À partir de cette différence de phase, il est possible d'extraire l'angle d'arrivée θ selon :

$$\theta = \arcsin\left(\frac{\lambda\Delta\phi}{2\pi d}\right) \quad (2.8)$$

Cette mesure angulaire précise est indispensable pour les applications de cartographie, localisation et suivi de cibles.

2.1.2 Représentation I/Q et exploitation des signaux FMCW

Dans un radar FMCW, le signal reçu est converti en deux voies base-bande déphasées de 90° : la composante I (In-phase) alignée sur l'oscillateur local et la composante Q (Quadrature) décalée d'un quart de période. Ces deux signaux, numérisés par des convertisseurs analogique-numérique, forment ensemble le signal complexe $s(t) = I(t) + jQ(t)$, assurant la conservation intégrale de l'information d'amplitude et de phase. Cette représentation I/Q est le fondement des traitements radar : la distance est extraite de la fréquence de battement entre l'onde émise et le retour, les signatures micro-Doppler sont isolées grâce à l'analyse conjointe de la magnitude $\sqrt{I^2 + Q^2}$ et de la phase, et la localisation angulaire s'appuie sur la comparaison de ces composantes sur plusieurs antennes (voir Figure 2.3)

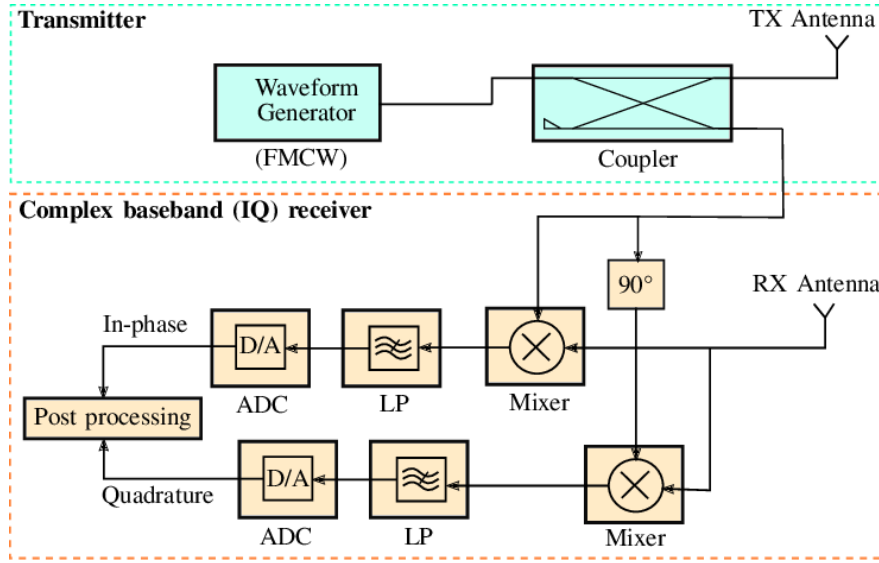


FIGURE 2.3 Représentation des composantes I/Q et schéma de leur exploitation dans un radar FMCW.

Figure utilisée avec permission. © 2017 Texas Instruments extrait de « Using a complex-baseband architecture in FMCW radar systems »

2.1.3 Bases méthodologiques du traitement spatio-temporel dans les radars FMCW

La transformation mathématique du signal reçu constitue une étape clé du traitement numérique dans un radar FMCW. Après la génération d'un signal de battement obtenu par mélange entre le chirp émis et le chirp réfléchi, une première Transformée de Fourier rapide (FFT) est appliquée sur ce signal mixé afin de convertir l'information temporelle en une représentation fréquentielle (voir Figure 2.4). Cette étape permet d'extraire des pics distincts correspondant directement aux distances des différentes cibles, en associant chaque fréquence de battement à une distance précise.

Ensuite, lorsque le radar est équipé d'un réseau d'antennes spatiales, une seconde FFT est effectuée spatialement sur les signaux issus de ces multiples antennes (illustrée dans la Figure 2.5). Cette transformée exploite l'évolution spatiale de la phase du signal, suivant une sinusoïde dont l'expression générale est donnée par $\sin(2\pi f \cdot n\Delta\tau)$, où f représente la fréquence spatiale dépendant directement de l'angle d'incidence, et $\Delta\tau$ est l'espacement entre les antennes. Cette analyse spatiale permet de calculer l'angle d'arrivée du signal réfléchi pour la localisation des cibles dans l'espace.

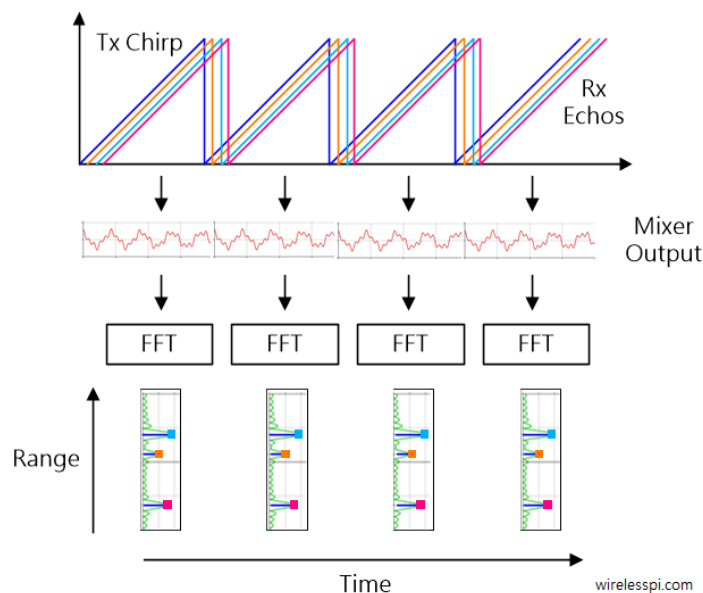


FIGURE 2.4 Illustration du principe mathématique de base du radar FMCW montrant la relation entre le signal chirp émis, l'écho réfléchi et la fréquence de battement, proportionnelle à la distance cible. Source : wirelesspi.com.

Figure utilisée avec autorisation de Wireless Pi (<https://wirelesspi.com>). © Wireless Pi.

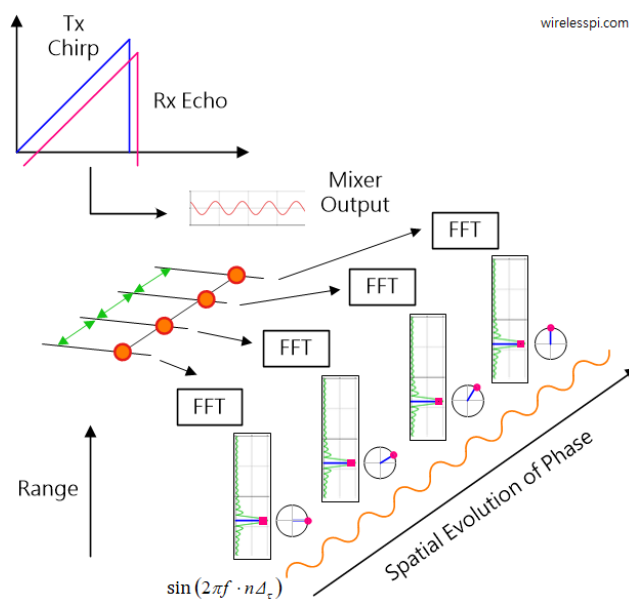


FIGURE 2.5 Schéma mathématique montrant l'application de la FFT spatiale sur les signaux issus de multiples antennes pour déterminer précisément l'angle d'arrivée grâce à l'évolution spatiale de la phase. Source : wirelesspi.com.

Figure utilisée avec autorisation de Wireless Pi (<https://wirelesspi.com>). © Wireless Pi.

Enfin, pour obtenir une estimation de la vitesse des objets détectés, le radar réalise une série successive des signaux à fréquence modulée linéairement (chirps en anglais) formant un cadre temporel défini (Frame Duration, $T_f = NT_c$, voir Figure 2.6). Une troisième FFT, cette fois effectuée sur l'évolution temporelle des fréquences de battement, extrait les variations fréquentielles dues au phénomène Doppler. Cette opération mathématique permet d'isoler les décalages de fréquence induits par les mouvements relatifs des cibles, fournissant ainsi une mesure de leur vitesse radiale. En combinant ces trois niveaux de transformation par FFT (portée, angle et vitesse), le radar FMCW permet une caractérisation complète et précise de la dynamique spatio-temporelle des cibles dans son environnement immédiat.

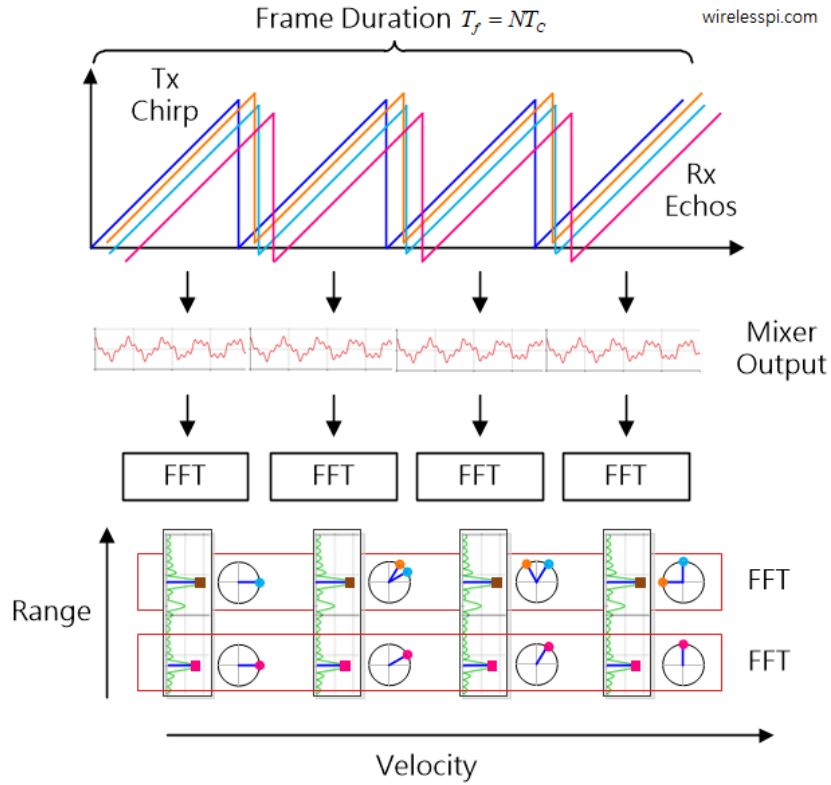


FIGURE 2.6 Illustration de la mesure de vitesse par FFT Doppler sur une série temporelle de chirps FMCW, permettant l'extraction des variations fréquentielles induites par le déplacement relatif des cibles. Source : wirelesspi.com.

Figure utilisée avec autorisation de Wireless Pi (<https://wirelesspi.com>). © Wireless Pi.

2.1.4 Définition et présentation des cartes Range-Doppler, Range-Azimuth et Range-Angle

Dans la figure 2.11, le traitement part d'un cube de données radar à trois dimensions (temps rapide, temps lent, canaux d'antenne) : une FFT le long de l'axe « temps rapide » génère la

carte de distance, une FFT le long de l'axe « temps lent » produit la carte de vitesse, et une formation de faisceaux spatiale adaptative (beamforming) sur les canaux d'antenne permet d'estimer la direction d'arrivée pour obtenir la carte d'angle. La méthode Capon (MVDR, Minimum Variance Distortionless Response) [13] minimise la puissance des interférences provenant de directions non ciblées tout en assurant un gain unitaire vers la cible, tandis que MUSIC (Multiple Signal Classification) [14] décompose la matrice de covariance en sous-espaces signal et bruit pour localiser les directions d'arrivée avec une résolution angulaire très élevée.

Donc, un premier passage en Transformée de Fourier (1D, puis 2D) permet de former la *Range-Doppler Map* (RDM), c'est-à-dire une représentation où chaque cellule indique la combinaison distance-vitesse d'une cible potentielle. Ensuite, des techniques de formation de faisceaux (par exemple Capon) appliquées aux canaux d'antenne servent à obtenir des cartes *Range-Azimuth* (RAM) ou *Range-Elevation* (REM), chacune traduisant l'intensité des échos en fonction de la distance et de l'angle considéré (azimutal ou en élévation). Une étape de détection CFAR (Constant False Alarm Rate) vient ensuite isoler les cibles significatives dans l'espace temps-fréquence, avant de concaténer ces cartes au fil des trames pour extraire des signatures globales ou des trajectoires dans la dimension temporelle.

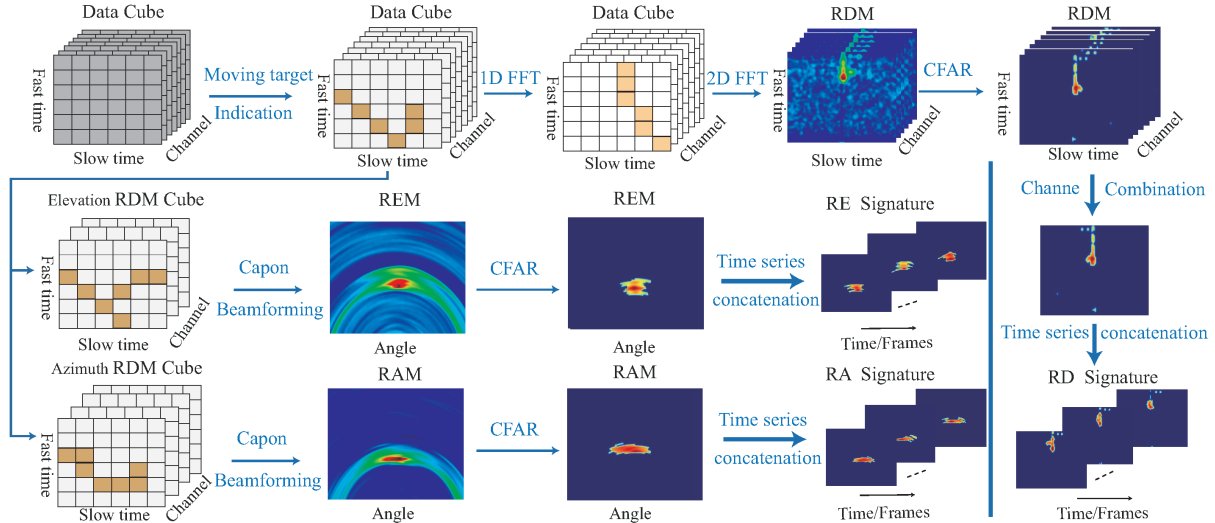


FIGURE 2.7 Obtention des cartes Range-Doppler (RDM), Range-Azimuth (RAM) et Range-Elevation (REM) à partir d'un cube de données radar mmWave FMCW. Illustration tirée de l'article *Multi-Feature Transformer-Based Learning for Continuous Human Motion Recognition with High Similarity Using mmWave FMCW Radar* [3], publié dans *Sensors* (2022), vol. 22, n°21, p. 8409 (doi:10.3390/s22218409). Figure utilisée avec permission. Sous licence Creative Commons Attribution (CC BY 4.0). © 2022 MDPI.

Range-Doppler Map (RDM). La *Range-Doppler Map* est obtenue en appliquant une première FFT sur l’axe du temps rapide pour extraire la distance (via la fréquence de battement) et une seconde FFT sur l’axe du temps lent pour mettre en évidence l’effet Doppler lié à la vitesse des cibles. Le résultat est une grille deux dimensions où chaque point représente l’intensité d’un écho identifié par sa distance radiale et sa vitesse relative, permettant de distinguer des cibles immobiles de celles en mouvement.

Range-Azimuth or Angle Map/ Range-Evaluation Map (RAM/REM). Une fois les informations de distance et de vitesse déterminées, le traitement s’étend à la dimension spatiale pour estimer l’angle de provenance du signal réfléchi. Grâce aux canaux d’antenne (ou à un réseau MIMO), on applique des algorithmes de formation de faisceaux ou de beam-forming suivis d’une FFT spatiale pour obtenir une carte *Range-Azimuth* ou *Range-Elevation*, selon l’orientation de l’antenne. Chaque cellule dans ce plan représente l’intensité du signal en fonction de la distance et de l’angle, ce qui facilite la localisation de la cible dans le plan horizontal ou vertical, ainsi que sa classification éventuelle.

2.2 Localisation des cibles sur la carte Range-Doppler : Approches classiques et méthodes basées sur l’apprentissage profond

2.2.1 Méthodes classiques et mathématiques

Transformée de Fourier, Traitement Doppler

Saponara et Neri [15] décrivent une plate-forme FMCW sur FPGA qui s’appuie entièrement sur des transformées de Fourier multidimensionnelles pour extraire, en temps réel, les cartes Range-Doppler nécessaires à la localisation de cibles ; localisation de piétons pour une application en transport. La chaîne de traitement (voir Figure 2.8) commence par une FFT 1D “fast-time” sur chaque rampe de chirp pour convertir le décalage de fréquence en distance (range), puis par une FFT 1D “slow-time” sur les amplitudes extraites au case (bin en anglais) de portée pour isoler la composante Doppler (vitesse). L’assemblage de ces deux étapes génère une carte Range-Doppler 2D dans laquelle la position radiale et la vitesse des cibles apparaissent sous forme de pics localisés ; la simple détection de l’indice du maximum de magnitude dans cette carte suffit alors à estimer la distance et la vitesse des humains sans recourir à des techniques de seuillage adaptatif.

La méthode proposée par Kim et al. (2021) [16] se fonde sur une première estimation de distance issue d’une FFT unidimensionnelle du signal battement, où l’indice du pic de magnitude fournit une localisation grossière, puis sur une étape d’interpolation locale qui affine cette

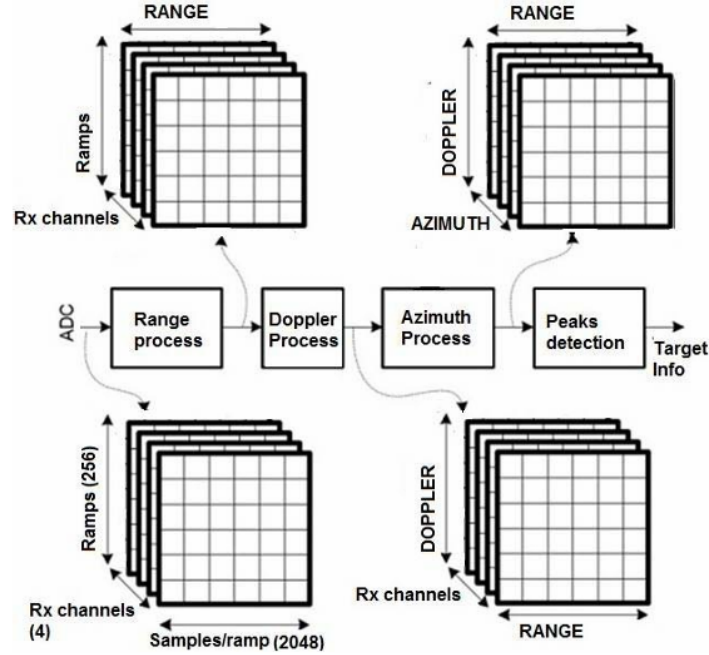


FIGURE 2.8 Chaîne de traitement FFT multidimensionnelle pour l'extraction de la portée, du Doppler et de l'azimut (adapté de Saponara & Neri, 2017).

Figure utilisée avec permission. © 2017 IEEE

estimation jusqu'à une résolution super-résolutive sans recourir à un seuillage CFAR. Dans un travail antérieur, Choi, Jang et Roh (2015) [17] ont montré qu'un radar altimètre FMCW (à onde continue modulée en fréquence) peut atteindre une très faible erreur de mesure sur de larges distances en détectant le pic FFT, puis en appliquant une interpolation parabolique autour de ce pic. De leur côté, Scherr et al. (2015) [18] ont combiné la localisation du pic FFT avec un ajustement de phase inspiré de la borne de Cramér-Rao, obtenant ainsi une précision de distance et de vitesse notablement accrue tout en conservant une implémentation temps réel.

Détection CFAR par seuillage : une approche classique de localisation Range-Doppler

En traitement des signaux radar, les détecteurs à taux constant de fausses alarmes **CFAR** [19] sont couramment utilisés pour identifier des cibles potentielles en ajustant dynamiquement les seuils de détection en fonction du bruit environnant. Ces détecteurs analysent les données via une fenêtre glissante, applicable à des données de toute dimension, composée de trois éléments illustrés à la figure 2.9 : la cellule testée (CUT) au centre, des cellules de garde

formant une zone tampon autour de la CUT pour éviter l'interférence des pics de signal, et des cellules de référence mesurant les conditions locales de bruit.

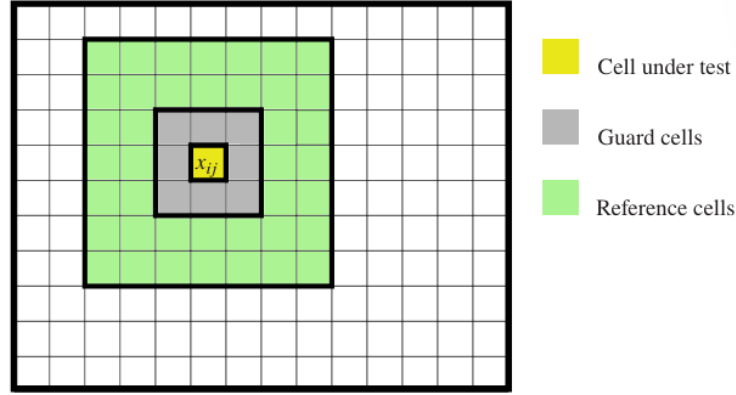


FIGURE 2.9 Visualisation de la cellule test (CUT), de la fenêtre de garde et de la fenêtre de référence dans les détecteurs CFAR.

