Capturing Head Poses Using FMCW Radar and Deep Neural Networks

Kumchaiseemak, Nakorn; Fioranelli, Francesco; Wilaiprasitporn, Theerawit

DOI
10.1109/TAES.2025.3529412

Publication date
2025

Document Version
Final published version

Published in
IEEE Transactions on Aerospace and Electronic Systems

Capturing Head Poses Using FMCW Radar and Deep Neural Networks

NAKORN KUMCHAISEEMAK, Graduate Student Member, IEEE
Vidyasirimedhi Institute of Science and Technology, Rayong, Thailand
Delft University of Technology (TU Delft), Delft, The Netherlands

FRANCESCO FIORANELLI, Senior Member, IEEE
Delft University of Technology (TU Delft), Delft, The Netherlands

THEERAWIT WILAIPRASITPORN, Senior Member, IEEE
Vidyasirimedhi Institute of Science and Technology, Rayong, Thailand

This article presents the first subject-specific head pose estimation approach using only one frequency-modulated continuous wave radar data frame. Specifically, the proposed method incorporates a deep learning framework to estimate head pose rotation and orientation frame-by-frame by combining a convolutional neural network operating on range-angle radar plots and a PeakConv network. The proposed method is validated with an in-house collected dataset, including annotated head movements that varied in roll, pitch, and yaw, and these were recorded in two different indoor environments. It is shown that the proposed model can estimate head poses with a relatively small error of approximately 6.7°–14.4° for all rotational axes and is capable of generalizing to unseen, new environments when trained in one scenario (e.g., lab) and tested in another (e.g., office), including in the cabin of a car.

Received 28 June 2024; revised 18 December 2024; accepted 8 January 2025. Date of publication 13 January 2025; date of current version 11 June 2025.

DOI. No. 10.1109/TAES.2025.3529412

Refereeing of this contribution was handled by C. Clemente.

The work of Francesco Fioranelli was supported by the Dutch Research Council NWO under KLEIN *RAD-ART* project.

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by TU Delft HREC committee.

Authors' addresses: Nakorn Kumchaiseemak is with the School of Information Science and Technology, Vidyasirimedhi Institute of Science and Technology, Rayong 21210, Thailand, and also with the Microwave Sensing, Signals and Systems (MS3) Group, Department of Microelectronics, Delft University of Technology (TU Delft), 2628 CD Delft, The Netherlands, E-mail: (N.Kumchaiseemak@tudelft.nl, nakorn.k_s18@vistec.ac.th); Francesco Fioranelli is with the Microwave Sensing, Signals and Systems (MS3) Group, Department of Microelectronics, Delft University of Technology (TU Delft), 2628 CD Delft, The Netherlands, E-mail: (F.Fioranelli@tudelft.nl); Theerawit Wilaiprasitporn is with the School of Information Science and Technology, Vidyasirimedhi Institute of Science and Technology, Rayong 21210, Thailand, E-mail: (theerawit.w@vistec.ac.th). *(Corresponding authors: Francesco Fioranelli; Theerawit Wilaiprasitporn.)*

0018-9251 © 2025 IEEE

I. INTRODUCTION

Monitoring human movements and activities through nonvision-based sensors plays a crucial role in various applications [1], including medical imaging [2] and automotive [3], [4]. Its significance lies in the enhanced privacy offered by sensors such as radar, lidar, or thread-based devices [5], in contrast to vision-based sensors such as cameras. Specifically, radar sensors excel in providing privacy and perception, including in scenarios with optically obscured paths. This makes them highly valuable for tasks such as human detection through opaque materials [6] or the estimation of human body motions [7], [8].

In applying body pose estimation using radar, the literature has made rapid progress. For instance, recent work has shown some results by utilizing deep learning (DL) techniques to map radar signals to body skeletons, guided by camera vision [9]. Following this approach, a series of studies have emerged, demonstrating techniques to enhance the estimation accuracy of the posture of the different body parts in the skeleton. For example, recent research has proposed using temporal information and attention mechanisms to improve body posture estimation with radar data [10], [11], [12]. Other research works have shown that using point cloud representations derived from raw radar signals with DL techniques can effectively estimate pose while keeping computational costs manageable [13], [14], [15]. Another key challenge in using DL techniques for body pose estimation is achieving generalization capability. Recent developments suggest this is to some extent possible [16], [17].