Figure utilisée avec permission. © 1960 S.-S. Lee extrait de « Digital Constant False Alarm Rate Radar Detection in Clutter », IRE Transactions on Information Theory.

De nombreuses variantes de CFAR ont été proposées qui utilisent des détecteurs CFAR modulables pour ajuster dynamiquement leurs seuils de détection en fonction du bruit ambiant, plutôt que de s'appuyer sur des valeurs fixes. La version de base, le **CA-CFAR** [20], calcule ce seuil en faisant la moyenne du bruit mesuré dans les cellules voisines de la cible analysée. Pour répondre à des situations plus complexes, d'autres variantes ont été développées : le **GO-CFAR** [21] utilise la valeur de bruit la plus élevée détectée autour de la cible, ce qui le rend efficace près de gros obstacles, tandis que le **OS-CFAR** [22] privilégie la valeur la plus faible, idéal pour des environnements déjà saturés en interférences. Le **VI-CFAR** [23] combine ces deux approches en alternant entre elles selon le contexte, et le **CM-CFAR** [24] innove en s'adaptant en temps réel via une carte mémorisant l'évolution du bruit.

Après avoir défini de manière générale les variantes CA-CFAR et OS-CFAR, Wang et al. [25] appliquent ces détecteurs directement sur les cartes Range-Doppler $I_{RD}(r, \dot{r})$, comme illustré dans la Figure 2.10. Après l'acquisition des données brutes, chaque carte range-doppler est d'abord soumise au CFAR pour produire une image binaire $I_{RD}^d(r, \dot{r})$ où les pixels à «1» marquent les détections. Plutôt que de se limiter à la simple extraction des pics distance/vitesse, l'approche consiste à associer à chaque point détecté son angle via une estimation DOA en azimuth. Avant de lancer l'algorithme de groupement pour séparer les cibles, ils convertissent ces coordonnées polaires (r, \dot{r}, θ) en coordonnées cartésiennes (x, y, z) afin de pouvoir appliquer un critère de distance uniforme. Les nuages de points ainsi obtenus

sont enfin regroupés pour suivre individuellement chaque personne, même en situation de «groupes» où deux sujets évoluent côte à côte ou l'un derrière l'autre, cas où la résolution en distance et en Doppler seule serait insuffisante pour discriminer tous les individus de manière fiable.

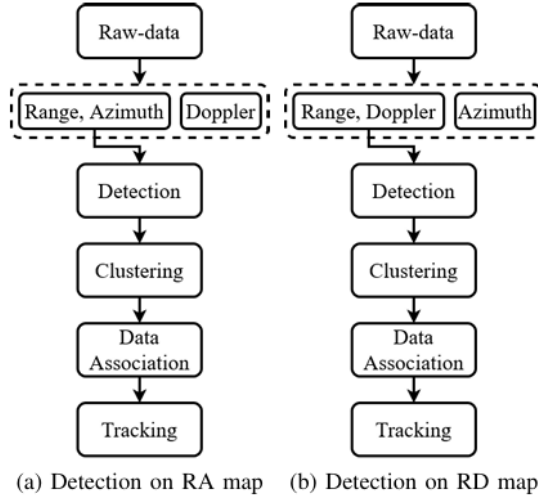


FIGURE 2.10 Chaînes de traitement pour la localisation avec un radar FMCW : (a) détection, clustering, association de données et suivi sur carte distance–azimut ; (b) détection, clustering, association de données et suivi sur carte distance–Doppler.

Figure utilisée avec permission. © 2024 IEEE

Dans la même lignée, Jiang et al. (2023) [26] proposent une méthode robuste de suivi de cibles en environnement intérieur dense (voir Figure 2.11), en s'appuyant également sur les cartes Range-Doppler issues d'un radar FMCW MIMO à 77 GHz. Après génération de la carte RD, un filtre ICM (Indicateur de cibles mobiles) est d'abord appliqué pour supprimer les cibles statiques. Un OS-CFAR bidimensionnel identifie ensuite les zones d'énergie significative correspondant à des cibles mobiles. Chaque détection est associée à une estimation angulaire par un formation de faisceau adaptatif à réponse sans distorsion et à variance minimale (MVDR) qui calcule des poids pour minimiser la puissance résiduelle (variance) tout en maintenant un gain unitaire dans la direction de la cible, pour former un nuage de points radar $(r, \dot{r}, \theta, \phi, \text{SNR})$. L'ensemble est ensuite traité par un filtre de Kalman étendu, suivi d'un algorithme de regroupement spatial permettant le suivi multi-cible. Les auteurs montrent, à l'aide de représentations visuelles, que l'utilisation systématique des cartes Range-Doppler dès les premières étapes de traitement permet d'assurer une localisation robuste, même en présence de signatures croisées ou fortement bruitées.

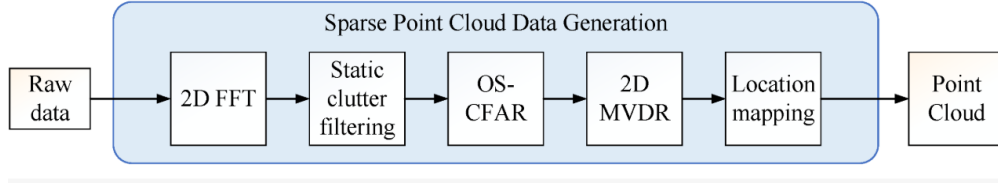


FIGURE 2.11 (a) Diagramme du workflow de prétraitement du signal montrant la FFT 2D, le filtrage des échos statiques, l'OS-CFAR, le beamforming MVDR et le mapping de position pour générer un nuage de points.

Figure utilisée avec permission. © 2022 IEEE

Dans le cadre de notre analyse des artefacts en détection radar, plusieurs travaux récents ont proposé des adaptations du CFAR (Taux Constant de Fausses Alarmes) ou des méthodes complémentaires visant à améliorer la robustesse face aux interférences, aux échos parasites non gaussien, ou encore à la présence de cibles parasites. Zheng (2023) [27] propose une méthode de détection utilisant un prétraitement par régions d'intérêt (ROI) afin de réduire significativement la complexité computationnelle, suivie par l'application successive de trois étapes de CFAR unidimensionnel (1D CA-CFAR). Cette approche améliore la probabilité de détection tout en réduisant les ressources nécessaires par rapport aux méthodes classiques bidimensionnelles. Tran et al. (2024) [28] se concentrent sur la détection et la classification de drones avec un radar FMCW haute résolution, traitant spécifiquement le problème du clutter discret et du bruit de phase parasite par modélisation avancée et suppression morphologique, complétée par une détection CFAR 2D finale. Xiao et al. (2024) [29] présentent une méthode adaptative particulièrement efficace dans des environnements de clutter non gaussien, utilisant des critères d'estimation alternée tels que le test de vraisemblance généralisé et divers autres critères statistiques pour améliorer la robustesse de la détection.

Enfin, Kim et Lee (2018) [28] explorent l'utilisation du méthode de regroupement (clustering) DBSCAN pour regrouper efficacement les cibles détectées par OS-CFAR dans les systèmes radar automobiles, permettant ainsi de réduire les fausses alarmes et d'améliorer la précision globale de détection.

2.2.2 Exploitation des séquences range–Doppler par apprentissage profond

Face aux limites des approches traditionnelles basées sur des heuristiques ou des traitements manuels (FFT, seuillage CFAR, etc.), plusieurs travaux récents ont exploré l'utilisation de l'apprentissage profond pour tirer pleinement parti des cartes Range–Doppler (RDM) dans des tâches de localisation et même de reconnaissance d'activités humaines.

Decourt et al. [30] ont exploité les méthodes d'apprentissage dans le modèle **DAROD**, un détecteur d'objets radar inspiré de Faster R-CNN. En s'appuyant uniquement sur les RDM (résolution typique : 256×64), leur méthode contourne les coûts computationnels élevés liés à la dimension angulaire des RAD. Comme illustré dans la Figure 2.12, leur approche interrompt le pipeline classique avant la transformation angulaire, se concentrant uniquement sur les spectres Range-Doppler. L'architecture est constituée d'un réseau convolutionnel léger (7 couches) et d'ancres spécialement conçues pour les échelles radar. Cette implémentation atteint des temps d'inférence de 25 ms/image tout en surpassant significativement les méthodes radar traditionnelles.

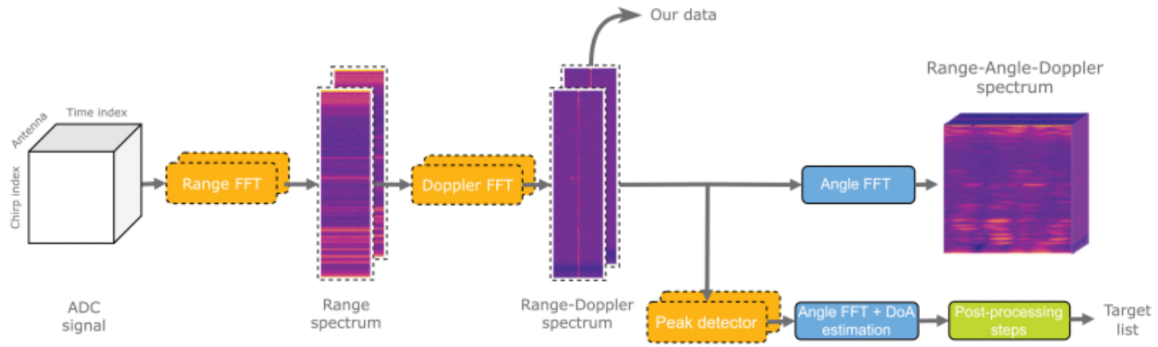


FIGURE 2.12 Pipeline de traitement radar pour la génération des cartes Range-Doppler (RDM) et des tenseurs Range-Angle-Doppler (RAD). La méthode DAROD utilise les données à la sortie du spectre Range-Doppler, réduisant ainsi la complexité computationnelle en omettant l'estimation angulaire.

Figure utilisée avec permission. © 2022 IEEE

Dans une approche complémentaire, Ding et al. [31] proposent **ST-ConvLSTM**, un réseau de type Convolutional LSTM (voir Figure 2.13) conçu pour extraire simultanément les dynamiques spatiales et temporelles des séquences RDM. Les cartes Range-Doppler dynamiques (DRDF), générées par transformée de Fourier 2D, sont traitées comme un volume spatio-temporel. L'architecture repose sur des cellules ConvLSTM, renforcées par des modules d'attention spatiale et temporelle, qui focalisent le traitement sur les zones et les instants les plus informatifs. Cette approche améliore la détection d'activités complexes dans des scènes bruitées ou avec occlusions partielles.

Poussant plus loin l'analyse des cartes RDM, l'approche hybride **HOOD** [32] adopte une séparation explicite entre les dynamiques macroscopiques (déplacement global de la cible) et les micro-mouvements (gestes fins capturés dans les spectres Doppler). Comme illustré

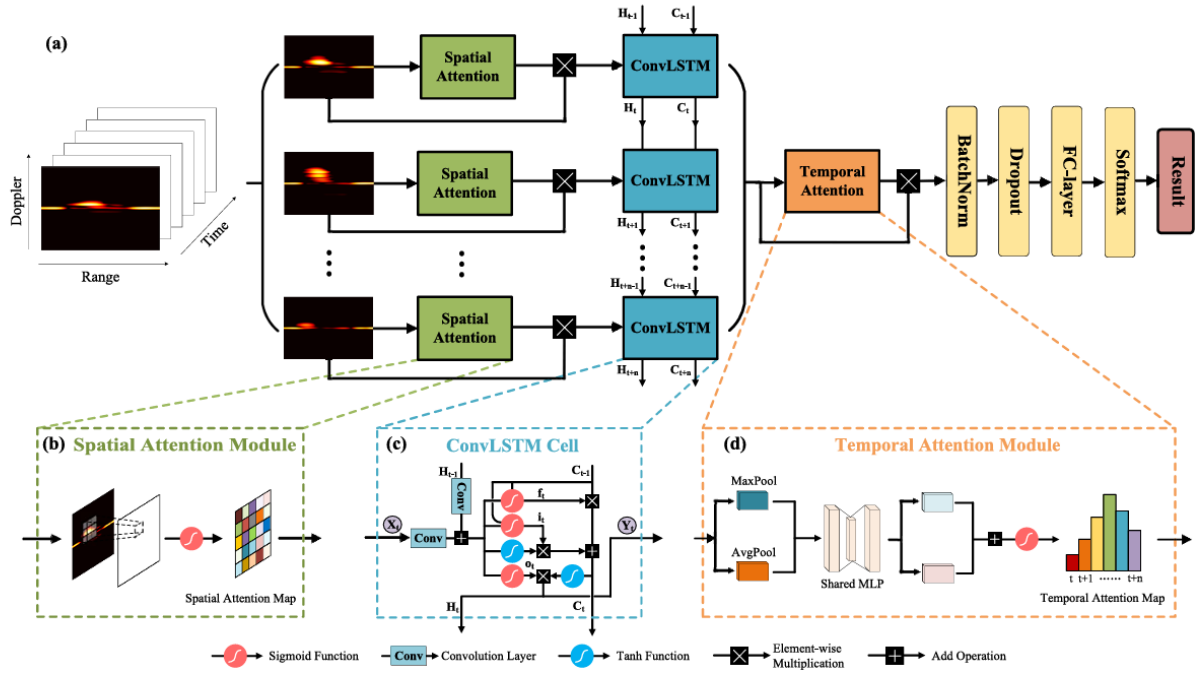


FIGURE 2.13 Architecture du réseau ST-ConvLSTM pour la reconnaissance des mouvements humains à partir des cartes Range-Doppler dynamiques. (a) Flux de traitement du ST-ConvLSTM. (b) Module d'attention spatiale pour l'amélioration des caractéristiques spatiales. (c) Structure de la cellule ConvLSTM combinant convolutions et mémoire à long terme. (d) Module d'attention temporelle pondérant les trames les plus informatives.

Figure utilisée avec permission. © 2022 IEEE

à la Figure 2.14, la méthode suit les trajectoires des pics de réflexion pour reconstruire la cinématique globale, tandis qu'elle extrait parallèlement la structure fine des spectres pour détecter les micro-gestes tels qu'un mouvement de bras ou un battement de main. Les deux flux sont fusionnés lors de la classification, ce qui permet une validation croisée entre niveau global et local, réduisant ainsi les fausses alarmes. Cette fusion bi-niveaux s'avère particulièrement efficace dans les environnements encombrés, où des mouvements ambigus peuvent survenir.

Référence	Algorithme / Architecture	Entrée principale	Contributions clés
Méthodes FFT directes			
Saponara et Neri (2017) [15]	Chaîne FFT 1+1 D sur FPGA	FMCW brut \rightarrow RDM	Pipeline temps réel; localisation piétons sans seuillage (pic max RD).
Kim <i>et al.</i> (2021) [16]	Pic FFT + interpolation locale	Signal battement 1-D	Super-résolution distance sans CFAR; précision métrique accrue.
Choi, Jang et Roh (2015) [17]	Pic FFT + interpolation parabolique	Altimètre FMCW	Erreur ± 0.01 m sur longues portées.
Scherr <i>et al.</i> (2015) [18]	Pic FFT + ajustement de phase (CRLB)	RDM	Gain portée / vitesse tout en restant temps réel.
Détection CFAR (Range-Doppler)			
Wang <i>et al.</i> (2024) [25]	CA/OS-CFAR + DOA + clustering	RDM $\rightarrow (r, \dot{r}, \theta)$	Conversion en (x, y, z) ; suivi multi-cible (personnes adjacentes).
Jiang <i>et al.</i> (2023) [26]	OS-CFAR + MVDR + EKF	RDM (77 GHz MIMO)	Filtre ICM \rightarrow suppression statiques; suivi indoor dense.
Zheng (2023) [27]	ROI + $3 \times 1D$ CA-CFAR	RDM	Complexité $\div 2$; P_D maintenu vs 2D-CFAR.
Tran <i>et al.</i> (2024) [28]	Suppression morphologique + 2D CFAR	RDM haute rés.	Détection / classification drones malgré clutter discret.
Xiao <i>et al.</i> (2024) [29]	CFAR adaptatif (GLRT, etc.)	RDM non gaussien	Seuils optimisés \rightarrow robustesse clutter fort.
Kim & Lee (2018) [28]	OS-CFAR + DBS-CAN	RDM automobile	Regroupement spatial; fausses alarmes $\downarrow 30\%$.
Apprentissage profond			
Decourt <i>et al.</i> (2022) [30]	DAROD (CNN 7c + RPN)	RDM 256×64	Inspiré Faster R-CNN; ancres radar; 25 ms/image.
Ding <i>et al.</i> (2022) [31]	ST-ConvLSTM + attentions	Séquence DRDF	Capture dynamiques spatio-temporelles; robuste bruit/occlusion.
Hou <i>et al.</i> (HOOD, 2024) [32]	HOOD (flux macro + micro)	RDM + trajectoires	Fusion bi-niveaux; fausses alarmes \downarrow en scènes encombrées.

TABLEAU 2.1 Tableau de synthèse des approches de localisation sur cartes Range-Doppler (Section 2.2).

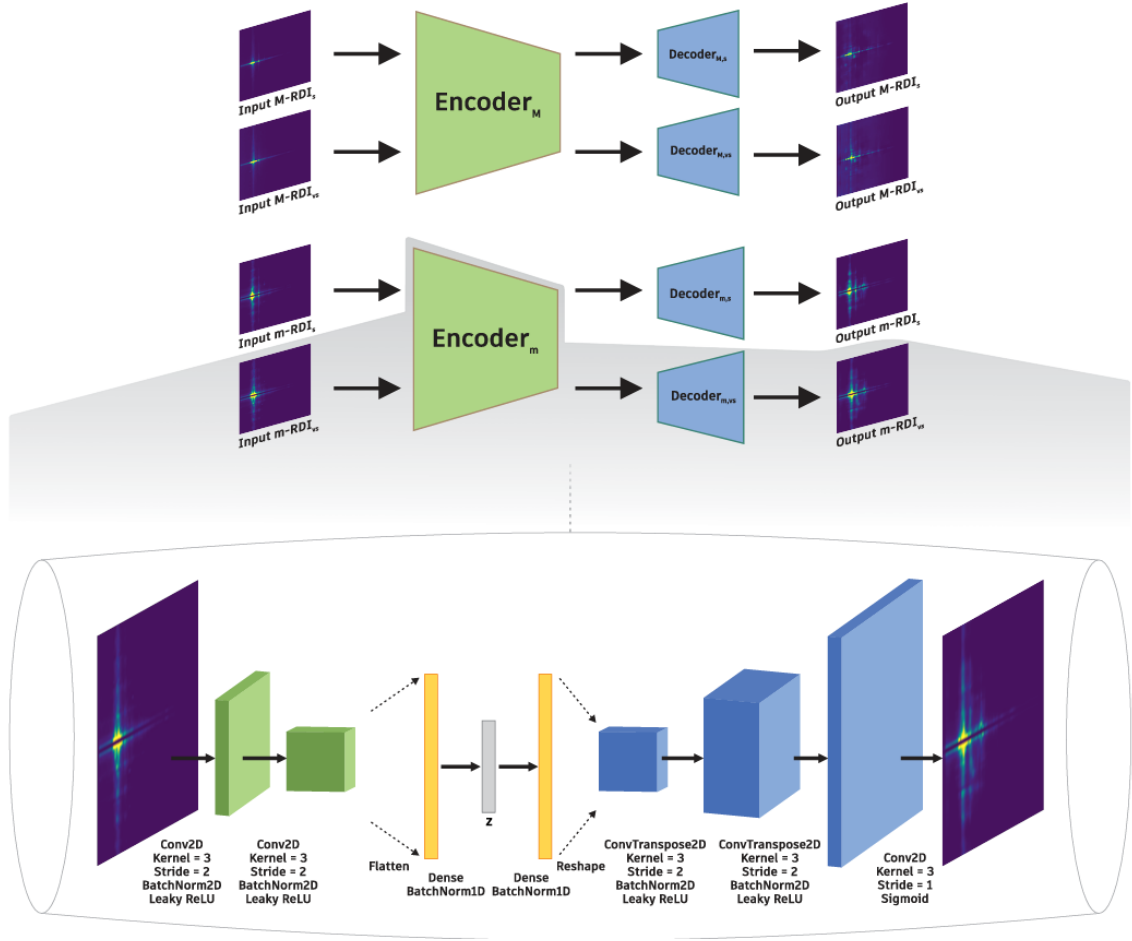


FIGURE 2.14 Architecture du modèle HOOD pour la reconnaissance d'activités à partir des cartes Range-Doppler. Le traitement est séparé en deux flux : un pour les mouvements globaux via le suivi des pics de réflexion, et un autre pour les micro-mouvements à partir de l'analyse fine des spectres Doppler. Les deux branches sont fusionnées lors de la classification.

Figure utilisée avec permission. © 2024 IEEE

2.3 Localisation des cibles sur la carte Range-Angle/Azimuth : Approches classiques et méthodes basées sur l'apprentissage profond

2.3.1 Méthodes de traitement du signal RAM

Les techniques d'analyse des cartes distance-angle (Range-Angle Map, RAM), représentations bidimensionnelles liant la distance et l'angle des cibles radar, reposent sur deux approches méthodologiques : les algorithmes de sur-résolution (super-resolution), conçus pour dépasser les limites physiques de résolution, et les méthodes de formation de faisceaux conventionnelle (conventional beamforming en anglais) et formation de faisceaux de Capon (Capon beamforming en anglais) pour optimiser la précision angulaire.

Algorithmes de Super-Résolution

Les algorithmes de super-résolution (SR) sont des techniques conçues pour estimer des paramètres physiques (distance, vitesse, angle, fréquence) ou reconstruire des signaux avec une précision supérieure aux limites théoriques des systèmes d'acquisition.

Les algorithmes de super-résolution, comme **MUSIC** [33], exploitent la structure mathématique des signaux radar pour séparer les composantes d'intérêt des artefacts. En décomposant les données en sous-espaces de signal et de bruit, ils identifient des fréquences Doppler plus subtiles, comme les harmoniques respiratoires ou cardiaques, même dans des environnements encombrés. Des adaptations spécifiques, comme le MUSIC 2D, étendent cette logique aux dimensions distance-vitesse, permettant une localisation précise des cibles [34]. Cet algorithme a été adapté pour séparer efficacement les harmoniques respiratoires et cardiaques, même dans des scénarios à faible rapport signal-bruit [35]. Parallèlement, des développements techniques exploitent l'inverse de la CZT (Chirp Z-Transform, transformée en Z chirpée), qui permet de calculer la transformée en Z d'un signal sur des contours de fréquence arbitraires pour un réglage plus fin de la résolution spectrale, afin de localiser des cibles biomédicales. Cette approche étend ainsi les applications du radar en télédétection pour la santé [36]. Enfin, des implémentations embarquées innovantes intègrent des filtres spectraux et des FFT 2D pour générer des RAM en temps réel pour des applications comme la détection de chutes [37].

ESPRIT [38] (Estimation of Signal Parameters via Rotational Invariance Techniques) est privilégié pour sa résolution élevée et son efficacité calculatoire dans les applications radar pour environnements intérieurs. Son approche matricielle exploite l'invariance rotationnelle des sous-espaces signaux, éliminant le besoin de recherche de pics spectraux comme avec MUSIC. En effet, MUSIC est une méthode de localisation de sources qui s'appuie sur la décomposition de la matrice de covariance des mesures d'un réseau de capteurs en deux sous-espaces : l'un contenant l'information utile (les signaux) et l'autre ne contenant que du bruit. En projetant un vecteur de direction hypothétique sur le sous-espace bruit, on obtient presque zéro pour la vraie direction d'une source et des valeurs plus élevées sinon. En balayant toutes les directions possibles et en calculant l'inverse de cette projection, on fait apparaître des pics très marqués aux angles correspondant aux sources réelles. Ainsi, MUSIC permet de repérer avec une grande précision les angles d'arrivée des signaux, même quand ils sont très proches les uns des autres.

Une alternative efficace à ESPRIT est l'algorithme **Root-MUSIC** [39], qui améliore l'estimation de l'angle en recherchant directement les racines d'un polynôme formé à partir de la matrice de covariance des signaux. Cette approche élimine la nécessité de rechercher des pics dans le spectre pseudo-spatial, ce qui réduit la complexité computationnelle tout en aug-

mentant la précision. Les simulations Monte-Carlo ont montré que Root-MUSIC surpasse ESPRIT dans des conditions de faible rapport signal/bruit (SNR),

Formation de faisceau conventionnelle (Beamforming), Méthode de Capon (Beamforming Capon) et les méthodes basées sur MVDR (Variance Minimale sans Distorsion de la Réponse)

Au sein des méthodes classiques de traitement des cartes range–azimut (ou range–angle), on retrouve deux approches phares de la formation de faisceau : Le **Formation de faisceau conventionnelle (CBF)** consiste à appliquer sur chaque capteur un décalage de phase permettant d’aligner les échos issus d’une direction donnée, puis à sommer les signaux ainsi pondérés [40]. Cette simple sommation renforce la contribution de la source visée tout en atténuant celles hors axe, mais sa résolution angulaire demeure limitée lorsque plusieurs cibles sont proches. Le **Formation de faisceau de Capon**, ou *Réponse Sans Distorsion à Variance Minimale (MVDR)*, calcule des poids qui minimisent la puissance totale des signaux hors direction cible tout en garantissant une réponse unitaire vers celle-ci [41]. Il offre une meilleure précision angulaire, au prix d’une complexité de calcul accrue.

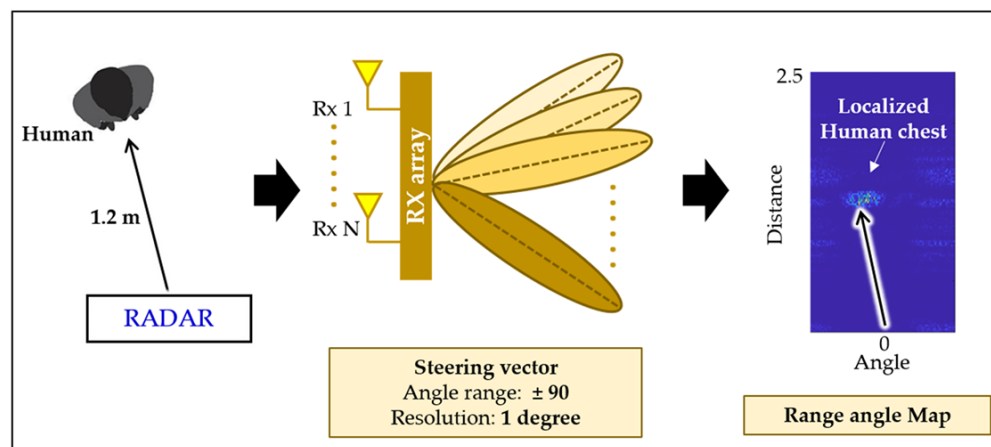


FIGURE 2.15 Illustration de la formation de faisceau pour la construction d’une carte range–angle : application d’un vecteur de pilotage (steering vector) sur une plage angulaire de $\pm 90^\circ$ pour localiser le torse humain.

Figure utilisée avec permission. © Texas Instruments

L’impact de ces techniques dans un contexte biomédical reste à valider. Ahmed et al. [42] ont montré qu’elles améliorent l’extraction de la fréquence cardiaque pour des angles d’incidence supérieurs à 20° , démontrant ainsi le potentiel du beamforming pour la détection de signes vitaux par radar FMCW.

2.3.2 Localisation par apprentissage profond et génération de cartes de confiance

Apprentissage Profond et méthodes hybrides pour l'exploitation des cartes RA

Parmi ces travaux, **DeepMUSIC** [43] s'appuie sur des *cartes range-angle* issues d'un radar FMCW. Après transformée de Fourier en distance puis en angle sur les signaux d'antenne, on calcule pour chaque faisceau de distance (beam en anglais) la matrice de covariance R_y , qui constitue l'entrée du DNN (voir Fig. 2.16). Le réseau génère alors un *spectre MUSIC* « profond » (spectres profonds), sur lequel une détection de pics identifie les directions d'arrivée estimées $\{\theta_k\}$. En tirant parti de la capacité d'apprentissage du DNN, DeepMUSIC améliore la résolution angulaire et réduit les artefacts des méthodes classiques, tout en restant tributaire du prétraitement de la carte range-angle.

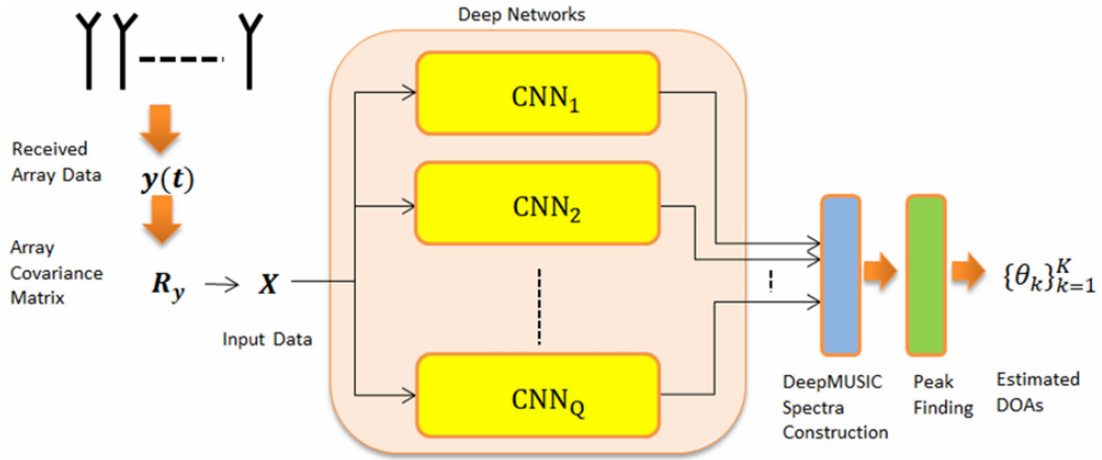


FIGURE 2.16 Architecture de DeepMUSIC : le réseau profond (Deep Networks) traite la matrice de covariance R_y pour produire un spectre MUSIC amélioré (Deep Spectra), suivi d'une étape de détection de pics (Peak Finding) pour extraire les angles DoA estimés.

Figure utilisée avec permission. © 2020 IEEE.

Les travaux de Meiresone et al. [44] visent à segmenter directement les cartes Range-Azimuth (RAM) issues d'un radar FMCW embarqué, afin d'estimer la distance au premier obstacle pour chaque angle d'arrivée. Pour ce faire, ils génèrent d'abord les RA maps par FFT en distance suivie d'un beamforming de type Capon, puis annotent manuellement, à l'aide de mesures LiDAR de référence, la distance réelle au premier obstacle pour chaque case (bin en anglais) azimuthal. Le réseau proposé (voir Fig. 2.17) est un CNN compact constitué de plusieurs blocs contractants (Conv2D, BatchNorm, LeakyReLU, MaxPool) qui réduisent progressivement la résolution spatiale tout en extrayant des caractéristiques pertinentes, puis

de deux couches fully connected (**fc1**, **fc2**) qui régressent un vecteur de distances normalisées.

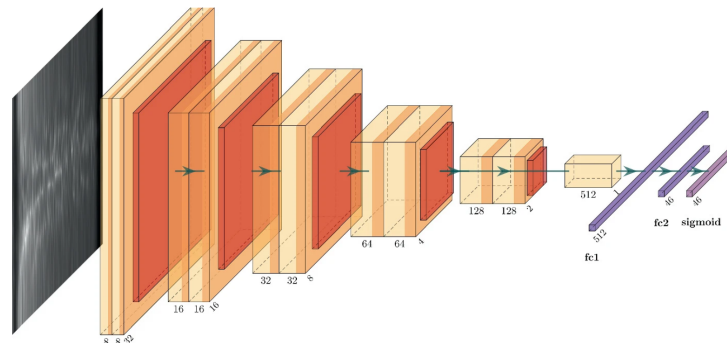


FIGURE 2.17 Architecture du CNN proposé par Meiresone et al. pour la segmentation des cartes Range–Azimuth : après beamforming de Capon, la carte RA est traitée par des blocs convolutifs contractants (Conv2D, BatchNorm, LeakyReLU, MaxPool), puis deux couches entièrement connectées (**fc1**, **fc2**) prédisent, pour chaque case (bin en anglais) azimuthal, la distance normalisée au premier obstacle.

Figure utilisée avec permission. © 2023 Springer.

Génération et exploitation des cartes de confiance pour la localisation

La détection par génération de cartes de confiance avec les modèles d'apprentissage profonds constitue une stratégie prometteuse pour la reconnaissance d'objets à partir de données radar. Ces cartes fournissent une représentation spatiale explicite indiquant la probabilité de présence d'un objet dans l'espace observé.

En utilisant cette approche, **RODNet** [45] exploite exclusivement les cartes portée–azimut (Range Azimuth Maps) issues du radar FMCW pour produire directement des ConfMaps via trois variantes spécialisées. Comme illustré à la Figure 2.18, un extrait de RA-Maps est d'abord injecté dans le modèle, qui génère ensuite des cartes de confiance représentant la position et la classe des cibles, avant qu'une étape de post-traitement n'en délivre les résultats finaux de détection.

Malgré leur puissance, les cartes de confiance générées par RODNet peuvent présenter des artefacts, notamment à courte distance, entraînant des faux positifs. Pour y remédier, Wen et al. [46] introduisent **GANomaly**, un modèle antagoniste qui associe un *réseau génératif* et *réseau discriminatif* (voir Figure 2.19). Le réseau génératif encode d'abord la carte de confiance d'entrée en un vecteur latent, reconstruit l'image via un décodeur puis réencode la reconstruction pour contraindre les deux vecteurs latents à se rapprocher (latent vector loss) tout en préservant la fidélité de l'image (image loss). Le réseau discriminatif, quant à

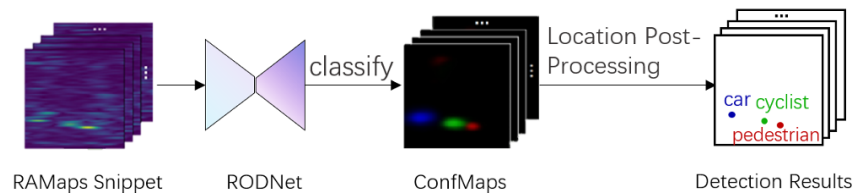


FIGURE 2.18 Pipeline de RODNet : un extrait de RA-Maps est traité par le réseau RODNet pour produire des cartes de confiance (ConfMaps), qui sont ensuite post-traitées pour générer les résultats finaux de détection.

Figure utilisée avec permission. © 2021 IEEE.

lui, extrait des caractéristiques de la carte reconstruite via un encodeur dédié et les transmet à un classifieur qui distingue les cartes réelles des reconstructions factices, renforçant ainsi la détection des anomalies par la discrimination loss. Grâce à cette architecture, GANomaly identifie et corrige automatiquement les zones peu fiables dans les ConfMaps, réduisant significativement le risque de fausses détections.

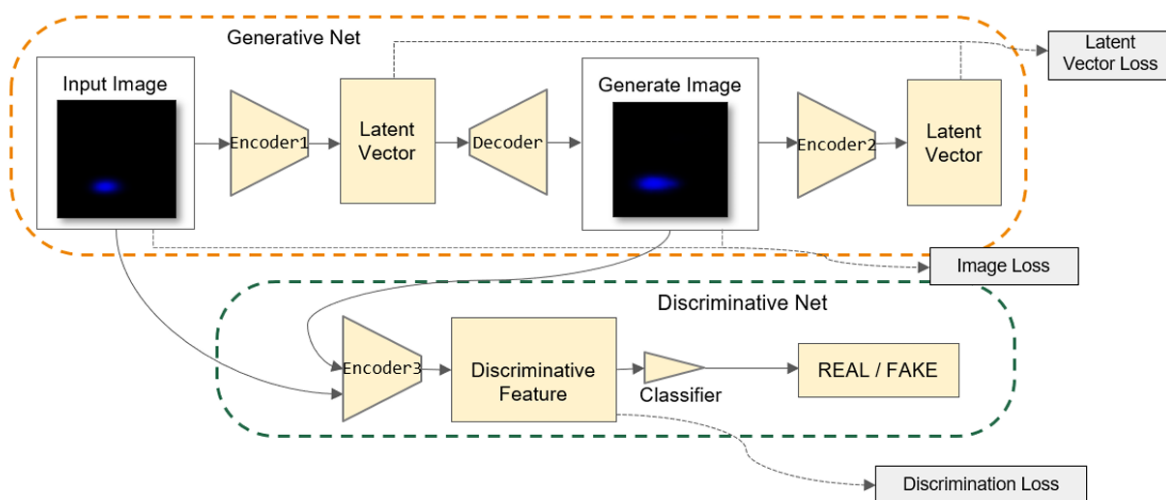


FIGURE 2.19 Architecture de GANomaly pour la détection et la correction d'anomalies dans les ConfMaps. Le Generative Net (orange) assure la reconstruction et le transfert latent, tandis que le Discriminative Net (vert) différencie les cartes réelles des reconstructions

Figure utilisée avec permission. © 2023 IEEE.

Après les avancées de RODNet en détection d'objets radar et les approches GANomaly pour la repérage d'anomalies, TransRadar [47] se positionne comme une solution de segmentation sémantique multi-vues en temps réel, tirant parti d'un réseau auto-attentif adaptatif-directionnel pour exploiter directement les données FMCW. Illustré par la figure 2.20), Trans-

Radar [47] est une méthode de segmentation sémantique de scènes radar basée sur un réseau auto-attentif spécialement adapté aux données multi-vues de radars FMCW. En sortie, le modèle génère pour chaque vue (Range–Doppler, Range–Angle, Angle–Doppler) non seulement une carte de segmentation pixel-à-pixel indiquant la présence et la classe des objets (par exemple « humain » vs « fond »), assurant ainsi leur localisation en distance et en angle, mais aussi une carte de confiance associée à chaque prédiction, fournissant pour chaque pixel le degré de certitude de son affectation à une classe. Face au bruit, à la forte parcimonie et au déséquilibre entre avant-plan et arrière-plan, les auteurs proposent une architecture à deux étages : un encodeur parallèle qui traite plusieurs trames de ces trois représentations, puis un bloc d’attention adaptative-directionnelle, capable d’échantillonner dynamiquement lignes et colonnes pour ne retenir que les informations les plus pertinentes tout en limitant la complexité. La rétropropagation s’appuie enfin sur une combinaison de pertes sur mesure (focal loss centrée sur les objets, localisation agnostique de classe, Soft Dice et perte de cohérence multi-vues) pour contrer la sparsité et l’inégalité des classes dans les données radar.

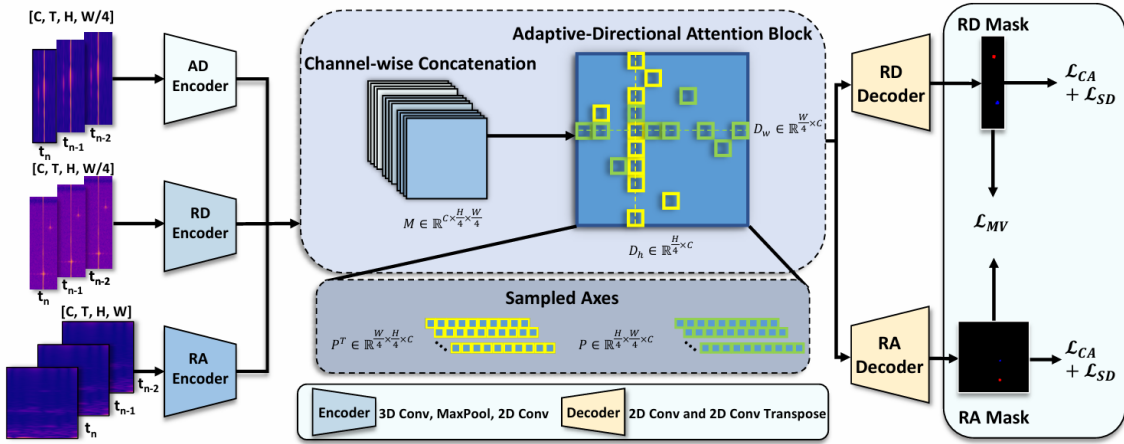


FIGURE 2.20 Schéma de l’architecture de TransRadar : Adaptive-Directional Attention Block pour la segmentation sémantique multi-vues radar FMCW.

Figure utilisée avec permission. © 2024 IEEE.

Bien que ces méthodes aient initialement été développées pour des contextes externes ou automobiles, leur principe de génération de cartes de confiance et d’exploitation des architectures profondes peut également être étendu à la détection radar en environnement intérieur.

2.4 Estimation de la localisation à partir des données brutes

Peu de méthodes de segmentation sont capables de localiser la personne à partir des données brutes. **ADCNet** [48] y arrive en traitant directement les échantillons bruts du convertisseur

analogique-numérique (ADC) pour reconstruire un cube Range–Azimuth–Doppler (RAD). Comme illustré à la Figure 2.21, l’architecture débute par un module de *Learnable Signal Processing* qui apprend, via un réseau convolutif pré-entraîné par distillation de connaissances, à extraire une représentation latente RApprentissage Profond er sans aucune FFT. Cette représentation est ensuite alimentée dans le backbone basé sur FFTRadNet [49]. Le backbone d’ADCNet est basé sur FFTRadNet [49], un réseau convolutionnel rapide conçu pour traiter directement les cubes Range–Doppler (RD). Concrètement, FFTRadNet [49] prend en entrée le tenseur RD issu d’une Transformée de Fourier appliquée aux données ADC, puis apprend à estimer implicitement l’angle d’arrivée (azimuth) sans recourir aux algorithmes de traitement de signal classiques, en s’appuyant sur les signaux de supervision des têtes de détection et de segmentation un réseau convolutionnel rapide chargé de modéliser les dépendances spatio-temporelles. Enfin, deux têtes spécialisées interviennent : la première génère des cartes de chaleur (heatmaps) et des paramètres de régression pour la détection et la localisation fine des cibles, tandis que la seconde produit un masque de segmentation du cube RAD reconstitué. En supprimant l’étape FFT et en apprenant directement sur les données ADC, ADCNet offre une perception radar plus rapide et résiliente face aux artefacts et au bruit, tout en promettant une reconstruction précise des informations de distance, d’angle et de vitesse.