However, a tradeoff exists, as none of these techniques can provide exceptionally high spatial resolution. This is because the labels are derived from skeleton poses in 3-D space and are limited to the positions of key points without considering rotational angles, posing challenges in precisely gauging motion within predefined target areas—this includes, for example, using radar to estimate the precise location of the human hands or orientation of the head, i.e., where the person is looking at or in other words the angular orientation of the head.

In contrast to typical body pose recognition, we are specifically interested in capturing head movement. As found in the literature, most previous works have focused on capturing the head movement during driving tasks, such as in the cabins of vehicles [18], [19], [20], aiming to enhance our understanding of human behavior and monitor attentiveness and mental focus. Recent efforts have shown the incorporation of deep neural networks or DL to improve the performance of the earlier works [21], [22]. However, most of these studies remain application-centric, primarily on classification such as head movement direction and specific types of head pose, while leaving a regression task with adaptability and generalization abilities not concretely explored until nowadays. This year's latest work reports

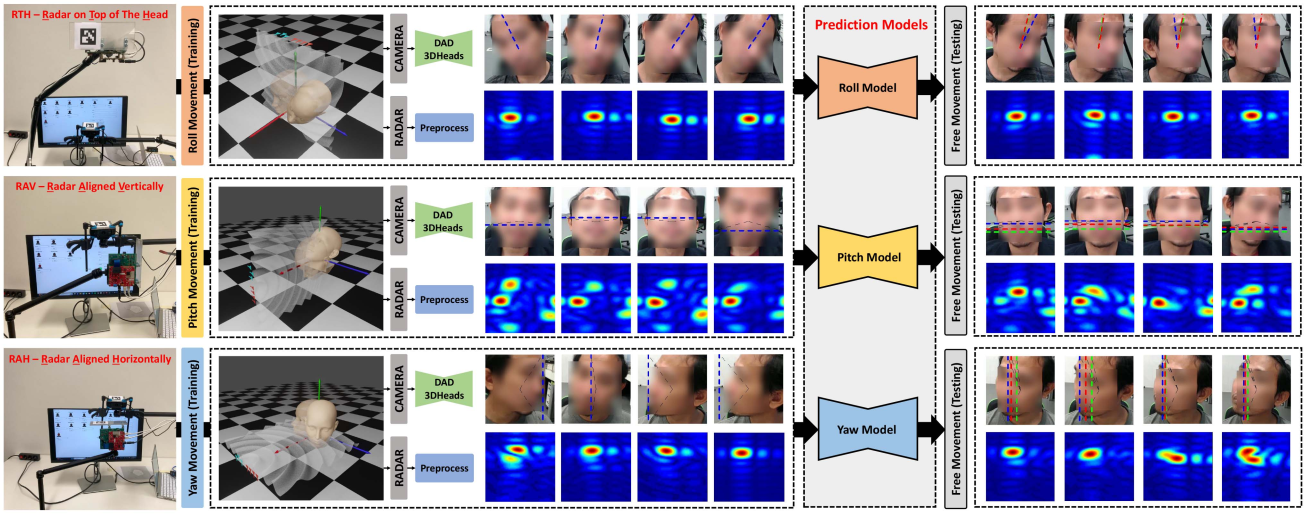


Fig. 1. Overview of our proposed framework for radar-based head pose estimation. The framework consists of a data collection process, a preprocessing step to acquire training data and corresponding ground-truth head poses, and the prediction model to predict head pose and rotations.

the algorithm for ultra-wideband radar used to estimate the pitch and yaw of the head in the car scene [23]. These motivated us to narrow down and conduct the proposed studies in this research direction, especially regression tasks or the estimation of head movement.

This article presents a feasibility study on estimating human head pose rotation and orientation based on a dedicated DL framework using frequency-modulated continuous wave (FMCW) radar. The proposed approach can work on just a single frame of radar data and provide subject-specific head pose estimations (i.e., training one specific model for individual subjects). Unlike focusing on human head classification, our approach is formulated as a regression task, predicting head motion rotation values parameterized by the three variables of roll, pitch, and yaw. In addition, we demonstrate the capability of our approach to generalize well to different environments by training the model in one scene and testing it in another completely different environment.

A specific dataset was collected with a 77-GHz FMCW radar to validate the proposed approach, and vision-based techniques were used to label the data. Precisely, to predict the head pose rotation, we propose to leverage the range-azimuth convolutional neural network (RA-CNN) model based on a fully convolutional network (FCN) [24], demonstrating its potential in localizing tiny targets in FMCW radar data within a limited dataset. In addition, we incorporate the response different aware PeakConv (ReDA-PKC) network [25] before the RA-CNN network. The kernel convolution of ReDA-PKC is designed to achieve a continuous false alarm rate (CFAR), aiding RA-CNN in focusing on meaningful features. Combining these two networks demonstrates the effectiveness of our proposed method in predicting subject-specific head pose rotations.