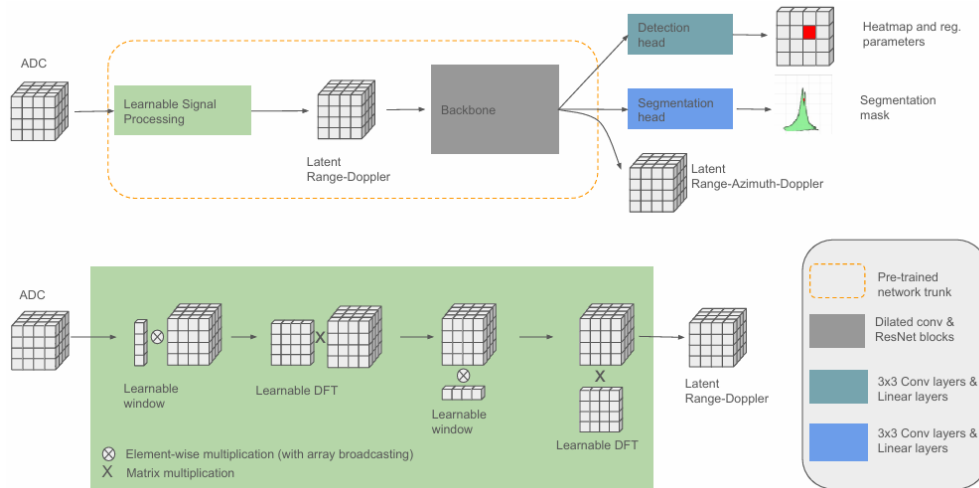


FIGURE 2.21 Architecture d’ADCNet : le module de traitement du signal apprenable (*Learnable Signal Processing*) transforme les données ADC brutes en une représentation distance–Doppler latente, qui est ensuite affinée par un backbone profond. Deux têtes de sortie assurent respectivement la détection (heatmap + régression) et la segmentation du cube distance–azimut–Doppler reconstitué.

Figure utilisée avec permission. © 2023 Bo Yang, Ishan Khatri, Michael Happold & Chulong Chen —
extrait de « ADCNet : Learning from Raw Radar Data via Distillation », arXiv :2303.11420.

Dans DCNN qui est un réseau de neurones convolutionnel profond (Deep Convolutional Neural Network) [4], illustrée à la Figure 2.22, cible des radars FMCW distribués et associe apprentissage profond et apprentissage par transfert. Ce dernier consiste à utiliser un modèle initialement pré-entraîné sur un vaste jeu de données d'images (p. ex. ImageNet) ; seuls ses paramètres des couches supérieures sont ensuite affinés sur les cartes radar, ce qui accélère la convergence et renforce la robustesse lorsque les données radar sont limitées ou bruitées. Les signaux bruts sont d'abord convertis via un module de rétroprojection qui aligne les phases, normalise les amplitudes et génère une représentation spatiale exploitable. Le DCNN extrait ensuite, par des couches convolutives et des opérations de normalisation par lot, les motifs spatio-temporels pertinents, avant qu'un aplatissement et des couches entièrement connectées ne produisent en sortie une carte bidimensionnelle distance/azimut. Applicables en intérieur, ces réseaux apprennent à distinguer les véritables échos cibles des artefacts dus aux réflexions multiples et au bruit, offrant ainsi une localisation de haute résolution dans des environnements encombrés.

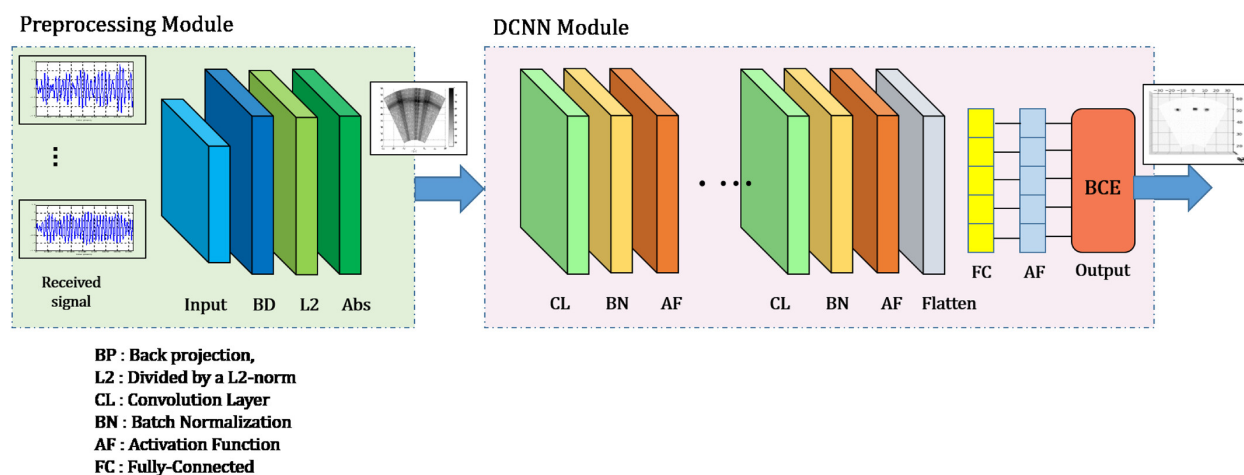


FIGURE 2.22 Architecture du DCNN proposé par Seo et al. [4] pour l'imagerie radar : après la rétroprojection (BP), la normalisation L2 et la prise de valeur absolue (Abs), le réseau applique des blocs convolutionnels (CL) avec normalisation par lot (BN) et fonctions d'activation (AF), suivis d'un aplatissement (Flatten) et de couches entièrement connectées (FC). La sortie est optimisée par une perte d'entropie croisée binaire (BCE). Figure utilisée avec permission

Figure utilisée avec permission. © 2021 IET Radar, Sonar et Navigation

L'approche basée sur les CVNNs (réseaux de neurones à valeurs complexes, en anglais c'est Complex-Valued Neural Networks) [50], qui sont des réseaux de neurones complexes, capables de traiter directement les composantes complexes des signaux, à savoir leur amplitude et leur phase. En modélisant avec précision la relation entre ces propriétés du signal et la

direction d'arrivée, les CVNNs permettent une estimation angulaire plus précise et robuste, même dans des conditions où la qualité des signaux est dégradée. Cette méthode se distingue par sa capacité à optimiser les performances tout en réduisant les besoins en ressources computationnelles, ce qui en fait une solution particulièrement adaptée pour les systèmes radar embarqués et les environnements contraints en énergie et en puissance de calcul. Bien que plusieurs méthodes exploitent déjà les signaux bruts d'un radar FMCW, aucune n'offre une localisation simultanée en distance et en angle pour des applications intérieures.

2.5 Conclusion

Dans ce chapitre, nous avons présenté les principes fondamentaux du radar FMCW, les méthodes de traitement spatio-temporel, ainsi que les cartes Range–Doppler, Range–Azimuth et Range–Angle, qui constituent les représentations clés des signaux radar. Nous avons ensuite analysé, pour les cartes Range–Doppler, les approches classiques (FFT, CFAR, MUSIC, ESPRIT) et mis en évidence leurs limites en termes de résolution et de robustesse, avant d’explorer les méthodes d’apprentissage profond (Bi-LSTM, SGRUN, ST-ConvLSTM, HOOD) qui offrent des gains significatifs tout en nécessitant souvent des prétraitements manuels et des volumes de données importants.

Pour les cartes Range–Azimuth, nous avons examiné les techniques de beamforming (CBF, Capon) et leur impact sur l’estimation angulaire, puis nous avons présenté les architectures hybrides et les réseaux profonds (DeepMUSIC, RODNet, DCNN, ADCNet) capables d’automatiser ces traitements et de traiter les données brutes (IF ou ADC) pour limiter les pertes d’information. Nous avons montré que les solutions end-to-end, opérant directement sur les échantillons radar, sont très prometteuses pour renforcer la robustesse face aux artefacts et au bruit ambiant.

Ces analyses confirment la nécessité de concevoir une démarche intégrée, dans laquelle nous exploitons directement les données radar brutes, tirons parti des dimensions spatiales et temporelles via des architectures profondes, et essayons d’obtenir une généralisation robuste aux environnements réels. C’est dans cette perspective que nous introduisons, au chapitre suivant, notre méthodologie générale pour réaliser un système de localisation radar performant, compact et adapté aux conditions intérieures complexes.

CHAPITRE 3 DÉMARCHE GÉNÉRALE

Pour atteindre l’objectif de ce projet, ce chapitre expose notre démarche en deux volets dégagées de la problématique et de la revue de littérature : d’une part, l’acquisition expérimentale de données radar FMCW dans des environnements variés avec des personnes différentes (section 3.1) ; de l’autre part, la conception et l’entraînement des réseaux de localisation, depuis le traitement des signaux I/Q bruts via cartes de chaleur et DBSCAN (section 3.2) jusqu’à la prédiction finale (section 3.3).

3.1 Acquisition de données radar FMCW pour des scénarios variés

En dépit de l’intérêt croissant pour la localisation humaine, il n’existe à ce jour aucun jeu de données public dédié à la localisation spatiale de personnes en intérieur par radar FMCW multi-antennes. Par exemple, le jeu de données FMCW Radar Dataset de Nguyen et al. [51] fournit des échantillons d’I/Q brutes, c’est-à-dire les composantes d’un signal complexe mesuré directement par l’échantillonneur analogique-numérique : la partie en phase (I pour In-phase) et la partie décalée de 90° (Q pour Quadrature), avant tout traitement ou transformation. Ces données simulées couvrent 11 activités humaines (rotation, marche, chute, etc.), mais reposent uniquement sur des signaux synthétiques et ne prennent pas en compte les configurations multi-antennes ni les environnements intérieurs réalistes encombrés. De même, le corpus *HOOD* de Kahya et al. [52] se concentre sur la détection de présence et l’identification hors distribution à l’aide d’un radar FMCW 60 GHz, mais ses données ne sont pas publiquement disponibles. Enfin, la revue *Awesome Radar Perception* souligne l’absence de jeux de données FMCW pour la localisation de personnes dans un espace intérieur, malgré de nombreux travaux en détection et tracking radar [53].

Pour combler ce manque et démontrer empiriquement la variabilité des signaux radar en présence d’artefacts, nous avons partagé en accès libre notre propre base de données : elle comprend des enregistrements multi-antennes dans deux salles de laboratoires, avec trois profils de participants et des scénarios un et deux personnes (distances 0,5–5 m, angles 0° – 136°), afin de favoriser la reproductibilité et l’évaluation comparative de futurs algorithmes de détection et de localisation.

Cette étude a reçu l’autorisation du Comité d’éthique de la recherche de Polytechnique Montréal (CER-2324-01-D).

Nos expériences reposent sur un radar FMCW monté sur un trépied (voir figure 3.1), sélection-

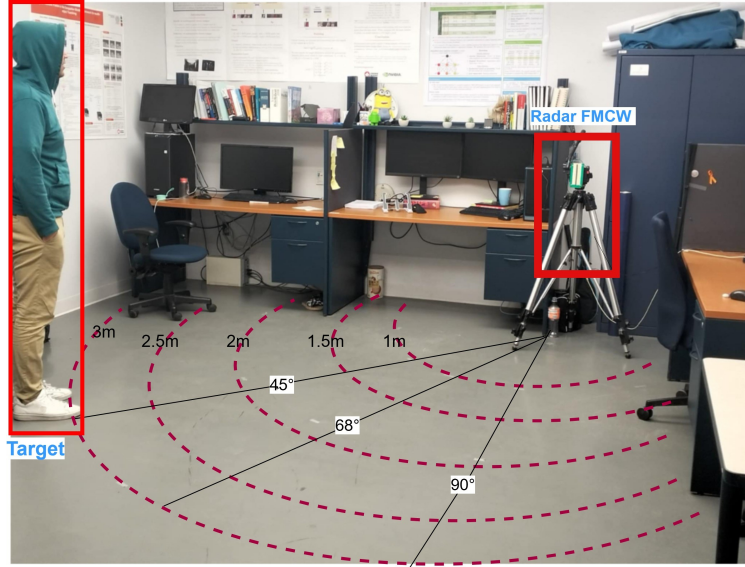


FIGURE 3.1 Configuration d'acquisition des données en laboratoire : le radar FMCW est monté sur un trépied et les positions cibles sont à différentes distances dans l'intervalle de (0,5–5 m) et pour les angles de (0° – 136°).

tionné pour ses performances adéquates en résolution spatiale en distance et en angle au sein d'une pièce de taille moyenne. Le système radar utilisé, opérant à la vitesse de la lumière $c = 3 \times 10^8$ m/s, fonctionne avec une bande passante $B = 200$ MHz et une fréquence d'échantillonnage $f_s = 320$ kHz. Chaque balayage fast-time fournit $N_{fast} = 2048$ échantillons tandis que $N_{slow} = 128$ balayages slow-time permettent de réduire le bruit et les interférences.

TABLEAU 3.1 Paramètres clés du radar FMCW utilisé

Paramètre	Valeur
Vitesse de la lumière c	3×10^8 m/s
Bande passante B	200 MHz
Fréquence d'échantillonnage f_s	320 kHz
Échantillons fast-time (N_{fast})	2048
Balayages slow-time (N_{slow})	128

Pour assurer une représentation équitable des participants, des positions spatiales et des configurations de laboratoire, nous avons appliqué une stratégie de partitionnement stratifié en nous référant aux paramètres expérimentaux présentés dans le Tableau 3.1. Les données des Personnes 1 et 2 couvrent 32 positions en simple-individu dans le laboratoire LAB1 et 22 positions dans le laboratoire LAB2, tandis que la Personne 3 a contribué 6 positions en simple-individu dans LAB1. Pour le scénario à deux personnes (Personnes 1 et 2 simultanément), 15 positions ont été enregistrées dans LAB1 et 10 dans LAB2.

3.2 Localisation des personnes basée sur les cartes de chaleurs et DBSCAN

La méthode que nous proposons consiste à générer directement des cartes de chaleurs (heatmaps) de probabilité sur le plan distance-angle à partir des séquences I/Q brutes, éliminant ainsi tout prétraitement par FFT ou filtrages heuristiques.

Pour ce faire, un réseau de convolution hybride (1D pour extraire les variations de phase liées aux micro-mouvements, puis 2D pour cartographier spatialement ces informations) produit une carte discrète où chaque pixel reflète la probabilité de présence humaine. Afin d’agréger ces informations en points de localisation précis, nous appliquons un DBSCAN auto-adaptatif : le rayon d’épsilon est ajusté dynamiquement en fonction de la densité locale de la heatmap, ce qui permet de regrouper les activations pertinentes tout en éliminant les faux positifs dus aux réflexions parasites. Cette combinaison heatmaps apprises + DBSCAN multi-échelle permet une détection plus robuste quelles que soient les variations d’encombrement et de comportement, sans étape manuelle. À notre connaissance, il s’agit de la première solution de bout en bout pour la localisation radar FMCW indoor, offrant un compromis entre précision et adaptabilité.

3.3 Architecture du modèle

L’architecture, illustrée dans la figure 4.2fig :architecture conçue pour la localisation via des signaux radar repose sur une chaîne de traitement intégrant une analyse multimodale des données et une modélisation contextuelle profonde. Les signaux radar complexes, représentés sous forme de composantes I/Q, sont initialement séparés en deux flux distincts correspondant à leur magnitude et à leur phase.

Cette décomposition permet une exploitation indépendante des informations portées par l’amplitude du signal et de celles contenues dans la phase, qui encapsule des micro-variations temporelles souvent indicatrices de mouvements subtils. Chacune de ces composantes est traitée par un encodeur, nommé PM-Encodeur (Phase-Magnitude Encodeur), constitué de blocs de convolution 1D aux propriétés complémentaires. La branche dédiée à la magnitude/amplitude utilise des filtres larges et peu profonds pour capter les tendances globales et les motifs à large échelle, tandis que la branche phase s’appuie sur des convolutions profondes à noyaux étroits, optimisées pour extraire des détails temporels fins.

Les caractéristiques issues de ces deux flux sont normalisées et projetées dans un espace latent commun avant d’être fusionnées, formant ainsi une représentation enrichie qui alimente le module Transformeur. Ce dernier, composé de plusieurs couches successives, exploite un mécanisme d’attention multi-têtes pour modéliser les dépendances à longue portée au sein des

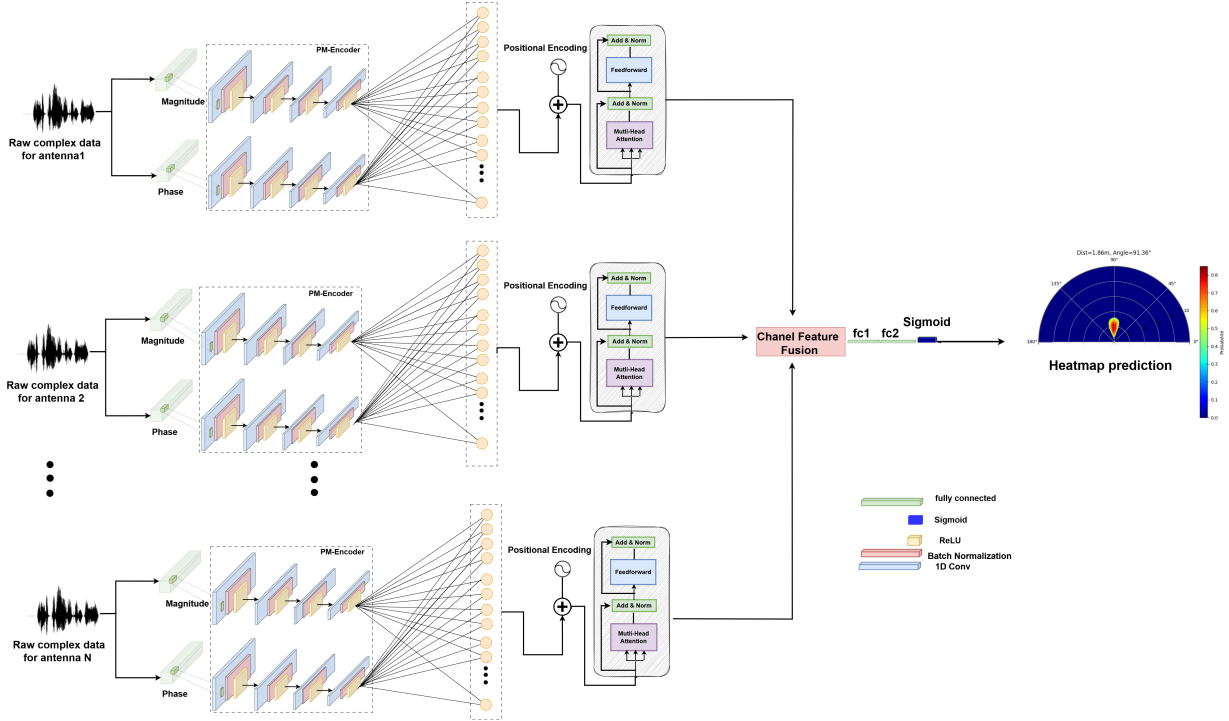


FIGURE 3.2 Architecture proposée, fondée sur les transformeurs, pour la localisation par radar. Les signaux complexes bruts sont décomposés en amplitude et en phase, puis traités par le PM-Encodeur avant d’être affinés par un module à base de transformeur. Les caractéristiques issues de plusieurs antennes sont fusionnées pour générer une carte de chaleur représentant les probabilités de localisation des personnes.

séquences radar, tout en préservant la cohérence temporelle grâce à un encodage positionnel. Le réseau auto-attentif agit comme un raffineur contextuel, pondérant dynamiquement les caractéristiques spatiales et temporelles pour la tâche de localisation.

Les sorties du module transformeur, provenant de multiples antennes radar, sont ensuite harmonisées via un mécanisme d’attention croisée. Ce processus permet de synthétiser les perspectives spatiales divergentes en un cadre unifié, intégrant les informations de chaque capteur pour compenser les angles morts ou les ambiguïtés locales. La représentation consolidée est finalement transformée, via une opération de déconvolution transposée, en une carte de chaleur bidimensionnelle.

Cette carte de chaleur, générée par le réseau puis affinée par un post-traitement DBSCAN, reflète des probabilités normalisées sur la grille (r, θ) ; chaque pixel correspond donc directement à une position angulaire et radiale dans l’espace surveillé. L’opérateur visualise aussitôt les « blobs » les plus probables — c’est-à-dire les zones où la présence humaine est détectée, sans avoir à interpréter des activations internes du modèle. Cette correspondance un-pour-un

entre intensité de la carte (valeurs de 0 à 1) et localisation physique rend la méthode interprétable : la sortie finale n'est pas un score abstrait, mais une indication spatiale explicite qui peut être exploitée pour le suivi et la localisation.

L'ensemble du pipeline allie la précision locale des convolutions, adaptée aux motifs radar complexes, et la capacité des transformeurs à raisonner sur des contextes étendus, permettant une robustesse accrue face aux artefacts et aux bruits inhérents aux environnements réels.

CHAPITRE 4 ARTICLE 1 : MULTI-BRANCH FEATURE FUSION NETWORK FOR RADAR-BASED HUMAN LOCALIZATION

Wissal Zarrami¹ Guillaume-Alexandre Bilodeau¹ François Nougrou²
Sylvain Azarian³ Bruno Duperrier³ Pape Sanoussy Diao³

¹LITIV, Polytechnique Montréal, Montréal, Canada
{wissal.zarrami, gabilodeau}@polymtl.ca

²Département de génie électrique et informatique,
Université du Québec à Trois-Rivières (UQTR), Québec, Canada
Francois.Nougrou@uqtr.ca

³Morphée+. Inc, Québec, Canada

This article was submitted to the Journal IEEE Transactions On Signal Processing, April the 14th, 2025.

Abstract

Accurate and privacy-preserving indoor human localization is a requirement for numerous applications, especially for elder care, to detect falls and for security monitoring. This paper presents an end-to-end deep learning approach that processes raw Frequency-Modulated Continuous Wave (FMCW) radar data for robust human localization. Unlike traditional methods that rely on extensive signal preprocessing (e.g., windowing, Fourier transforms), our method operates directly on raw intermediate frequency (IF) signals to capture both amplitude and phase information. We introduce a Phase-Magnitude Encoder to extract features from raw complex signals and an Antenna-wise Transformer Encoder to independently process and capture long-range dependencies across multiple antennas. The proposed method is adaptable to various antenna configurations. As well, we introduce a new dataset with our FMCW radar system that includes two indoor environments with different clutter conditions, as well as several individuals exhibiting varied body shapes and movement patterns. Experimental results demonstrate that our method outperforms baseline and state-of-the-art techniques, achieving precise localization with sub-meter level distance accuracy and low angular errors in one or two persons scenarios. Code and Dataset are available at link.

keywords : FMCW radar, human localization, privacy-preserving technologies, radar signal processing, indoor positioning.

4.1 Introduction

Indoor Localization Systems (ILSs) [54] have shown great promise in improving indoor monitoring capabilities and helping in delivering effective elderly care. They can be used to locate elderly and identify falls. These systems have become more prevalent in recent years [55]. According to the United Nations (UN), the global elderly population is expected to increase to 1.5 billion by 2050 [56], with 28-35% of them expected to experience a fall annually [57]. This demonstrates the importance of using technology in elder care for those who are losing their autonomy, but still want to live in their own homes and require ongoing, in-the-moment support [55].

Nevertheless, these monitoring system faces several challenges : (1) they should be none obtrusive, (2) they should be privacy-preserving, and (3) they should be reliable. Therefore, passive indoor localization systems, which localize individuals without the need for wearable devices, have gained increasing attention. WiFi-based localization is a popular choice due to its wide availability and easy deployment. However, it mainly relies on angle-of-arrival (AoA) and time-of-flight (ToF) estimations, which have low resolution because of limited bandwidth and a small number of antennas in standard WiFi devices [58]. In comparison, camera-based systems offer high accuracy but face the challenges of privacy concerns and sensitivity to lighting and obstacles [59, 60].

In this context, radar-based systems, particularly Frequency-Modulated Continuous Wave (FMCW) radar, have emerged as a promising solution. Unlike visual sensors, radar systems inherently respect privacy by capturing non-visual information, and they remain effective under poor lighting and challenging environmental conditions [61]. In conventional FMCW radar systems, target detection typically relies on preprocessing steps such as windowing and filtering, followed by Fourier-based transformations (e.g., DFT) to extract distance and velocity information [62]. These methods are often coupled with beamforming techniques like digital beamforming (DBF) [63] or Capon's algorithm to estimate the angle-of-arrival (AOA) by analyzing spatial phase differences across antenna arrays [64].

The task of estimating range (or distance) and angle in radar systems can be approached by dividing it into signal processing and target parameter extraction. However, this approach can fail in scenarios requiring direct angle and distance estimation from raw, unprocessed signals corrupted by noise because : (1) traditional beamforming assumes clean, high signal-to-noise ratio (SNR) signals to resolve phase differences, but noise and interference distort spatial coherence, leading to angular ambiguities and degraded resolution, (2) without pre-processing or Fourier-based spectral analysis, extracting precise distance information from

raw intermediate frequency (IF) signals becomes challenging, as noise obscures subtle frequency shifts caused by target reflections, (3) Multipath propagation effects, exacerbated by cluttered environments, further complicate angle estimation by introducing false spatial signatures, and (4) the reliance on predefined array geometries and static beamforming weights limits adaptability to dynamic noise conditions, resulting in poor robustness.

To overcome these limitations, recent works have explored data-driven techniques that bypass traditional preprocessing and directly extract target parameters from raw radar data [4]. Inspired by applications in fields like gesture recognition [65] and autonomous driving [66], where raw data is used as input to neural networks, we can adopt a similar strategy for range and angle estimation. By feeding raw intermediate frequency (IF) signals into deep learning models, the system can learn to identify patterns and extract range and angle information without relying on manual parameter tuning or predefined transformations. This approach is particularly advantageous in noisy environments, where traditional methods have difficulties to distinguish target reflections from interference and clutter. Neural networks can adapt to dynamic conditions and complex signal distortions, offering a more robust and flexible solution for range and angle estimation.

Unlike several methods from previous works that impose rigid, domain-specific transformations (e.g., Fourier-based spectral decomposition) to reduce raw signals to engineered features, our method retains the full high-dimensional structure of raw FMCW waveforms. By processing the raw signal in its native time-frequency representation, our framework avoids collapsing critical interactions between temporal, spectral, and spatial dimensions. This preservation of the original signals as input to our neural network model enables it to learn robust representations that takes into account multipath reflections, micro-Doppler signatures, and environmental perturbations directly, bypassing the information-limiting abstractions of traditional radar processing. Therefore, our decision to use the raw signal from each antenna as input is driven by the need to retain the full signal information without introducing distortions or losing phase relationships.

The main contributions of our work are :

- We propose a new data-driven method that enables our model to learn the best feature representations without the need for manual adjustments by processing raw signals directly.
- We introduce a Phase–Magnitude Encoder (PME), which jointly processes magnitude and phase information from raw FMCW radar signals. Unlike traditional methods that discard phase early, the PME preserves phase information while extracting meaningful spatial features.

- We integrate an Antenna-wise Transformer Encoder, a Transformer-based multi-head attention mechanism that learns contextual relationships across spatial and temporal radar data to capture long-range dependencies.
- We constructed a new publicly available FMCW People Localization Dataset using our radar system, capturing real-world multi-person localization data in two environments with different clutter levels.

4.2 FMCW Radar Background

4.2.1 Operating Principle of FMCW Radar

Frequency Modulated Continuous Wave (FMCW) radars are well-known for achieving high range resolution with low power consumption. As illustrated in Fig. 4.1, the radar transmits a *chirp* signal whose instantaneous frequency increases linearly from f_c to $f_c + B$ over a period T_c . We define the chirp rate S and the instantaneous frequency $f(t)$ by :

$$f(t) = f_c + St, \quad S = \frac{B}{T_c}, \quad (4.1)$$

where f_c is the initial (carrier) frequency and B is the total bandwidth swept during each chirp. By controlling B and T_c , one can tune the radar's range resolution and maximum unambiguous range, respectively. Typically, the radar hardware includes :

- A **frequency synthesizer** or Phase-Locked Loop (PLL) that generates the chirp.
- A **transmitter** that amplifies and radiates the chirp signal through a TX antenna.
- One or more **receive** antennas and front-end circuits that down-convert and digitize the reflected signals.

When the chirp encounters a target at range d , the received signal is delayed by $\tau_1 = \frac{2d}{c}$ (accounting for the round-trip time). This delay produces a constant frequency offset, or *beat frequency*, directly proportional to the distance of the target.

4.2.2 Beat Frequency and IF Signal

The transmitted chirp is modeled as :

$$s_{\text{tx}}(t) = A \cos\left(2\pi\left[f_c t + \frac{1}{2}St^2\right]\right), \quad (4.2)$$

where A is the amplitude. For a target at range d , the received signal is :

$$s_{\text{rx}}(t) = A \cos\left(2\pi\left[f_c(t - \tau_1) + \frac{1}{2}S(t - \tau_1)^2\right]\right). \quad (4.3)$$

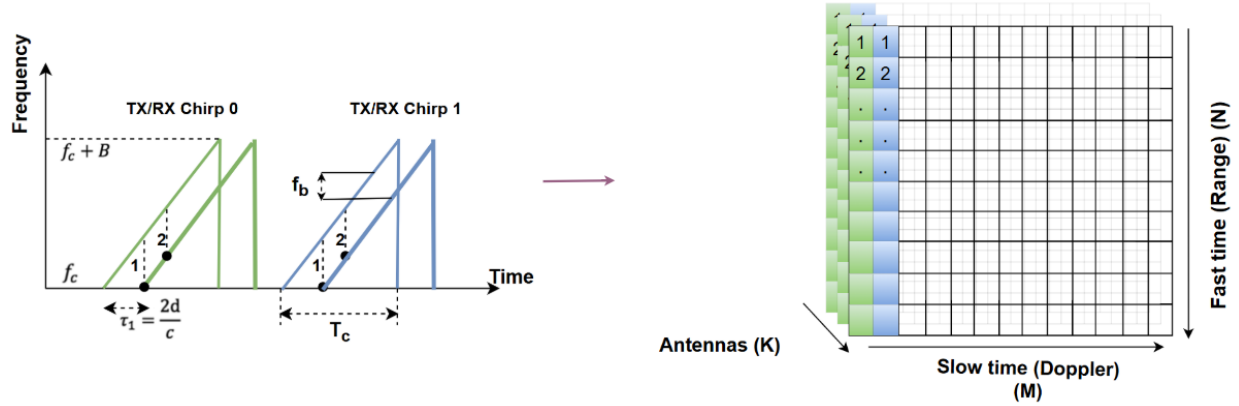


FIGURE 4.1 (Left) Transmitted (TX) and received (RX) chirps : The beat frequency f_b arises from the delay τ_1 between TX and RX signals. (Right) Data cube structure : Dimensions include antennas (spatial), slow time (Doppler), and fast time (range).

The delay τ_1 introduces a slight difference in instantaneous frequency between $s_{\text{tx}}(t)$ and $s_{\text{rx}}(t)$. In hardware, the transmitted and received signals are *mixed* (multiplied) and then low-pass filtered to yield the **intermediate frequency** (IF) signal :

$$s_{\text{IF}}(t) = A \cos(2\pi f_b t + \phi), \quad (4.4)$$

where ϕ is an initial phase term, and the *beat frequency* f_b is :

$$f_b = S\tau_1 = \frac{2Bd}{cT_c}. \quad (4.5)$$

Because f_b is directly proportional to the target's range d , measuring f_b enables the radar to estimate distance. The amplitude of $s_{\text{IF}}(t)$ also encodes information about reflectivity and target cross-section.

4.2.3 Raw Data Acquisition and Signal Processing

In a typical radar frame, the system transmits M chirps (often called “slow-time” samples) and samples N points (“fast-time” samples) per chirp over K antennas. As illustrated in Fig. 4.1, this results in a three-dimensional data cube :

$$\underbrace{\text{Antennas (K)}}_{\text{spatial/Angle}} \times \underbrace{\text{Slow time (M)}}_{\text{Doppler/velocity}} \times \underbrace{\text{Fast time (N)}}_{\text{range}},$$

where :

- **Fast time (N)** : Captures the IF signal samples within each chirp. An FFT across these samples reveals the beat frequencies f_b and thus the range profile.
- **Slow time (M)** : Corresponds to the sequence of chirps in one frame. Phase shifts across consecutive chirps reveal Doppler information (i.e., target velocity).
- **Antennas (K)** : Represents spatial sampling across the receive antenna array. Phase differences between antennas are used to perform angle estimation.

4.3 Related Works

To localize people using a radar, traditional approaches typically process beat signals with Fast Fourier Transforms (FFT) to extract distance information from frequency shifts, but they can have difficulties with noise and multiple paths in complex indoor settings. Park et al. [67] addressed these challenges by employing a feedforward neural network that takes FFT-processed FMCW radar data as input to classify distances from 1–10 meters into discrete categories. Despite its improved accuracy over classic FFT-based techniques, this method is limited to controlled indoor settings and discrete distance outputs. Likewise, Yang and Kim [68] transformed range FFT snapshots from a single radar into 2D images, which they then feed into a CNN that classifies discrete grid positions, effectively framing human detection as an image classification problem. Their method, however, remains limited by single-target assumptions.

While range estimation localizes targets along the radial axis, angular positioning is needed for resolving spatial ambiguities in multi-target indoor scenarios. This necessitates robust angle-of-arrival (AoA) estimation techniques. To improve over MUSIC [69] and beamforming [70], Elbir et al. [43] introduced DeepMUSIC, which partitions the angular domain into subranges, assigning a dedicated CNN to each subrange to learn the MUSIC spectrum from the real, imaginary, and phase components of the covariance matrix. While this design handles multiple targets with reduced complexity, its limitation is that it assumes at most one target per subrange, and thus depends on both accurate subrange partitioning and a representative training set for reliable performance. Meanwhile, Sang et al. [71] proposed a 3D CNN operating directly on a cropped data cube (range–antenna–chirp) to capture real, imaginary, and magnitude information. Although these deep learning methods often surpass classical solutions by preserving local phase features, they rely on diverse training sets and can degrade under unrepresented angles or small SNR conditions.

While the aforementioned works focus on angle estimation, high-resolution systems often need joint range–angle estimation to avoid resolution loss from range–angle coupling. Addressing

this issue, Han and Hong [72] proposed a wavelength-dependent beamforming approach that decouples distance and angle by adjusting DFT lengths per frequency bin, thereby restoring angular resolution in off-boresight directions without sacrificing range accuracy. Similarly, Xing et al. [73] proposed a pipeline that processes FMCW data into range-angle spectrograms, applying DBSCAN clustering and multipath analysis to localize static humans with a 0.2 meter accuracy in indoor settings with plants, curtains, or furniture. However, it relies on partial scene knowledge, limiting adaptability in dynamic environments, and has a high computational cost $O(N^2)$ due to DBSCAN.

Both range and angle estimation are inherently sensitive to noise and interference in indoor settings, where multipath reflections and dynamic clutter dominate. Effective noise suppression is essential to isolate human-generated signals from static clutter and random noise. In this context, Peng et al. [74] proposed a detection system that fuses multiple radar representations, such as range-Doppler and range-angle images, into a 3D input, which is then processed by a 3D CNN enhanced with deep metric learning. This approach effectively distinguishes subtle human motions (e.g., respiration) from the similar movements of swinging blinds, thereby achieving higher detection accuracy even in complex indoor environments. HOOD, introduced by Kahya et al. [52] used both macro- and micro range-Doppler maps to capture motion information. By processing these maps through a reconstruction-based multi-encoder multi-decoder framework, the system extracts prominent and subtle features (such as respiration), and uses reconstruction errors with multi-threshold criteria to differentiate human presence from out-of-distribution clutter. However, the method success hinges on finely tuned thresholds and an extensive training dataset, and its complex architecture can impose significant computational overhead.

4.4 Methodology

Figure 4.2 illustrates our proposed Multi-Branch Feature Fusion Network for radar-based human localization.

The architecture processes raw complex-valued signals from any number of receiver antennas using three core components :

1. A Phase Magnitude-Encoder (PME) that independently extracts spatial-frequency features from each antenna. Each antenna data stream is processed by identical neural subnetworks without parameter sharing. This design preserves spatial diversity inherent to radar signals, such as antenna-specific phase variations required for angle-of-arrival estimation, while mitigating biases that can be introduced by shared weights. By maintaining independent parameters for each branch, the architecture adapts to

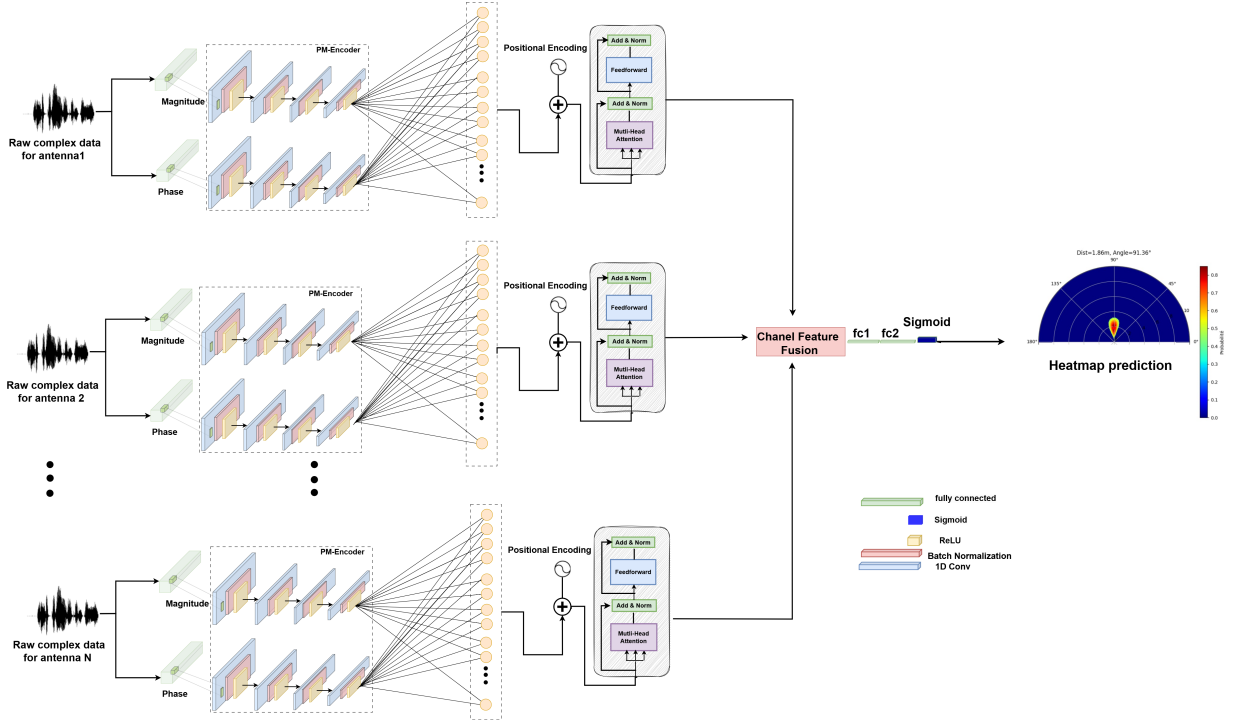


FIGURE 4.2 Proposed Transformer-based architecture for radar-based localization. Raw complex signals are decomposed into magnitude and phase, processed through the PM-Encoder, and refined using a transformer-based module. Features from multiple antennas are fused to generate a heatmap of prediction probabilities for human localization.

- hardware imperfections (e.g., antenna gain mismatches) and environmental artifacts (e.g., spatially distinct reflections).
- 2. an Antenna-wise Transformer Encoder that refines these representations by modeling long-range dependencies across time and frequency, and
- 3. A Channel Feature Fusion Module that adaptively aggregates features across antennas using learnable weights.

The network predicts a probability heatmap in polar coordinates relative to the radar origin (positioned at (0,0)), enabling human localization by jointly estimating angle and radial distance.

4.4.1 Magnitude-Phase vs. Real-Imaginary Representations

A limitation of Cartesian representations is that they split complex signals into separate real and imaginary parts. This disrupts the natural structure of the complex plane, reducing wavefront coherence, which should be preserved for detecting small path differences in radar signals. As a result, Cartesian models have difficulty capturing phase variations, especially

in situations with closely spaced targets [75].

The polar decomposition of complex radar signals into amplitude and phase streams aligns better with the physical interpretability of these modalities in radar perception tasks. Our PME design explicitly prioritizes the inherent geometric separability of phase and magnitude in radar reflected signals, because the magnitude polar component correlates with energy-based scattering properties, while the phase polar component encodes wave propagation dynamics.

Figure 4.3 shows how feature representation differs when 1D convolutions are applied to real and imaginary components independently as compared to employing magnitude and phase components. The top row of the magnitude-phase spectrograms shows smooth transitions and well-structured frequency responses with phase continuity across time. The phase misalignment errors seen in the real-imaginary spectrograms (bottom row), on the other hand, may result in less instructive spatial feature extraction.

4.4.2 Phase–Magnitude Encoder (PME)

Each antenna captures the radar signal as a complex sequence :

$$z_n = x_n + j y_n, \quad (4.6)$$

where x_n and y_n denote the real (in-phase) and imaginary (quadrature) components, respectively. We decompose it into its magnitude and phase representations :

$$|z_n| = \sqrt{x_n^2 + y_n^2}, \quad \theta_n = \arctan 2(y_n, x_n). \quad (4.7)$$

We introduce a novel Phase–Magnitude Encoder (PME) that processes raw magnitude-phase radar data. The magnitude and phase are separate 1D input vectors (e.g. size (224756, 1) in our experiments) by decomposing them into parallel magnitude and phase streams. Each stream is independently fed into identical convolutional neural network (CNN) branches with non-shared weights, ensuring modality-specific feature learning while maintaining architectural symmetry to preserve consistency in feature space alignment, allowing the model to effectively fuse phase and magnitude representations at later processing stages without introducing distortions. The magnitude branch focuses on extracting energy distribution patterns (e.g., target reflectivity and clutter intensity), while the phase branch captures spatial dependencies (e.g., antenna-specific delays and interference signatures) needed for resolving angular ambiguities. Both branches employ the same four convolutional layers followed by

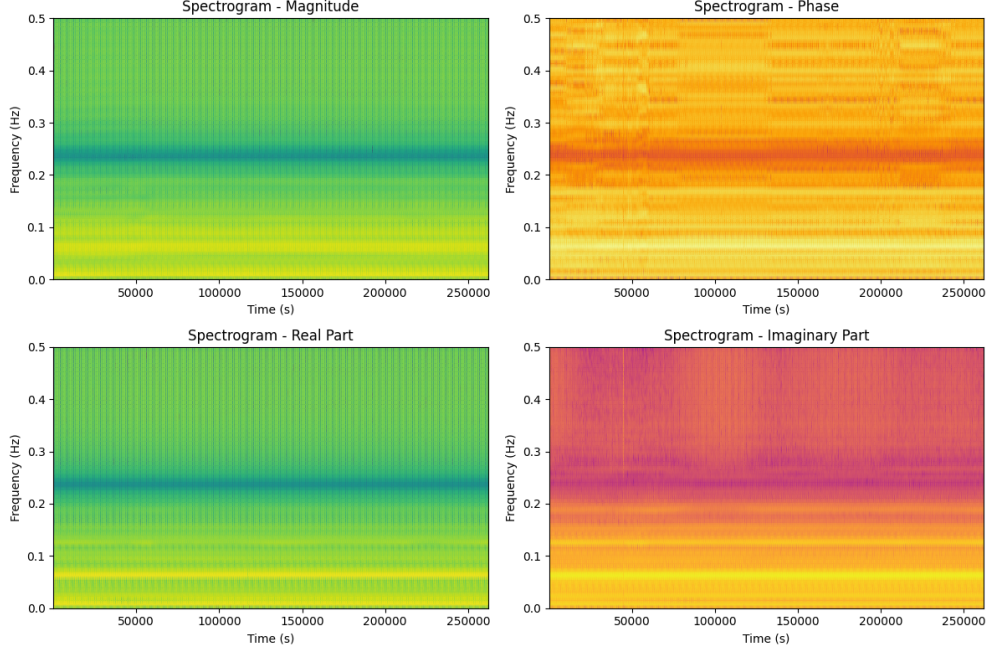


FIGURE 4.3 Comparison of spectrogram representations after applying 1D convolution. (Top-Left) Magnitude spectrogram, (Top-Right) Phase spectrogram, (Bottom-Left) Real part spectrogram, (Bottom-Right) Imaginary part spectrogram.

batch normalization, and nonlinear activations RELU. The encoded magnitude and phase representations are fused only at the final stage via channel-wise concatenation, preserving their distinct inductive biases until higher-level abstractions are formed. By decoupling the processing of magnitude, the PME mitigates noise amplification in phase data and enhances robustness to environmental artifacts.