To summarize, an overview of our study is depicted in Fig. 1, and the main contributions are outlined as follows.

1) To the best of our knowledge, this is the first work to estimate subject-specific head pose rotation with only FMCW radar data utilizing a DL regression approach. It should be noted that this is a more challenging estimation problem than simply locating or tracking the position of the head in the skeleton [13], [15], as the angular orientation of the head is regressed in the three dimensions of yaw, pitch, and roll.
2) The proposed approach is comprehensively validated with an experimental dataset with head rotation information, comprising over 198 000 frames of movement and ten participants recorded in two different environments.

The rest of this article is organized as follows. Section II elaborates on the proposed methodology. Following this, Section III outlines the experimental setup and data collection process. Next, in Sections IV and V, we evaluate our models' performance under various settings, compare them to alternative models inspired by the literature, and discuss the limitations of our work. We show an example of our model's usage in a vehicle's cabin in Section V-C. Finally, Section VI concludes this article.

II. Proposed Methodology

We aim to predict subject-specific head pose rotations using a DL regression model based on a combination of ReDA-PKC and RA-CNN architectures. This section provides details of the proposed approach. First, in Sections II-A and II-B, we briefly overview the signal model, and then, describe the preprocessing steps for FMCW radar data, which serve as the input for our DL model. Second, our proposed model is described in Section II-C. Finally, Section II-D addresses the postprocessing aspects of the predictions provided by the model.

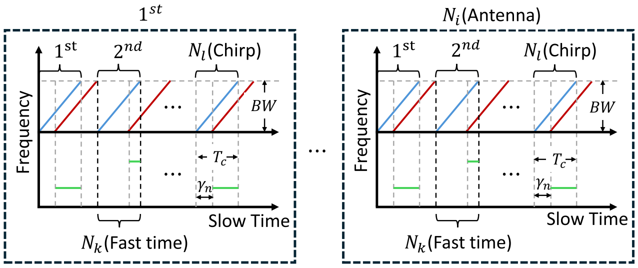


Fig. 2. Illustration of the principles of linear chirp FMCW radar; it shows how the frequency sweeps during the radar operations. The blue line represents the transmitting chirp, the red line denotes the receiving chirp, and the green line illustrates the IF signal derived by "beating" the received signal with the transmitted one.

A. Radar Signal Model

In our study, we used FMCW radar to gather head motion signatures from each participant. Essentially, FMCW radar estimates target distance by generating intermediate frequency (IF, or "beat" frequency) signals by multiplying a copy of the transmitted signal and the received signal. The transmitted signal's frequency is swept within a specified bandwidth, as shown in Fig. 2. In contrast, the received signal is assumed to be a delayed copy of the transmitted signal, with additional phase shifts in case of moving targets. The IF signal within a single radar frame can be modeled as follows:

$$z_{i,k,l} = \sum_{n=1}^{N} \alpha_n \exp\left(j2\pi f_0 \frac{id}{c} \sin\theta_n\right) \times \exp\left(-j2\pi\left(f_0 \frac{2v_n}{c} T_p l + \mu\gamma_n \frac{k}{f_s}\right)\right) \quad (1)$$

where $z_{i,k,l}$ represents specifically the discretized IF signal (dechirp). There, $i = [0, 1, \ldots, N_i - 1]$ denotes the index of the antennas in the radar linear virtual array, where N_i is the total number of virtual antennas. Similarly, $k = [0, 1, \ldots, N_k - 1]$ denotes the number of sampling indices (fast time), where $N_k = T_c f_s$ represents the maximum number of samples in one chirp. Here, T_c denotes the chirp duration, and f_s denotes the sampling frequency. Finally, $l = [0, 1, \ldots, N_l - 1]$ is the number of chirp indices (slow time), with N_l being the total number of chirps. Then, α_n is a constant complex amplitude related to the characteristics of the target n, while f_0 represents the starting frequency and c denotes the speed of light. θ_n denotes the azimuth of the target n, d represents the spacing between adjacent antennas, and v_n represents the radial velocity between the radar and target n. T_p stands for the pulse repetition time, and μ denotes the frequency modulation rate. Finally, $\gamma_n = \frac{2D_n}{c} \ll T_c$ represents the time delay of the echo signal relative to the transmitting signal, where D_n denotes the distance between the radar and target n.