For each antenna i , the intermediate features from the magnitude ($x_{\text{mag}}^{(i)}$) and phase ($x_{\text{phase}}^{(i)}$) streams are fused using a layer-specific multi-layer perceptron (MLP) using :

$$x_{\text{fused}}^{(i)} = \text{MLP}^{(i)} \left(\text{Concat} \left(x_{\text{mag}}^{(i)}, x_{\text{phase}}^{(i)} \right) \right). \quad (4.8)$$

The MLP generates a compact representation by transforming concatenated features into a lower-dimensional space through weighted mappings. It emphasize the relevant features while reducing dimensionality. By filtering out redundant and noisy information, it minimizes overfitting and lowers computational costs, enhancing overall model efficiency.

4.4.3 Antenna-wise Transformer Encoder

To capture global dependencies within the fused features, each PME output $x_{\text{fused}}^{(i)}$ is processed through a transformer encoder layer. Initially, query (Q), key (K), and value (V) embeddings are computed using learnable linear projections :

$$Q = W_Q x_{\text{fused}}^{(i)}, \quad K = W_K x_{\text{fused}}^{(i)}, \quad V = W_V x_{\text{fused}}^{(i)}, \quad (4.9)$$

where $W_Q, W_K, W_V \in \mathbb{R}^{C \times d}$ are learnable weight matrices.

Multi-head attention allows the model to jointly attend to information from different representation subspaces at different positions. In contrast, a single attention head typically averages these interactions, potentially limiting expressiveness. Multi-head attention is defined as :

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h) W_O, \quad (4.10)$$

where each head head_i computes :

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V). \quad (4.11)$$

The output of the multi-head self-attention (MHSA) is combined with a residual connection to enhance information flow and stabilize learning :

$$z' = \text{MHSA}(x_{\text{fused}}^{(i)}) + x_{\text{fused}}^{(i)}. \quad (4.12)$$

Subsequently, a feedforward network (FFN), augmented by Batch normalization (BN), refines the representation :

$$x_{\text{output}}^{(i)} = \text{BN}(W_2 \cdot \text{ReLU}(W_1 \cdot \text{LN}(z') + b_1) + b_2), \quad (4.13)$$

where W_1, W_2 are learned weight matrices, and b_1, b_2 represent bias terms enabling the network to introduce flexible shifts in activation, thereby improving expressiveness. By employing this transformer-based processing, the encoder effectively models long-range dependencies, capturing critical non-local correlations such as transient phase shifts and persistent magnitude patterns within the radar signals.

4.4.4 Channel Feature Fusion

Our architecture processes each antenna signal independently, so that each antenna retains its intrinsic characteristics before fusion. Each antenna data is passed through identical processing blocks, including the Phase–Magnitude Encoder (PME) and Antenna-Wise Transformer Encoder, generating unique feature vectors $x_{\text{output}}^{(1)}, x_{\text{output}}^{(2)}, \dots, x_{\text{output}}^{(N)}$ for N antennas. Instead of directly merging these features, we employ a channel-wise feature fusion strategy to enhance feature interactions while maintaining spatial and contextual integrity.

Each feature vector is treated as a separate channel in the feature space, and a channel-wise fusion operation is performed by concatenating them along the channel dimension as it is presented in the figure 4.4 with :

$$f_{\text{fused}} = \text{Concat}(x_{\text{output}}^{(1)}, x_{\text{output}}^{(2)}, \dots, x_{\text{output}}^{(N)}). \quad (4.14)$$

To effectively integrate complementary information from different antennas while reducing redundancy, we apply a 1×1 convolution across the concatenated channels :

$$f_{\text{final}} = \sigma(\text{BN}(\text{Conv}_{1 \times 1}(f_{\text{fused}}))), \quad (4.15)$$

where σ represents a non-linear activation function such as ReLU, and Batch Normalization (BN) ensures stable feature distribution.

Processing each antenna separately preserves its unique spatial and phase information, which would be otherwise lost if signals were fused too early. Each antenna captures different propagation effects and spatial perspectives, providing complementary information for human localization. Channel-wise feature fusion then integrates these distinct embeddings while maintaining their information content, enhancing target discrimination and localization accuracy.

4.4.5 Heatmap and Target Localization

A feed-forward network (FFN) in the final step of our architecture generates a prediction probability heatmap, which represents the spatial probability distribution of the human locations within the scene. Framing the localization task as a regression problem in a heatmap rather than a classification one allows the model to predict the position of several human simultaneously.

Then, a post-processing pipeline transforms raw heatmap predictions into target coordi-

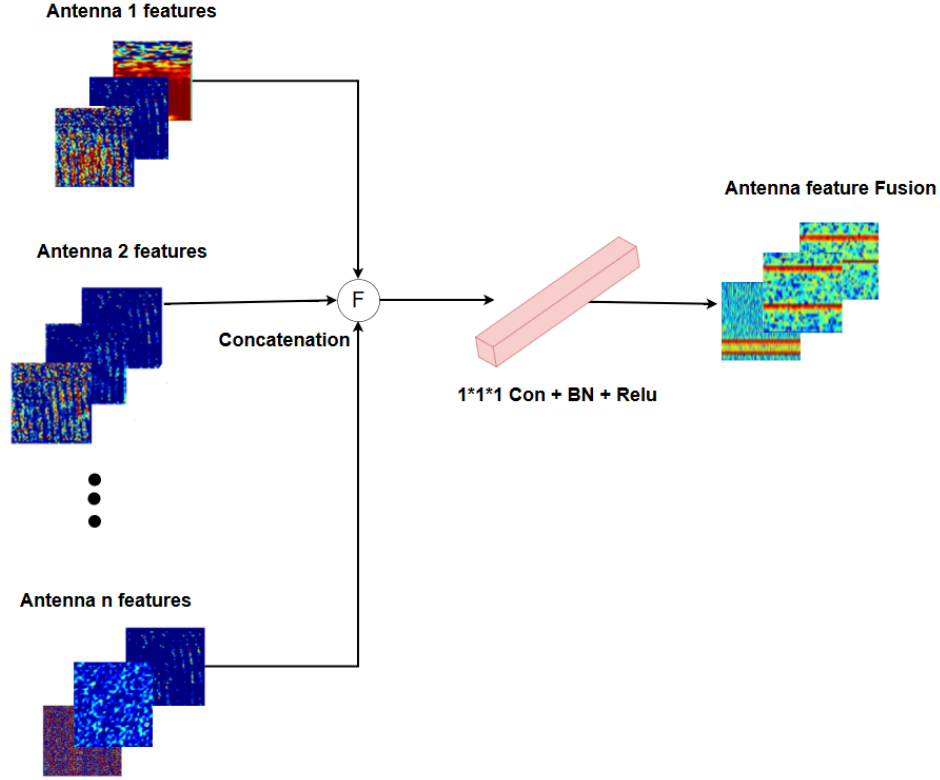


FIGURE 4.4 Antenna Feature Fusion : Features from multiple antennas are concatenated and processed through a $1 \times 1 \times 1$ convolution with batch normalization and ReLU.

nates. While the heatmap provides a spatial probability distribution of target presence, it often contains noise, overlapping activations, and ambiguous peaks that may obscure true human locations. By applying adaptive thresholding, low-confidence regions are suppressed, retaining only statistically significant areas likely to correspond to genuine human locations. Subsequent local maxima detection isolates precise candidate coordinates, resolving ambiguities in dense or overlapping activations, a common challenge in cluttered environments. Finally, DBSCAN clustering groups spatially close maxima while filtering outliers, ensuring robustness against false positives caused by transient artifacts or residual noise. This hierarchical refinement method processes coarse heatmap predictions into final locations for each human in the scene.

4.4.6 Polar Heatmap Ground Truth Generation

To supervise the training process, we generated polar heatmaps that encode the true human position in both range and angle. The ground truth is constructed by modeling a human location using the product of two Gaussian functions, one in the range dimension and one in

the angular dimension, centered at the human actual coordinates (r_0, θ_0) :

$$H(r, \theta) = \exp\left(-\frac{(r - r_0)^2}{2\sigma_r^2}\right) \cdot \exp\left(-\frac{(\theta - \theta_0)^2}{2\sigma_\theta^2}\right), \quad (4.16)$$

where σ_r and σ_θ determine the spread in the range and angular dimensions, respectively. The resulting heatmap is normalized to the interval $[0, 1]$, providing the supervisory signal during training. An example of this representation is shown in Figure 4.5, where a heatmap is generated to indicate the presence of a person at a specific range and angle.

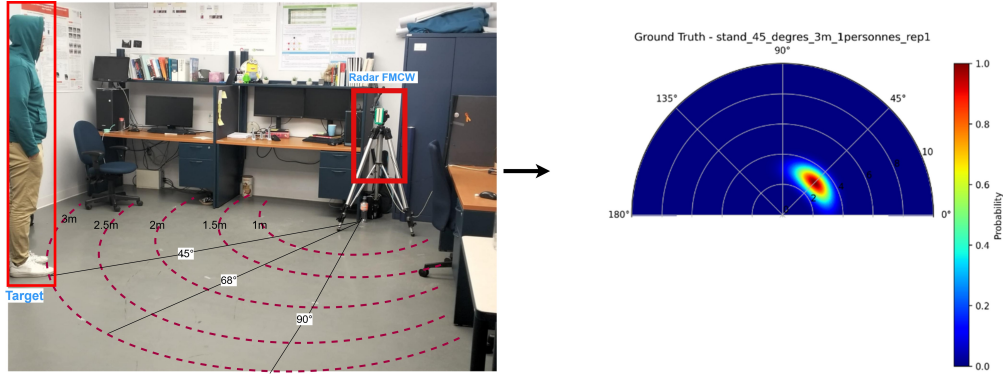


FIGURE 4.5 Heatmap generation for a person in the position range=3 and angle=45 degrees in LAB1

4.5 Experimental Results

4.5.1 Data Acquisition

In our experiments, data was acquired using an FMCW radar system configured to deliver a good spatial resolution in both range and angle for a medium size room. The radar operates using the speed of light $c = 3 \times 10^8$ m/s, a chirp bandwidth B of 200 MHz, and a sampling frequency f_s of 320 kHz. During each fast-time sweep, 2048 samples are collected. In addition, 128 sweeps are recorded during the slow-time period. The bandwidth B directly sets the range resolution ΔR of the system, which can be calculated as

$$\Delta R = \frac{c}{2B}. \quad (4.17)$$

For a 200 MHz bandwidth, this yields a theoretical range resolution of approximately 0.75 m. Table 4.1 summarizes the main parameters of the radar system. These settings have been chosen to make the system suitable for person detection and localization in practical

applications.

TABLEAU 4.1 FMCW Radar Parameters for Data Acquisition

Parameter	Value
c (Speed of Light)	3×10^8 m/s
B (Bandwidth)	200 MHz
f_s (Sampling Frequency)	320 kHz
$N_{\text{fast time}}$	2048 samples
$N_{\text{slow time}}$	128 samples

To build our dataset, we selected three individuals with distinct physical characteristics, with various shape, age, height, and weight. Those different human profiles allow us to study whether the model can generalize to different human signals. A person naturally introduces micro-movements such as hand gestures, slight posture adjustments, or involuntary actions, like coughing or shifting weight, all of which affect the reflected radar signal. Furthermore, the dataset includes both single-person and two-person scenarios, with up to three different individuals.

During the acquisition process, data was collected from both antennas of the radar, with each antenna capturing complex radar data in the form of a vector $x + iy$. To account for variations, each position was recorded three times across separate sessions, with 1-minute recordings per session, yielding 18 acquisition cycles (9 cycles per antenna) per position. Each cycle produced complex-valued radar data with a raw shape of (524288×2) , representing simultaneous signal capture from both antennas. This was refined into two complex-valued data vectors of shape (262144×1) , one per antenna, retaining full spatiotemporal resolution.

The entire process was conducted in two different laboratory environments to introduce diversity in scene-dependent reflections and radar signal propagation. Figure 4.6 illustrates the configurations of both acquisition setups, in which we have differences in spatial constraints and surrounding clutter. The chosen environments allow covering distances from 0.5 to 5 m and angles from 0° to 136° .

4.5.2 Dataset Partitioning

To ensure a good representation of individuals, spatial positions, and different lab conditions, we implemented a stratified partitioning strategy, where data from Person 1 and 2 totaled to 32 single-person positions in LAB1 and 22 single-person positions in LAB2. Person 3 contributed 6 LAB1 positions in the single-person scenario. Moreover, to evaluate the two-person scenario, we recorded 15 positions for two persons (person 1 and person 2) present at



FIGURE 4.6 Data acquisition environments : LAB1 (right), LAB2 (left). LAB1 has an higher clutter levels, with additional interference caused by furniture, walls, and electromagnetic reflections. The second environment (LAB2) exhibited lower clutter levels. Ventilation systems, and other active equipment contributed additional noise.

the same time in LAB1 and 10 two-person positions in LAB2.

For each person, the data was split into 70% training, 15% validation, and 15% testing subsets, ensuring all positions (and their three repeated sessions) were proportionally represented in each subset. By distributing repeated sessions across splits, natural variations from human micro-movements (e.g., posture shifts) and environmental fluctuations (e.g., sensor noise) were embedded into the dataset (available at link). We train our heatmap prediction model with four attention heads and a hidden layer size (hidden size=2048). In the training phase, we uses the Adam optimizer with learning rate=1e-4 and weight decay=1e-5. The loss function is **MSE Loss**, combined with a ReduceLROnPlateau scheduler (patience=5, factor=0.5). For implementation details, our code is available at the repository.

4.5.3 Experimental Setup

For evaluation, four experiments were conducted :

(1) **Location Estimation Errors in Individual and Two- Person Scenarios** where the model was trained on combined data from three individuals (Person 1, 2, and 3) from both labs LAB1 and LAB2 and tested separately on each person data. This experiment evaluates the performance of the proposed model in estimating locations and learning to distinguish between one and two persons where all the persons have been seen ; (2) **Inter-person**

testing where we tested all three combinations of training the model on two individuals (e.g., Persons 1+2, 1+3, 2+3) using mixed data from Lab1 and Lab2, then evaluated it on the excluded third person; **(3) Cross-environment testing**, where a model trained on all the persons in LAB1 is tested on the data of the LAB2 and vice versa, testing for scene-specific variability (e.g., room layout); and **(4) Evaluation of Predictions on Unseen angles or distances** where the model was trained on a predefined range of angles and distances, then tested on entirely new values outside this training scope.

4.5.4 Evaluation Metrics

To evaluate our model performance, we quantified the localization accuracy using the distance error and angular error. The distance error reflects the deviation between the true range and the estimated range of the person, and is defined as :

$$E_{\text{dist}} = |r_{\text{gt}} - r_{\text{pr}}|. \quad (4.18)$$

where :

- r_{gt} represents the ground truth range of the person.
- r_{pr} denotes the estimated range obtained from the model.

Similarly, the angular error measures the absolute difference between the true angle and the predicted angle :

$$E_{\text{ang}} = |\theta_{\text{gt}} - \theta_{\text{pr}}|. \quad (4.19)$$

where :

- θ_{gt} represents the ground truth angle of the person.
- θ_{pr} denotes the estimated angle obtained from the model.

To further evaluate our model performance, we compute the Mean Absolute Error (MAE) and the standard deviation (Std) for both distance and angle, defined as follows :

$$MAE_{\text{dist}} = \frac{1}{N} \sum_{i=1}^N E_{\text{dist}}^{(i)}, \quad MAE_{\text{ang}} = \frac{1}{N} \sum_{i=1}^N E_{\text{ang}}^{(i)}. \quad (4.20)$$

Finally, the standard deviation of the distance and angle errors is :

$$Std_{\text{dist}} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left(E_{\text{dist}}^{(i)} - MAE_{\text{dist}} \right)^2}. \quad (4.21)$$

$$Std_{\text{ang}} = \sqrt{\frac{1}{N-1} \sum_{i=1}^N \left(E_{\text{ang}}^{(i)} - MAE_{\text{ang}} \right)^2}. \quad (4.22)$$

4.5.5 Performance for the various test scenarios

Location Estimation Errors in Individual and Two-Person Scenarios

As shown in Table 4.2, our model under individual and multi-person scenarios, has the best distance estimation with Person 3 (MAE : 0.238, Std : 0.171), though this likely reflects overfitting as this person has very few samples, whereas Persons 1 and 2 demonstrate slightly better results in angle estimation (MAEs : 2.0°–2.3°), benefiting from larger and more diverse training examples (32+22 single-person and 25 two-person poses). This suggests that data volume and diversity directly influence angle prediction robustness.

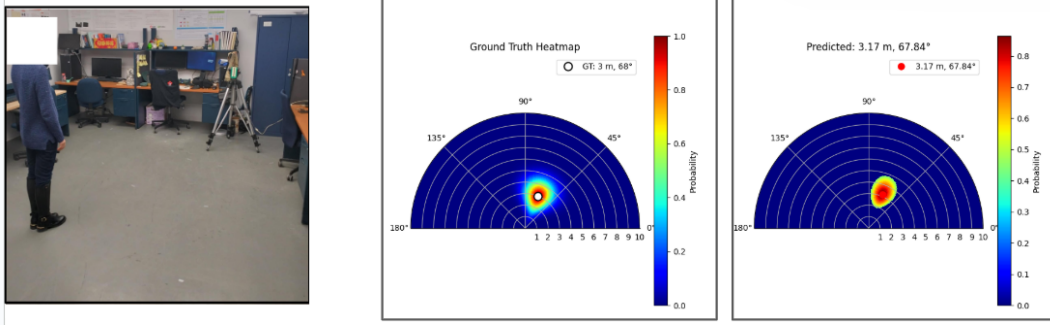
Table 4.3 reports the results for the multi-person scenario : Some limitations appear as simultaneous detections degrade distance accuracy (MAEs rise to 1.228 and 1.336 for Persons 1 and 2, respectively, 3–4× higher than in single-person cases) and degrades angle estimation (MAEs : 27.104° and 64.318°, with Stds exceeding 23°). This can be explained by the more complex scenario that may need more data for training. Example results including individual detection and simultaneous detection of two persons in scenarios are shown in the figure 4.7.

TABLEAU 4.2 MAE and Standard Deviation for Distance (m) and Angle (degrees) with person-based training on persons 1 (P1), 2 (P2), and 3 (P3), and separate testing per person.

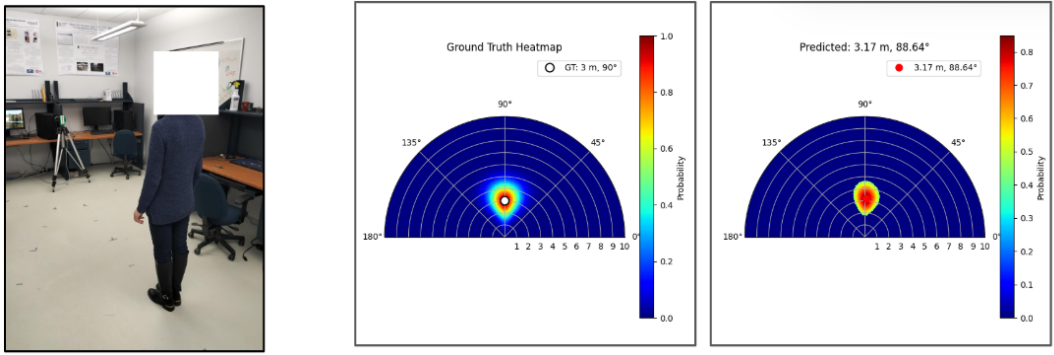
Train	Test	Distance (m)		Angle (degrees)	
		MAE	Std	MAE	Std
P1, P2, P3	P1	0.405	0.394	2.036	1.899
P1, P2, P3	P2	0.521	0.453	2.335	2.351
P1, P2, P3	P3	0.238	0.171	2.668	1.499
P1, P2, P3	Overall	0.423	0.401	2.168	1.989

Inter-Person Testing

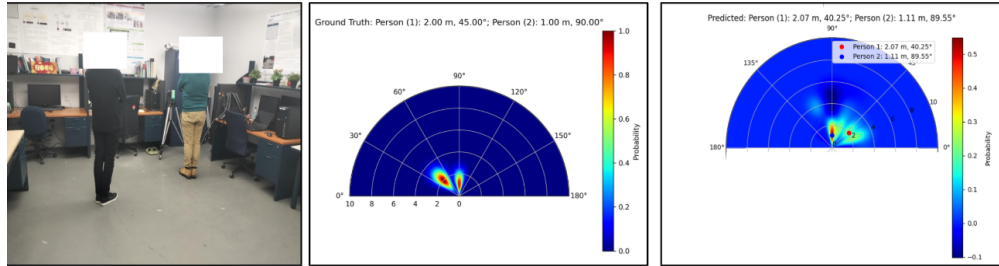
As shown in Table 4.4, our model shows varying performance when tested on individuals excluded from training. Distance estimation remains relatively stable (MAE : 0.864–1.115 m), with Person 3 achieving the lowest error (0.864 m). However, angle estimation varies across subjects, with MAEs ranging from 19.895° (Person 3) to 30.977° (Person 1), accompanied by high standard deviations (e.g., 20.866° for Person 3). Results with Person 3 show lower errors.



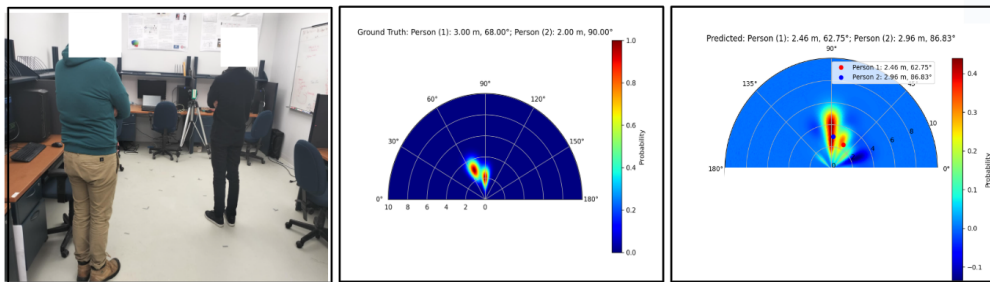
(a) Experimental evaluation in Lab1 : Comparison of ground truth at (3.00 m, 68°) and model output (3.17 m, 67.84°) using heatmaps.



(b) Experimental evaluation in Lab1 : Comparison of ground truth at (3.00 m, 90°) and model output (3.17 m, 88.64°) using heatmaps.



(c) Experimental evaluation in Lab1 : Comparison of ground truth at (2.00m, 45.00°) and (1.00m, 90.00°), respectively. and model output (2.07m, 40.25°) and (1.11m, 89.55°) using heatmaps.



(d) Experimental evaluation in Lab2 : Comparison of ground truth at (3.00m, 68.00°) and (2.00m, 90.00°), respectively. and model output (2.46m, 62.75°) and (2.96m, 86.83°) using heatmaps.

FIGURE 4.7 Ground truth vs. predicted location : The left image shows an image of the scene, while the middle and right heatmaps compare the actual and predicted positions.

TABLEAU 4.3 MAE and Std for Distance (m) and Angle ($^{\circ}$). Model trained on persons 1 (P1), 2 (P2), and 3 (P3) (including both P1 and P2 simultaneously (P1+2)). Testing involves pairwise predictions : errors are reported for one person while the other is present.

Training Data	Test Case	Distance (m)		Angle ($^{\circ}$)	
		MAE	Std	MAE	Std
P1, P2, P3, P1+2	P1 (with P2 present)	1.228	1.072	27.104	23.164
P1, P2, P3, P1+2	P2 (with P1 present)	1.336	1.107	64.318	27.526

Specifically, there seems to be a direct link between the quantity of available data and our model ability to generalize. These results emphasize the need for larger datasets with more diversity, like richer variations in posture, to improve robustness for real-world deployment.

TABLEAU 4.4 MAE and Std for Distance (m) and Angle ($^{\circ}$) under different train-test splits among Persons 1, 2, and 3 (P1,P2,P3).

Train	Test	Distance (m)		Angle ($^{\circ}$)	
		MAE	Std	MAE	Std
P1, P2	P3	0.864	0.470	19.895	20.866
P1, P3	P2	0.894	0.718	26.858	16.440
P2, P3	P1	1.115	0.782	30.977	22.893

Cross-environment testing

Table 4.5 summarizes the model generalization performance across scene changes between Lab1 and Lab2. When trained on Lab1 and tested on Lab2, the model achieves moderate distance prediction performance (MAE = 0.727, Std = 0.574), while angle estimation shows larger errors (MAE = 35.889, Std = 20.041), indicating greater sensitivity to scene variations in angle error metrics. The reverse configuration (train on Lab2 and test on Lab1) exhibits worst performance in both distance (MAE = 1.314, Std = 1.017) and angle predictions (MAE = 37.333, Std = 23.669), with higher standard deviations across metrics. This difference in performance suggests that Lab1 training data is more diverse and representative.

Evaluation of Predictions on Unseen angles or distances

Figure 4.8 illustrates our method performance for seen and unseen distances and angles. Predictions near trained angles (e.g., 34° and 55°) are highly accurate, with minimal distance

TABLEAU 4.5 MAE and Standard Deviation for Distance and Angle with Cross-Lab Training and Testing

Train	Test	Distance		Angle	
		MAE	Std	MAE	Std
Lab1	Lab2	0.727	0.574	35.889	20.041
Lab2	Lab1	1.314	1.017	37.333	23.669

errors (about 0.05 m) and correct quadrant alignment (e.g., front-left), reflecting robust performance in data-rich regions. However, at unfamiliar intermediate angles (e.g., 100°), errors increase (30° angular deviation, 1.2 m distance overestimation), revealing limitations in extrapolating beyond sparse training data. This shows the model dependence on sufficient training coverage for continuous angular predictions.

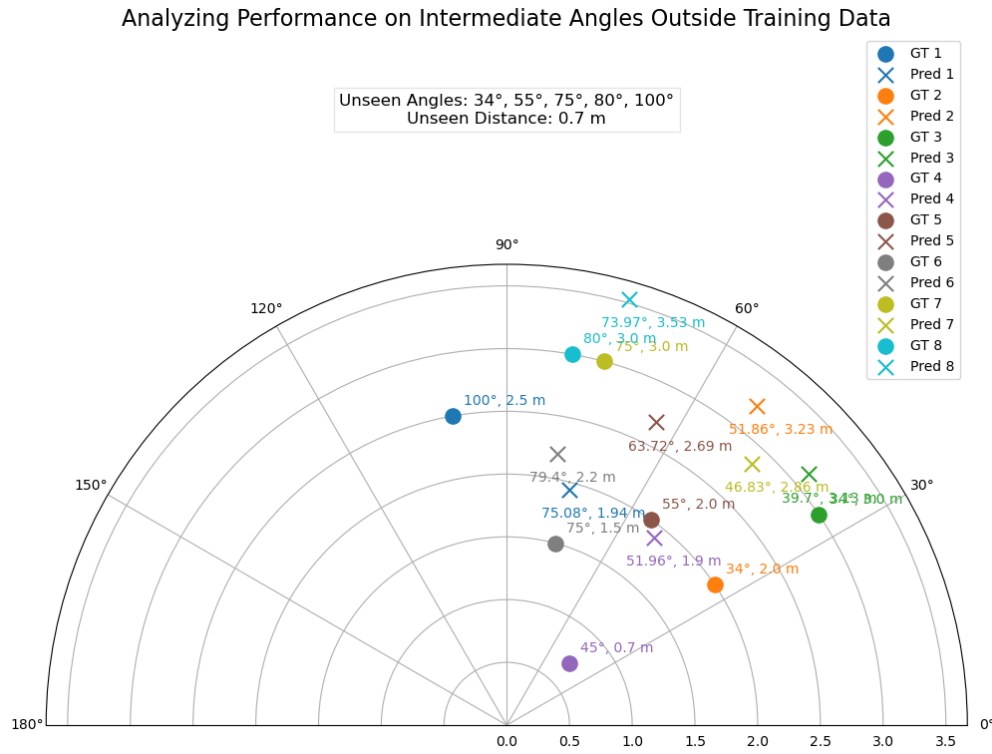


FIGURE 4.8 Polar Plot of Ground Truth vs. Predicted Points for Unseen angles or distances

4.5.6 Comparison with SOTA methods

Distance Estimation : MAE and Standard Deviation of Error Distribution Analysis

In this experiment and the next, for methods requiring training, we use the data split from the first experiment above (trained on combined data from three individuals (Person 1, 2, and 3) from both labs LAB1 and LAB2 and tested separately on each person data). To benchmark against state-of-the-art (SOTA) methods, we compared our approach to : (1) the DNN-based method by Park et al. [67], which classifies distances from 1–10 meters using a feedforward neural network on FFT-processed data, and (2) the conventional FFT method that estimates distance by detecting the maximum peak.

Figure 4.9 (from left to right) illustrates the performance of these three methods. The diagonal red line represents ideal results, with green dashed lines indicating a $\pm 0.5\text{m}$ error margin, and the color of each point shows its absolute error (warmer colors meaning higher error). The FFT-based method (left) exhibits a large scatter around the diagonal, resulting in a mean absolute error (MAE) of 1.67m and a standard deviation (Std) of 1.20m. Here, MAE quantifies how far the predictions deviate from the true distance on average, and Std measures how much the prediction errors vary across samples. A high MAE and Std imply that the method is both inaccurate and inconsistent. Moving to the DNN approach (middle), the results cluster more tightly near the diagonal, reducing the MAE to 0.77m (Std 0.79m); this indicates more accurate and more consistent predictions. Finally, our CNN+Transformer model (right) shows the narrowest clustering around the diagonal and keeps most predictions well within $\pm 0.5\text{m}$, yielding the lowest MAE of 0.423m and a smaller Std of 0.401m. Overall, the progression from FFT to DNN to CNN+Transformer reveals that adding deep feature extraction and attention mechanisms significantly enhances both accuracy (low MAE) and stability (low Std) in distance estimation.

Analysis of Angle Estimation Performance

We compare our method with two conventional methods from the literature for angle estimation. The first is the classical subspace-based method known as MUSIC [76], which estimates the angle of the target by taking use of the signal covariance matrix eigen structure. The second is DeepMUSIC [43], which uses CNN architecture to improve the conventional MUSIC technique. Results versus our proposed CNN+Transformer method are presented in Figure 4.10 (a). They show that MUSIC performs poorly, with errors often exceeding 100 degrees, which confirms its high sensitivity to real-world noise and signal distortions.

Comparison of Distance Estimation Performance

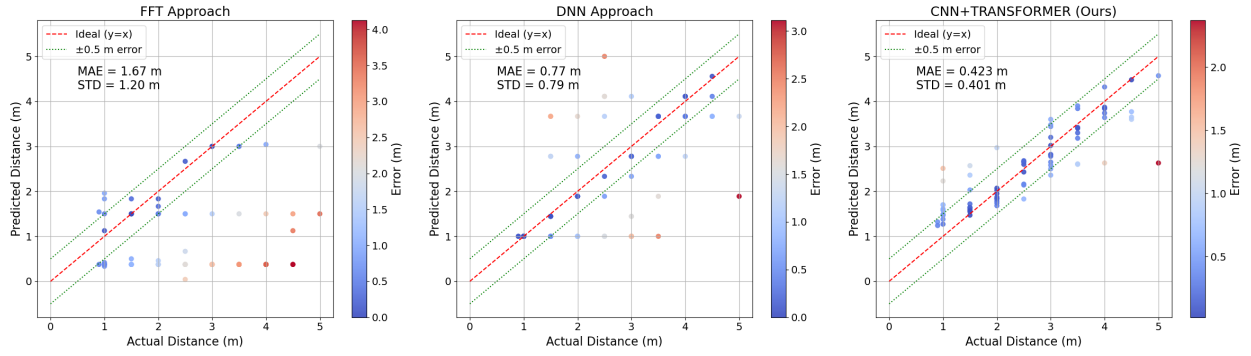


FIGURE 4.9 SOTA comparison : Std and MAE Analysis of the distance Estimation

DeepMUSIC improves this by integrating a CNN method with MUSIC subspace principles in order to reduce errors and confining most predictions within 50 degrees from the ground truth, leading to better accuracy. However, our CNN+Transformer model achieves the best precision, with errors tightly clustered near 0° and a rapid decrease in frequency as the error increases, demonstrating its superior ability to generalize across different scenarios and to suppress noise.

These results are presented in a different way in Figure 4.10 (b) showing similar conclusions. By extracting better spatial features, the CNN+Transformer model surpasses both alternatives, offering more robustness for practical radar-based localization. The CNN efficiently captures spatial information, while the transformer enhances global signal understanding, making our approach an effective solution for radar-based localization in real-world environments.

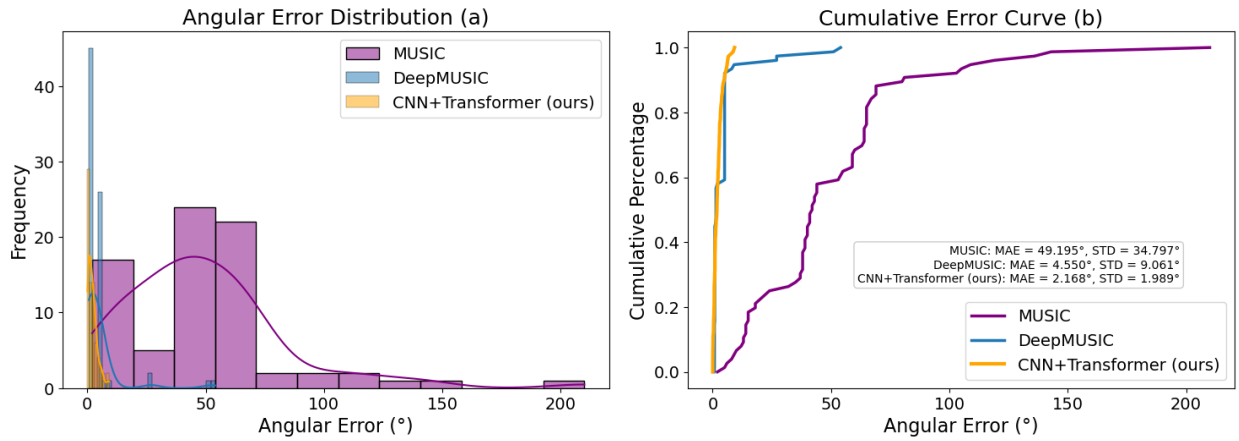


FIGURE 4.10 SOTA comparison : MAE and Std Analysis of the Angle Estimation

4.5.7 Discussion

Although our model shows strong performance in single-person environments, its drop in performance under more complex conditions, such as multi-person interactions or unseen environmental layouts, can be primarily explained by the need for a lot of training data rather than a limitation of the model itself. This means that to use a radar in several different rooms, our model should be trained from data coming from all rooms. This applies also to the possible location the person can occupy in the scene. With more diversified data, the model performance would generalize better. Properly trained, our propose model outperforms significantly a traditional FFT method, because it can better deal with adverse environmental conditions. A traditional FFT method can be applied directly to any scenarios without training, but its performance will be acceptable only in simpler environment.

Cross-environment testing illustrates the importance of diversified training sets, as models exposed to multiple distinct environments consistently outperform those relying on limited data.

The radar hardware itself imposes some constraints : a two-antenna setup limits angular resolution like in our case, hindering the system ability to distinguish closely spaced targets or generalize to unseen angles. Expanding to multi-antenna arrays would enhance spatial diversity, enabling finer angular discrimination through richer geometric signal relationships. Additionally, the system range resolution (for example in our case is 0.75 m, governed by a 200 MHz bandwidth) and accuracy at longer distances are constrained by inherent hardware limitations. While comprehensive training data covering diverse angles and distances is important, optimizing the radar design, through multi-antenna configurations and bandwidth scaling, is equally essential to overcome these resolution bottlenecks.

4.6 Conclusion

This study presents a framework for indoor human localization using raw FMCW radar data, eliminating the dependency on manual feature extraction. By combining CNN-based spatial analysis and Transformer-driven temporal modeling, the hybrid architecture effectively captures spatial-temporal relationships in radar signals in order to enable robust localization across diverse indoor environments. Experimental results demonstrate that this data-driven approach outperforms conventional methods in accuracy, while inherently preserving privacy by avoiding camera-based systems.

Acknowledgment

The authors would like to thank Morphee+ Inc. and Mitacs for their support and contributions to this research.

CHAPITRE 5 DISCUSSION GÉNÉRALE

Dans cette section, nous présentons des résultats expérimentaux additionnels obtenus avec notre méthode de localisation basée sur le radar FMCW. Les évaluations sont conduites selon différents protocoles : tests intra- et inter-personnes, entraînement sur un seul individu ainsi qu’un retour sur les résultats obtenus dans notre article.

5.1 Retour sur les résultats présentés dans l’article

5.1.1 Performances globales et métriques clés

Évaluation intra-personne et équilibrage des données

L’expérience du tableau 4.2 souligne les défis inhérents à la généralisation d’un modèle de localisation lorsqu’il est confronté à des individus différents. Bien que les performances globales restent acceptables, les disparités entre sujets révèlent une sensibilité marquée à la représentativité des données d’entraînement. En particulier, un individu bénéficiant d’un nombre limité d’échantillons lors de l’apprentissage tend à induire des prédictions précises en distance, suggérant un risque de surapprentissage. À l’inverse, les individus associés à des ensembles d’apprentissage plus riches et diversifiés, incluant des configurations multi-personnes, démontrent une meilleure régularité dans les prédictions angulaires. Ces observations renforcent l’idée que la variabilité posturale et contextuelle des données est un déterminant clé de la robustesse.

Généralisation inter-personne : limites et enjeux

Lorsque le modèle est testé sur des individus exclus de l’entraînement comme c’est montré dans le Tableau 4.4, les résultats mettent en évidence une divergence significative entre la précision en distance et celle en angle. Alors que la localisation spatiale conserve une certaine stabilité malgré le changement de sujet, l’estimation angulaire subit une dégradation prononcée, avec des erreurs variant fortement selon la diversité posturale des individus utilisés en entraînement. Ce phénomène souligne que le modèle peine à extrapoler les caractéristiques morphologiques ou comportementales non rencontrées précédemment. Ces limitations soulignent la nécessité d’intégrer des données couvrant un spectre plus large de profils physiques et de mouvements pour améliorer l’adaptabilité à des utilisateurs inconnus.

Impact des variations environnementales sur la robustesse

L’évaluation croisée illustrée dans le tableau 4.5 entre environnements distincts révèle une vulnérabilité du modèle aux changements de contexte spatial. La dégradation des performances,

surtout angulaires, lors du transfert entre laboratoires, indique une dépendance forte aux spécificités du lieu d'apprentissage, telles que la configuration des lieux. Fait notable, l'asymétrie des résultats selon la direction du transfert suggère que la composition des données d'entraînement (par exemple, leur hétérogénéité intrinsèque) joue un rôle prépondérant dans la capacité à généraliser. Ces résultats plaident pour l'enrichissement systématique des jeux de données avec des captures multi-sites, afin de réduire les biais contextuels et d'approcher une invariance aux conditions environnementales.

Enfin, ces analyses mettent en lumière des axes d'amélioration pour les systèmes de localisation basés sur l'apprentissage. Premièrement, l'équilibrage des jeux de données entre individus s'avère essentiel pour limiter les biais de prédiction. Deuxièmement, la généralisation à de nouveaux sujets requiert une représentation exhaustive des variations posturales et morphologiques, dépassant le cadre restreint d'un petit groupe d'entraînement. Enfin, la robustesse environnementale impose une diversification des scènes d'apprentissage. Ces pistes, bien que coûteuses en termes de collecte, sont indispensables pour déployer des modèles fiables dans des contextes réels, où l'hétérogénéité des utilisateurs et des espaces est la norme.

5.2 Résultats additionnels

5.2.1 Variabilité inter-individuelle et biais d'entraînement

Afin d'évaluer la capacité de généralisation du modèle dans un contexte de données restreintes et de variabilité inter-individuelle, nous avons adopté une validation croisée en laissant tour à tour deux personnes de côté. À chaque itération, le modèle est entraîné sur un unique sujet parmi les trois disponibles puis évalué sur les deux restants. Le sujet 3, présent uniquement dans le laboratoire LAB1 et enregistré sur seulement six positions, offre un scénario extrême pour vérifier si le réseau peut prédire une position jamais vue durant l'entraînement. Pour les deux autres sujets, dont les jeux de données couvrent l'ensemble des positions dans les deux laboratoires, nous balayons successivement toutes les positions avec une personne durant la phase de l'entraînement puis testons sur ces mêmes emplacements mais avec un autre individu, afin de quantifier l'impact de la variabilité physiologique sur la précision spatiale pour des positions identiques.

Les résultats du tableau 5.1 indiquent que ce mode d'apprentissage entraîne une perte significative de performance, en particulier pour l'estimation angulaire. Par exemple, lorsque le modèle est entraîné sur la personne 1, l'erreur angulaire moyenne absolue (MAE) dépasse 30 ° pour les deux sujets testés (30,945 ° pour P2 et 33,107 ° pour P3), témoignant d'une capacité de généralisation limitée dans ce cas, malgré des erreurs de distance modérées. À

TABLEAU 5.1 MAE et écart-type pour la distance (m) et l'angle (°) lorsque le modèle est entraîné sur une seule personne (P1, P2 ou P3) et testé sur les deux autres.

Train	Test	Distance (m)		Angle (°)	
		MAE	Écart-type	MAE	Écart-type
P3	P1	1.170	0.798	26.997	23.607
P3	P2	0.993	0.660	25.386	26.527
P2	P1	0.835	0.659	26.051	16.931
P2	P3	0.796	0.589	17.377	16.478
P1	P2	0.955	0.726	30.945	14.155
P1	P3	0.731	0.401	33.107	23.518

l'inverse, l'entraînement sur la personne 2 semble offrir un compromis plus équilibré, avec les meilleures performances globales en distance (MAE = 0.830 m) et des erreurs angulaires moindres (MAE global = 25.025°), notamment sur la personne 3 (17.377°).