For a more comprehensive discussion of the signal models and system considerations of FMCW radar, refer to the detailed explanation in [26] and[27].

B. Preprocessing

An overview of the preprocessing on the radar data is shown in Fig. 3 on the left-hand side, and the steps are detailed as follows.

1) Range-Angle Data Representation: To generate a range-angle map, we utilize the conventional approach based on the Fourier transform. Fast Fourier transform (FFT) is applied along both the fast time axis and antennas/channels axis of the radar signal previously defined in (1). The output of the range-angle Fourier transform is determined by

$$\hat{z}_{\hat{i},\hat{k},l} = \sum_{i=0}^{N_i-1} \sum_{k=0}^{N_k-1} z_{i,k,l} \cdot \exp\left(-j2\pi\left(\frac{\hat{i}}{N_i} i + \frac{\hat{k}}{N_k} k\right)\right) \quad (2)$$

where $\hat{i}$ and $\hat{k}$ are the angle and range bin after FFT, respectively. Next, the modulus of the signal mentioned previously is computed to generate the range-angle plot, where peaks in the data correspond to the distance (range bin) and azimuth angle (angle bin) of the target head. In addition, to minimize environmental noise and clutter, we limit the range bins to concentrate solely on the region containing the target head. Furthermore, it should be noted that zero padding is implemented in both FFTs beforehand to enhance the visual quality of the resulting range-angle map in both the range and angle direction.

2) Clutter Removal: As in this study, we aim to use a DL approach for head pose estimation across diverse environments, the proposed model should be able to maintain background invariance. This typically involves training on a large dataset covering multiple environments, enabling the network to capture various background clutter or multipath propagation effects. However, resource constraints led to a relatively small radar head pose estimation dataset. Hence, to address this background invariance, we use a classical technique to mitigate the static clutter effect through interframe processing. Specifically, this process involves averaging the range bins across the chirps within a single frame and subtracting this average from each chirp. The clutter removal equation is defined as follows:

$$\hat{z}'_{\hat{i},\hat{k},l} = \hat{z}_{\hat{i},\hat{k},l} - \frac{1}{N_l} \sum_{l=0}^{N_l-1} \hat{z}_{\hat{i},\hat{k},l} \quad (3)$$

where $\hat{z}'$ represents the range-angle feature after static clutter removal.

C. Estimation Framework With DL Model

To estimate the head pose, we take inspiration from two deep learning models [24], [25] that predict frame-by-frame range-angle features as defined in the previous subsection, i.e., after static clutter removal. Initially, we modify the RA-CNN model and propose a combination of ReDA-PKC with RA-CNN, as shown on the right hand of Fig. 3.

1) *RA-CNN:* We propose a modified FCN network derived from the RA-CNN network in [24] for estimating the target head pose rotation. The model comprises three branches for roll, pitch, and yaw prediction, each sharing

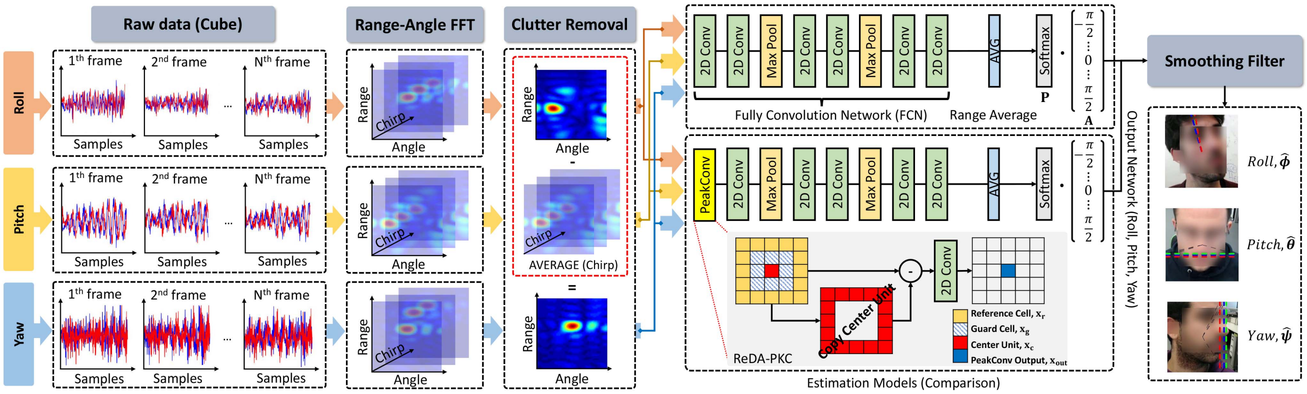


Fig. 3. Overview of our proposed model. The left side of the figure demonstrates the preprocessing step. After data collection, a double FFT operation generates range-angle plots to extract representative features. Next, clutter removal techniques are applied to reduce the impact of static background clutter. Our deep learning model for head pose estimation is depicted on the right side of the figure. We provide a comparison with two existing models: first, the RA-CNN network at the top, and second, the ReDA-PKC + RA-CNN model at the bottom. Finally, a smoothing filter is employed on the predictions to improve their smoothness over time.

an identical architecture. This architecture includes six 2-D convolution layers with rectified linear unit activations and two max pooling layers. The output of the last convolution layers is averaged along the range axis, yielding a dimensionality-reduced 1-D vector. This reduction preserves meaningful features, given that head pose rotation primarily relies on changes in the peak along the angle axis. Next, the logits from the last layer are converted into probabilities, **P**, using Softmax, representing the probability distribution as a function of angle. Finally, to convert these probabilities into the final prediction, we calculate the expectation of the angle bin index, **A**, over, **P** as

$$\hat{\mathbf{P}}_t = \mathbf{P}_t \cdot \mathbf{A}, \quad t = \{\hat{\phi}, \hat{\theta}, \hat{\psi}\} \tag{4}$$

and

$$\mathbf{A} = \left[-\frac{\pi}{2}, \ldots, 0, \ldots, \frac{\pi}{2}\right]^\top \tag{5}$$

where $\hat{\mathbf{P}}_t \in [-\frac{\pi}{2}, \frac{\pi}{2}]$ is the predicted output of the network for the head pose, and the elements of **A** are arranged based on the given range of **P**.

The utilization of the FCN model aims to maintain a property known as translation equivariance. This property ensures that when the peak in the range-angle feature translates by a certain amount, the output in each layer shifts by the same amount. This feature aids in the model's ability to learn, mainly when training data for some specific participants that may not cover all possible head movement angles. The convolutional kernels learn to detect peaks in a limited training region and can reuse such learned patterns across various movements, including unseen ones.

2) *ReDA-PKC + RA-CNN:* In addition to the modified RA-CNN, we introduced a fusion approach combining the RA-CNN with the ReDA-PKC [25]. This model replaces the initial layers of RA-CNN with ReDA-PKC to capture significant peak information from the range-angle features. The concept behind this is inspired by the CFAR algorithm, commonly used in radar systems for peak detection. However, unlike traditional CFAR methods, where the threshold for the cell under test is calculated solely based on the surrounding reference cells, leaving the center cell excluded in the threshold calculation, the ReDA-PKC method also involves the cell under test in the thresholding process. Consequently, the thresholding factor is calculated based on the variance between each reference cell and the center cell. This results in enhanced interference suppression, as the kernel convolution incorporates both the reference cells and the cell under test values in the thresholding process. The output of ReDA-PKC network, $\mathbf{x}_{\text{out}}$, can be determined by

$$\mathbf{x}_{\text{out}} = \text{Vec}\left(\left\{\sum_{i=1}^{N_r} \mathbf{w}_j^i * (\mathbf{x}_c - \mathbf{x}_r^i)\right\}_{j=1}^{C_{\text{out}}}\right), \quad \mathbf{w}_j^i \in \mathbb{R}^{C_{\text{in}}} \tag{6}$$

where $\mathbf{x}_c \in \mathbb{R}^{C_{\text{in}}}$ represents the cell under test, $\mathbf{x}_r^i \in \mathbb{R}^{C_{\text{in}}}$ denotes the set of reference cells at i elements, **w** represents the set of learning weights, and N_r is the number of reference cells. The symbol $*$ denotes the convolution operator. C_{in} and C_{out} represent the channel input and output, respectively, both set equal to the number of chirp indices, N_l, in this study. Finally, similar to traditional CFAR, a guard cell, $\mathbf{x}_g$, is necessary to establish buffer zones around the cell under test to prevent target signal leakage to the reference cells.

Subsequently, the ReDA-PKC's output will be directed to the RA-CNN model to estimate head pose rotation. As mentioned, adding ReDA-PKC before RA-CNN mitigates interference signals that may not be entirely removed during preprocessing and static clutter removal. Likewise, this technique helps the network prioritize significant peaks within the range-angle maps.