Cela suggère que certaines personnes, probablement en raison de la diversité de leurs morphologies, la composition corporelle ou les matériaux des vêtements dans les données, permettent au modèle de mieux abstraire des représentations transférables. L'entraînement sur la personne 3, quant à lui, engendre la pire performance globale en distance (MAE = 1.106 m) et des écarts angulaires très élevés, illustrant qu'un jeu d'apprentissage avec peu de données ou peu représentatif compromet lourdement la robustesse du modèle. Dans l'ensemble, ces résultats confirment que l'apprentissage sur un seul sujet limite fortement la généralisation, en particulier pour les angles, et soulignent la nécessité de disposer de données variées et multi-individuelles pour une application fiable en conditions réelles.

5.3 Perspectives futures : vers une détection robuste des chutes par radar FMCW

Pour étendre la localisation radar FMCW à la reconnaissance d'activité, et plus particulièrement à la détection fiable des chutes, il est impératif de constituer un jeu de données riche et rigoureusement annoté : (1) diversité des sujets : recruter des participants de tout âge, de différentes morphologies et capacités motrices. (2) variété des activités : enregistrer de vraies chutes (avant, arrière, latérales), des quasi-chutes, des transitions assis-debout, de la marche, des étirements et d'autres gestes brusques courants dans la vie quotidienne, afin d'équilibrer le nombre d'exemples par classe. (3) variété des artefacts : capter les données dans des pièces meublées, des couloirs, des escaliers et des espaces encombrés, avec différentes surfaces

réfléchissantes (murs, vitres, rideaux) et configurations de mobilier. (4) environnements domestiques variés : agréger des données provenant de plusieurs laboratoires ou domiciles pour garantir la robustesse aux différences de géométrie et de matériau, et publier des métadonnées standardisées (positions du radar, dimensions de la pièce, paramètres de balayage) pour faciliter la réutilisation. En structurant ainsi le corpus, il suffira ensuite de remplacer la sortie du modèle par une tête de classification d'activité et de procéder à l'entraînement supervisé pour obtenir une détection de chutes précise et généralisable.

CHAPITRE 6 CONCLUSION

Ce travail constitue une première étape vers le développement d'un système de localisation respectueux de la vie privée, destiné à améliorer la surveillance des personnes âgées vivant seules. Notre approche, fondée sur un cadre d'apprentissage profond traitant directement les signaux bruts du radar FMCW, intègre un encodeur double-branche Phase-Amplitude couplé à un module Transformeur par antenne. Cette architecture permet de capter les signatures humaines dans des environnements encombrés tout en préservant les informations spatio-temporelles importantes. Validée expérimentalement dans deux salles de laboratoires différentes, la méthode démontre des performances prometteuses en précision de localisation et en robustesse face des artéfacts environnementaux réalistes, notamment les réflexions parasites et les occlusions partielles.

Cependant, nos expérimentations révèlent des défis persistants. La généralisation du modèle reste tributaire de la diversité des sujets et des configurations spatiales testées lors de l'entraînement, soulignant la nécessité d'enrichir les jeux de données avec des profils morphologiques variés et des environnements domestiques réalistes. Par ailleurs, les limitations physiques du radar utilisé, notamment sa résolution angulaire limitée et son nombre réduit d'antennes, contraignent la discrimination fine des cibles dans des scènes densément encombrées. Une piste prometteuse consisterait à coupler ces algorithmes avec des radars nouvelle génération offrant une densité spatiale accrue et une sensibilité améliorée.

En exploitant les propriétés non intrusives du radar, contrairement aux systèmes caméra, notre approche esquisse les bases d'une surveillance continue préservant la dignité des utilisateurs. Ces avancées ouvrent la voie à une nouvelle génération de technologies assistives où l'intelligence artificielle, ancrée dans des principes de conception responsables, pourrait accompagner les transitions démographiques tout en honorant les choix de vie individuels.

Bien que sur le plan méthodologique l'exploitation directe des signaux bruts FMCW sans prétraitement manuel (FFT, filtrage..) constitue une approche originale, les résultats obtenus ouvrent la voie à plusieurs axes de recherche prometteurs pour le développement de systèmes radar intelligents et respectueux de la vie privée en environnements complexes : d'une part, l'augmentation du jeu de données est nécessaire, avec des scénarios plus variés comprenant plusieurs personnes, des objets mobiles et différents environnements. D'autre part, les limites du radar utilisé pourraient être compensées par l'utilisation de radars à antennes multiples ou multi-statiques, afin d'améliorer la précision de détection et la couverture de la scène.

RÉFÉRENCES

- [1] J. Zhang, Z. Zhang, Y. Li, C. Gu et J. Mao, “Target detection with short-range fmcw radar based on time–frequency characteristic-based leakage cancellation technique,” *IEEE Transactions on Microwave Theory and Techniques*, 2024.
- [2] Y. Li, Q. Long, Z. Wu et Z. Zhou, “Low-complexity joint 3d super-resolution estimation of range velocity and angle of multi-targets based on fmcw radar,” *Sensors*, vol. 22, n^o. 17, p. 6474, 2022.
- [3] Y.-S. Chen, K.-H. Cheng, Y.-A. Xu et T.-Y. Juang, “Multi-feature transformer-based learning for continuous human motion recognition with high similarity using mmwave fmcw radar,” *Sensors*, vol. 22, n^o. 21, p. 8409, 2022.
- [4] J. Seo, Y. Yang, Y.-g. Hong et J. Park, “Transfer learning-based radar imaging with deep convolutional neural networks for distributed frequency modulated continuous waveform multiple-input multiple-output radars,” *IET Radar, Sonar & Navigation*, vol. 15, n^o. 10, p. 1209–1220, 2021.
- [5] Y. Xu, X. Wang, J. Kupiainen, J. Sæe, J. Boutellier, J. Nurmi et B. Tan, “Tracking the occluded indoor target with scattered millimeter wave signal,” *IEEE Sensors Journal*, 2024.
- [6] M. Zhao, S. Chen, Y. Zuo, H. Xie et R. Liu, “Leakage and background clutters suppression for human detection indoors in fmcw radar with digital frequency domain technology,” *Available at SSRN 4693608*.
- [7] C. Will, P. Vaishnav, A. Chakraborty et A. Santra, “Human target detection, tracking, and classification using 24-ghz fmcw radar,” *IEEE Sensors Journal*, vol. 19, n^o. 17, p. 7283–7299, 2019.
- [8] J.-K. Park, J.-H. Park et K.-T. Kim, “Multipath signal mitigation for indoor localization based on mimo fmcw radar system,” *IEEE internet of things journal*, vol. 11, n^o. 2, p. 2618–2629, 2023.
- [9] S. Yang et Y. Kim, “Single 24-ghz fmcw radar-based indoor device-free human localization and posture sensing with cnn,” *IEEE Sensors Journal*, vol. 23, n^o. 3, p. 3059–3068, 2022.
- [10] M. E. Yanik et S. Rao, “Radar-based multiple target classification in complex environments using 1d-cnn models,” dans *2023 IEEE Radar Conference (RadarConf23)*. IEEE, 2023, p. 1–6.

- [11] B. Friedlander et A. J. Weiss, "Direction finding in the presence of mutual coupling," *IEEE transactions on antennas and propagation*, vol. 39, n^o. 3, p. 273–284, 1991.
- [12] H. Krim et M. Viberg, "Two decades of array signal processing research : the parametric approach," *IEEE signal processing magazine*, vol. 13, n^o. 4, p. 67–94, 1996.
- [13] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, n^o. 8, p. 1408–1418, 1969.
- [14] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, n^o. 3, p. 276–280, 1986.
- [15] S. Saponara et B. Neri, "Radar sensor signal acquisition and multidimensional fft processing for surveillance applications in transport systems," *IEEE Transactions on Instrumentation and Measurement*, vol. 66, n^o. 4, p. 604–615, 2017.
- [16] B.-s. Kim, Y. Jin, J. Lee et S. Kim, "High-efficiency super-resolution fmcw radar algorithm based on fft estimation," *Sensors*, vol. 21, n^o. 12, p. 4018, 2021.
- [17] J.-H. Choi, J.-H. Jang et J.-E. Roh, "Design of an fmcw radar altimeter for wide-range and low measurement error," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, n^o. 12, p. 3517–3525, 2015.
- [18] S. Scherr, S. Ayhan, B. Fischbach, A. Bhutani, M. Pauli et T. Zwick, "An efficient frequency and phase estimation algorithm with crb performance for fmcw radar applications," *IEEE Transactions on Instrumentation and Measurement*, vol. 64, n^o. 7, p. 1868–1875, 2014.
- [19] W. Zhang, H. Li, G. Sun et Z. He, "Enhanced detection of doppler-spread targets for fmcw radar," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, n^o. 4, p. 2066–2078, 2019.
- [20] A. Jalil, H. Yousaf et M. I. Baig, "Analysis of cfar techniques," dans *2016 13th International Bhurban Conference on Applied Sciences and Technology (IBCAST)*, 2016, p. 654–659.
- [21] V. G. Hansen, "Constant false alarm rate processing in search radars," 1973. [En ligne]. Disponible : <https://api.semanticscholar.org/CorpusID:63449050>
- [22] G. Trunk, "Range resolution of targets using automatic detectors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-14, n^o. 5, p. 750–755, 1978.
- [23] M. Smith et P. Varshney, "Intelligent cfar processor based on data variability," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 36, n^o. 3, p. 837–847, 2000.
- [24] R. Nitzberg, "Clutter map cfar analysis," *IEEE Transactions on Aerospace and Electronic Systems*, vol. AES-22, n^o. 4, p. 419–421, 1986.

- [25] D. Wang, F. Fioranelli et A. Yarovoy, “Analysis of processing pipelines for indoor human tracking using fmcw radar,” dans *2024 IEEE Radar Conference (RadarConf24)*. IEEE, 2024, p. 1–6.
- [26] M. Jiang, S. Guo, H. Luo, Y. Yao et G. Cui, “A robust target tracking method for crowded indoor environments using mmwave radar,” *Remote Sensing*, vol. 15, n°. 9, p. 2425, 2023.
- [27] Q. Zheng, L. Yang, Y. Xie, J. Li, T. Hu, J. Zhu, C. Song et Z. Xu, “A target detection scheme with decreased complexity and enhanced performance for range-doppler fmcw radar,” *IEEE Transactions on Instrumentation and Measurement*, vol. 70, p. 1–13, 2020.
- [28] T. Vu Hop, T. Cao Quyen et N. Van Loi, “A design of a high-resolution frequency modulated continuous wave radar for drone detection based on spurious phase noise and discrete clutter reduction,” *IET Radar, Sonar & Navigation*, vol. 18, n°. 5, p. 704–714, 2024.
- [29] D. Xiao, W. Liu, H. Chen, H. Li et B. Li, “An adaptive radar target detection method based on alternate estimation in power heterogeneous clutter,” *Remote Sensing*, vol. 16, n°. 13, p. 2508, 2024.
- [30] C. Decourt, R. VanRullen, D. Salle et T. Oberlin, “Darod : A deep automotive radar object detector on range-doppler maps,” dans *2022 IEEE Intelligent Vehicles Symposium (IV)*. IEEE, 2022, p. 112–118.
- [31] C. Ding, L. Zhang, H. Chen, H. Hong, X. Zhu et C. Li, “Human motion recognition with spatial-temporal-convlstm network using dynamic range-doppler frames based on portable fmcw radar,” *IEEE Transactions on Microwave Theory and Techniques*, vol. 70, n°. 11, p. 5029–5038, 2022.
- [32] S. M. Kahya, M. S. Yavuz et E. Steinbach, “Hood : Real-time human presence and out-of-distribution detection using fmcw radar,” *IEEE Transactions on Radar Systems*, vol. 3, p. 44–56, 2025.
- [33] R. Schmidt, “Multiple emitter location and signal parameter estimation,” *IEEE Transactions on Antennas and Propagation*, vol. 34, n°. 3, p. 276–280, 1986.
- [34] —, “Multiple emitter location and signal parameter estimation,” *IEEE Transactions on Antennas and Propagation*, 1986.
- [35] C. Hsieh et T. Tsai, “Harmonic music method for mmwave radar-based vital sign estimation,” *arXiv preprint arXiv :2408.01951*, 2024. [En ligne]. Disponible : <https://arxiv.org/abs/2408.01951>
- [36] J. Doe et A. Smith, “Inverse czft for subsurface biomarker detection,” US Department of Energy, Rapport technique OSTI-10110067, 2015. [En ligne]. Disponible :

<https://www.osti.gov/servlets/purl/10110067>

- [37] A. Smith et B. Lee, “Real-time fall detection using embedded radar signal processing,” *IEEE Transactions on Biomedical Circuits and Systems*, 2022.
- [38] R. Roy et T. Kailath, “Esprit-estimation of signal parameters via rotational invariance techniques,” *IEEE Transactions on acoustics, speech, and signal processing*, vol. 37, n^o. 7, p. 984–995, 1989.
- [39] H. Hwang, A. Zakeriya, G. Marshall et A. Yakovlev, “Direction of arrival estimation using a root-music algorithm,” *Lecture Notes in Engineering and Computer Science*, vol. 2169, 03 2008.
- [40] W. Wei, R. Liu, X. Yu, Q. Lu, G. Cui, X. Fang et L. Zhang, “Doa estimation of distributed mmwave radar system via fast iterative adaptive approach,” dans *2021 International Conference on Control, Automation and Information Sciences (ICCAIS)*, Oct 2021, p. 414–418.
- [41] J. Li, P. Stoica et Z. Wang, “On robust capon beamforming and diagonal loading,” *IEEE Transactions on Signal Processing*, vol. 51, n^o. 7, p. 1702–1715, July 2003.
- [42] S. Ahmed, J. Park et S. H. Cho, “Effects of receiver beamforming for vital sign measurements using fmcw radar at various distances and angles,” *Sensors*, vol. 22, n^o. 18, p. 6877, 2022.
- [43] A. M. Elbir, “Deepmusic : Multiple signal classification via deep learning,” *IEEE Sensors Letters*, vol. 4, n^o. 4, p. 1–4, 2020.
- [44] P. Meiresone, D. Van Hamme et W. Philips, “Segmentation of range-azimuth maps of fmcw radars with a deep convolutional neural network,” dans *International Conference on Advanced Concepts for Intelligent Vision Systems*. Springer, 2023, p. 136–147.
- [45] Y. Wang, Z. Jiang, X. Gao, J.-N. Hwang, G. Xing et H. Liu, “Rodnet : Radar object detection using cross-modal supervision,” dans *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, January 2021, p. 504–513.
- [46] X. Wen, Y. Li et Y. Wang, “Mmw radar confmap abnormality detection based on generative adversarial network,” dans *Proceedings of the 2023 7th International Conference on Electronic Information Technology and Computer Engineering*, 2023, p. 831–835.
- [47] Y. Dalbah, J. Lahoud et H. Cholakkal, “Transradar : Adaptive-directional transformer for real-time multi-view radar semantic segmentation,” dans *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2024, p. 353–362.
- [48] B. Yang, I. Khatri, M. Happold et C. Chen, “Adcnet : learning from raw radar data via distillation,” *arXiv preprint arXiv :2303.11420*, 2023.

- [49] J. Giroux, M. Bouchard et R. Laganier, “T-fftradnet : Object detection with swin vision transformers from raw adc radar signals,” dans *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, p. 4030–4039.
- [50] K. Kaiser, J. Daugalas, J. López-Randulfe, A. Knoll, R. Weigel et F. Lurz, “Complex-valued neural networks for millimeter wave fmcw-radar angle estimations,” dans *2022 19th European Radar Conference (EuRAD)*, 2022, p. 145–148.
- [51] B. Nguyen, “FMCW radar dataset,” 5 2024. [En ligne]. Disponible : https://figshare.com/articles/dataset/FMCW_radar_dataset/25874515
- [52] S. M. Kahya, M. S. Yavuz et E. Steinbach, “Hood : Real-time human presence and out-of-distribution detection using fmcw radar,” *IEEE Transactions on Radar Systems*, 2024.
- [53] Y. Zhou, “Awesome radar perception : A curated list of radar datasets, detection, tracking and fusion,” GitHub repository, 2022, <https://github.com/ZHOUYI1023/awesome-radar-perception>.
- [54] P. Roy et C. Chowdhury, “A survey on ubiquitous wifi-based indoor localization system for smartphone users from implementation perspectives,” *CCF Transactions on Pervasive Computing and Interaction*, vol. 4, n^o. 3, p. 298–318, 2022.
- [55] M. Z. Khan, M. Farooq, A. Taha, A. Qayyum, F. Alqahtani, A. N. Al Hassan, K. Arshad, K. Assaleh, S. Ansari, M. Usman *et al.*, “Indoor localization technologies for activity-assisted living : opportunities, challenges, and future directions,” *Advances in Computers*, vol. 133, p. 59–98, 2024.
- [56] K. Navaneetham et D. Arunachalam, “Global population aging, 1950–2050,” dans *Handbook of Aging, Health, and Public Policy*. Springer, 2023.
- [57] Q. Xu, X. Ou et J. Li, “The risk of falls among the aging population : A systematic review and meta-analysis,” *Frontiers in public health*, vol. 10, p. 902599, 2022.
- [58] G. Zhang, D. Zhang, Y. He, J. Chen, F. Zhou et Y. Chen, “Multi-person passive wifi indoor localization with intelligent reflecting surface,” *IEEE Transactions on Wireless Communications*, vol. 22, n^o. 10, p. 6534–6546, 2023.
- [59] R. Han, W. Feng, F. Wang, Z. Qian, H. Yan et S. Wang, “Benchmarking the complementary-view multi-human association and tracking,” *International Journal of Computer Vision*, vol. 132, n^o. 1, p. 118–136, 2024.
- [60] Y. Jiang, T. Gong, L. He, S. Yan, X. Wu et J. Liu, “Fall detection on embedded platform using infrared array sensor for healthcare applications,” *Neural Computing and Applications*, vol. 36, n^o. 9, p. 5093–5108, 2024.

- [61] K. Park, J. Lee et Y. Kim, “Deep learning-based indoor two-dimensional localization scheme using a frequency-modulated continuous wave radar,” *Electronics*, vol. 10, n^o. 17, p. 2166, 2021.
- [62] Z. Benyahia, M. Hefnawi, M. Aboufatah, H. Abdelmounim et T. Gadi, “A two-stage support vector machine and squeezenet system for range-angle and range-speed estimation in a cluttered environment of automotive mimo radar systems,” dans *ITM Web of Conferences*, vol. 48. EDP Sciences, 2022, p. 01010.
- [63] Q. Xin, Z. Jiang, P. Cheng et M. He, “Signal processing for digital beamforming fmcw sar,” *Mathematical Problems in Engineering*, vol. 2014, n^o. 1, p. 859890, 2014.
- [64] B. Ślesicki, A. Ślesicka et A. Kawalec, “Analysis of the accuracy of the estimation of signal arrival angle in monostatic mimo radar using the capon algorithm and its modifications,” dans *2022 23rd International Radar Symposium (IRS)*. IEEE, 2022, p. 283–287.
- [65] Y. Wang, A. Ren, M. Zhou, W. Wang et X. Yang, “A novel detection and recognition method for continuous hand gesture using fmcw radar,” *IEEE Access*, vol. 8, p. 167 264–167 275, 2020.
- [66] S. Lim, J. Jung, B.-h. Lee, J. Choi et S.-C. Kim, “Radar sensor-based estimation of vehicle orientation for autonomous driving,” *IEEE Sensors Journal*, vol. 22, n^o. 22, p. 21 924–21 932, 2022.
- [67] K.-E. Park, J.-P. Lee et Y. Kim, “Deep learning-based indoor distance estimation scheme using fmcw radar,” *Information*, vol. 12, n^o. 2, 2021. [En ligne]. Disponible : <https://www.mdpi.com/2078-2489/12/2/80>
- [68] S. Yang et Y. Kim, “Single 24-ghz fmcw radar-based indoor device-free human localization and posture sensing with cnn,” *IEEE Sensors Journal*, vol. 23, n^o. 3, p. 3059–3068, 2023.
- [69] B.-s. Kim, Y. Jin, J. Lee et S. Kim, “Low-complexity music-based direction-of-arrival detection algorithm for frequency-modulated continuous-wave vital radar,” *Sensors*, vol. 20, n^o. 15, p. 4295, 2020.
- [70] J. Xiong, H. Hong, L. Xiao, E. Wang et 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, n^o. 1, p. 83–92, 2023.
- [71] T.-H. Sang, F.-T. Chien, C.-C. Chang, K.-Y. Tseng, B.-S. Wang et J.-I. Guo, “Doa estimation for fmcw radar by 3d-cnn,” *Sensors*, vol. 21, n^o. 16, p. 5319, 2021.
- [72] K. Han et S. Hong, “Range-angle decoupling technique using wavelength-dependent beamforming for high-resolution mimo radar,” *IEEE Transactions on Microwave Theory*

and Techniques, 2023.

- [73] Z. Xing, P. Chen, J. Wang, Y. Bai, J. Song et L. Tian, “Millimeter-wave radar detection and localization of a human in indoor complex environments,” *Remote Sensing*, vol. 16, n° 14, 2024. [En ligne]. Disponible : <https://www.mdpi.com/2072-4292/16/14/2572>
- [74] X. Peng, M. Zhang, L. Servadei et R. Wille, “Enhancing indoor radar detection : An fmcw radar system for distinguishing human presence and swinging blinds,” dans *2024 IEEE Radar Conference (RadarConf24)*. IEEE, 2024, p. 1–6.
- [75] S. L. Borison, S. B. Bowling et K. M. Cuomo, “Super-resolution methods for wideband radar,” *Lincoln Laboratory Journal*, vol. 5, n° 3, p. 441–461, 1992.
- [76] F. Belfiori, W. van Rossum et P. Hoogeboom, “Application of 2d music algorithm to range-azimuth fmcw radar data,” dans *2012 9th European Radar Conference*. IEEE, 2012, p. 242–245.