In our study, we train both models using the mean absolute error loss function, ensuring that the predicted head rotation is optimized to match the ground-truth head pose, as defined by

$$L = \mathbb{E}_t[||\bar{\mathbf{P}}_t - \hat{\mathbf{P}}_t||_1] \tag{7}$$

where $\bar{\mathbf{P}}_t$ is the ground truth of head pose rotation, with the models being optimized independently for roll, pitch, and yaw.

D. Postprocessing

As our prediction model operates by providing predictions for every single frame, intermittent jitter trajectories may occur with unrealistic fluctuations in predictions on a frame-by-frame basis, i.e., unrealistic in terms of physical human head movement. To mitigate this, we use the Savitzky–Golay filter after prediction to improve the realism of head movement trajectories and obtain a smoothing effect. This is defined by

$$\tilde{\mathrm{P}}_t^i = \sum_{j=-\frac{W-1}{2}}^{\frac{W-1}{2}} C_f \cdot \hat{\mathrm{P}}_t^{i+j} \tag{8}$$

where $\tilde{\mathrm{P}}_t^i$ represents prediction values after filtering at position i, with W is the size of window used for filtering. C_f are the filter coefficients, and $\hat{\mathrm{P}}_t^i$ is the element of the vector $\hat{\mathbf{P}}_t$. For this study, we set the window size to 35, and the polynomial fitting order is set to 2. This choice is based on the observation that a larger window size and a lower order polynomial yield smoother results, effectively filtering out highly intermittent jitter trajectories and aligning more closely with realistic head movements.

E. Implementation Details

For the RA-CNN model, we employ six layers of 2-D convolution with filter sizes {32, 32, 32, 32, 32, 1}, all with a kernel size of 3. The number of layers and convolution channels are determined through grid search hyperparameter tuning. Given the extensive range of combinations and models to train, we restrict the search space to {4, 6, 8, 10} as the number of convolutional layers in the overall architecture and to {16, 32, 64} for the number of filters in each convolutional layer; the output of the last channel is set to 1. In addition, we incorporate two max-pooling layers after every two convolution layers, using a kernel size of 3 to minimize less significant numbers generated by the convolution operations. For the ReDA-PKC model, the bandwidth of the reference and guard cells are set to 1 and 2, respectively. Training is carried out using the Adam optimizer with a constant learning rate of 1×10^3 and a batch size of 100. The model undergoes training for 200 epochs, and on a single Nvidia GeForce RTX 2080 Ti, the process takes approximately 2 min for each run. Details of the different combinations of data used for training and testing are provided in the subsequent section with results.

III. RADAR HEAD POSE DATASET

This section presents the dataset to validate the proposed approach for head pose estimation. Section III-A details the equipment setup. Next, Section III-B describes how the data collection process was conducted. Finally, in Section III-C, we describe the process of generating ground truth and refining it through calibration to adjust pose positions within the dataset.

TABLE I
Radar Hardware Parameters

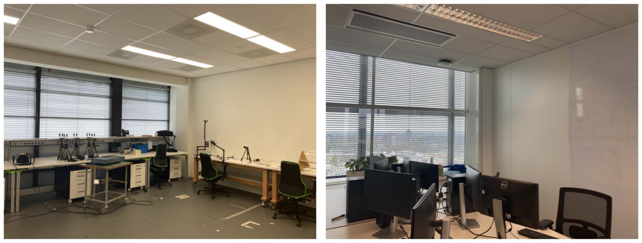
Item	Value
Bandwidth	≈ 3.19 GHz
Number of chirps	32
Frame rate	30 fps
Samples per chirp	256
Number of virtual antennas	12
Multiplexing techniques	TDM

A. Equipment Setup

To generate our radar-based head pose estimation dataset, we present a vision-based system for simultaneous data collection at each time step, capturing both radar signals and 2-D images at each time step. In this dataset, we use the Texas Instruments mmWave IWR1443 FMCW radar, which operates in the 76–81 GHz frequency band, and some of its detailed parameters are provided in Table I. We introduced ten types of movements detailed as follows to cover all possible trajectories of head motions. Plus, the radar array geometry/location concerning the head is optimized in three different configurations, namely, radar on top of the head (RTH), radar aligned vertically (RAV), and radar aligned horizontally (RAH). This is done to maximize the available angular resolution for each movement type, as shown in Fig. 1, essentially by aligning the length of the resulting radar virtual array along the direction of the dominant head movement for that specific movement. These head movement types include the following.

1) Roll (tilting head left ↔ right) with RTH.
2) Pitch (moving head up ↔ down) with RAV.
3) Yaw (moving head left ↔ right) with RAH.
4) Left tilting (tilting head left ↔ center) with RTH.
5) Right tilting (tilting head right ↔ center) with RTH
6) Center-up (moving head up ↔ center) with RAV.
7) Center-down (moving head down ↔ center) with RAV.
8) Center-left (moving head left ↔ center) with RAH.
9) Center-right (moving head right ↔ center) with RAH.
10) Free head movement, repeating collections to utilize the radar in all three positions (RTH, RAV, and RAH).

Our experiments were conducted in two environments with different levels of background clutter: laboratory and office scenes. The laboratory scene had minimal clutter, with a large empty space and no walls within the study range, resulting in reduced multipath effects on the collected data. In contrast, the office scene contained significantly more clutter, such as walls and multiple office chairs and desks, introducing higher levels of contamination into the



(a) (b)

Fig. 4. Photos of the environments used for data collection.
(a) Laboratory: Characterized by minimal clutter and wide spaces.
(b) Office: Featuring numerous background elements, such as office chairs, monitors, and desks close to each other, with limited space.

data in the form of clutter and multipath. Both environments are shown in Fig. 4 .

To maximize the signal-to-noise ratio (SNR) and maintain consistent model performance regardless of distance variations, we limited the study range in our work to approximately 15–30 cm, with the target positioned within the radar's line of sight.

B. Data Collection

We recruited ten volunteers for this experiment, with heights ranging from 170 to 184 cm and weights from 60 to 90 kg. We captured data for 11 s during each movement, yielding 330 frames from both radar and camera. The recordings were captured with a brief pause in between them. This entire cycle was repeated five times, resulting in 1650 frames per movement and 19 800 per volunteer for all movement types. This process was carried out in laboratory and office settings, where the total data were duplicated. We did not impose limitations on the range of head angles, head movement speed, or motion repetition in each cycle, leaving the participants free to achieve a broader range of movement patterns or maintain a closer resemblance to real-world scenarios. The data collection process received approval from the Human Research Ethics Committee, Delft University of Technology.

C. Ground-Truth Generation and Calibration Method

To create ground-truth data for training, 2-D images collected alongside radar data were processed using a large-scale dense, accurate, and diverse dataset for 3D head alignment from a single image (DAD-3DHeads) model [28], which outputs roll, pitch, and yaw values. However, these ground truths are based on the camera orientation, whereas we require ground truths corresponding to the radar orientation. Therefore, we also employed the quick response (QR) marker, which was used for calibration orientation to obtain the accurate pose of the human head, aligning it with the radar orientation .

We employ a vision-based system to determine the head pose rotation relative to the radar orientation, using a QR marker based on the augmented reality university of cordoba (ArUco) [29] library, as shown in the concept illustration of Fig. 5. The goal is to use the QR markers as references for the poses of each device (radar and camera), and to calculate the angular difference between them. This angular difference will be used to adjust the generated data to be used as ground truth. To do this, an additional camera is needed to estimate the poses of the two QR markers. Our technique comprises two steps. First, we attach QR markers to the radar and the camera to serve as reference poses. We then estimate the extrinsic of the camera and radar, Ec, $Er \in \mathbb{SE}(3)$, by solving the perspective n-point problem. The camera intrinsic parameters are obtained through standard calibration with an OpenCV checkerboard. Second, we calculate the relative pose between the camera and radar, defined as

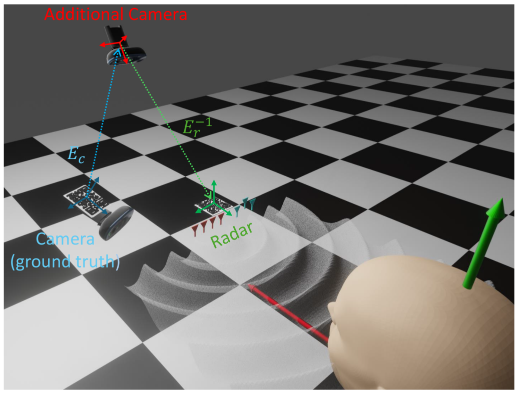


Fig. 5. Illustration of the vision-based system used to calibrate data generated from the DAD-3DHeads model to be used as ground truth.

$$E_{c\to r} = E_r^{-1} E_c \tag{9}$$

where $E_{c\to r}$ represents the relative pose from the camera to the radar. This indicates the angular difference between the radar and camera, serving as a compensatory angle for ground truths predicted by the DAD-3DHeads network.

IV. EXPERIMENTS AND RESULTS

In this section, we present the results of four experiments conducted to evaluate the performance of our model, all of which were conducted in a subject-specific manner, namely, the following:

1) we show the accuracy of our prediction model in estimating environment-specific head poses;
2) we evaluate the extent to which our model can generalize to different environments by training in one and testing in another;
3) we highlight the performance enhancements resulting from an increase in the size of the available training dataset;
4) we evaluate our model's performance against alternative models inspired by the literature.

Fig. 6 illustrates an overview of the first three experiments.

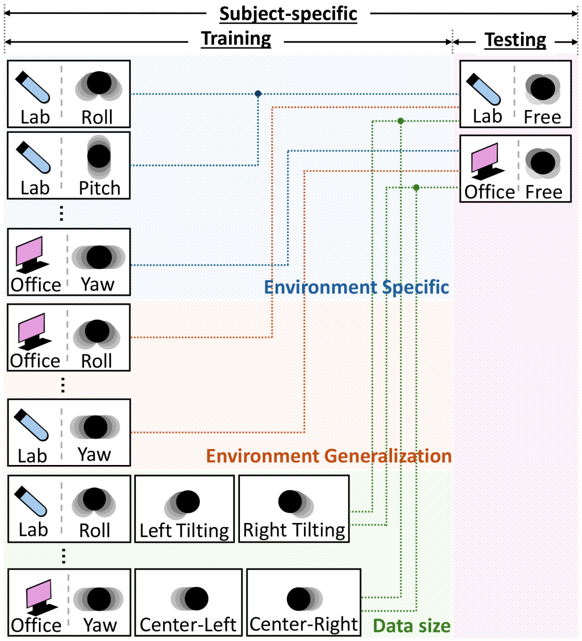


Fig. 6. Illustration of the different experiment configurations to demonstrate the training and testing process in different environments and for different head motions.

A. Environment Specific

We assess the efficacy of our proposed prediction model in fitting subject-specific Head pose estimations. The models are trained and tested on the same subjects and environment. Specifically, the models are trained on specific motion data sequences (roll, pitch, and yaw only), and then, tested on free-head movement sequences, with the antenna array geometry of the testing data matching that of the training data. For each training model, 50% of the data from the specific motion is allocated to training, and the remaining 50% from free head movement is allocated for testing data.

In Fig. 7 and Table II, the proposed ReDA-PKC + RA-CNN model shows slightly better performance than RA-CNN, with mean absolute errors (MAEs) of 6.7°, 10.10°, and 14.4° for roll, pitch, and yaw, respectively, in the lab–lab (training–testing) scene, and 7.2°, 8.7°, and 13.5° for roll, pitch, and yaw, respectively, in the office–office (training–testing) scene. For individual subjects, the lowest average MAE is 7.4° on P-10 for the lab–lab scene and 7.2° on P-2 for the office–office scene, while the highest MAE is 15.6° on P-1 for the lab–lab scene and 12.4° on P-7 for the office–office scene. Considering the typical head rotation angle ($\sim 180°$), the 7.4° and 7.2° errors translate to approximately 4.1% and 4.0% error at the minimum, while the 15.6° and 12.4° errors correspond to around 8.7% and 6.9% error in angle at maximum. These results indicate that the angle errors are relatively small compared to regular rotation, as observed in [23], where they achieved error angles ranging from 8.63° to 18.24° in terms of yaw and pitch movements.

TABLE II
Head Pose Estimation Results for Each Participant (Subject-Specific in Each Environment, Lab and Office)

Note:FCN = RA-CNN, P+FCN = ReDA-PKC + RA-CNN

	MAE Roll ↓		MAE Pitch ↓		MAE Yaw ↓	
Participants	FCN	P+FCN	FCN	P+FCN	FCN	P+FCN
P-1	10.3	10.3	14.8	13.3	23.4	23.2
P-2	8.4	8.4	9.2	9.1	10.4	8.4
P-3	6.3	5.9	8.2	8.6	12.1	12.4
P-4	5.3	5.8	14.1	13.1	6.6	7.5
P-5	4.6	4.7	4.9	4.6	16.1	15.4
P-6	10.4	6.6	9.5	9.3	11.2	9.9
P-7	9.1	7.2	13.9	14.7	16.7	18.4
P-8	5.6	5.7	14.4	12.3	23.0	23.1
P-9	7.6	8.4	11.2	12.2	16.7	17.2
P-10	4.2	3.8	13.7	10.0	7.8	8.3
Mean (μ)	7.2	**6.7**	11.4	**10.7**	14.4	14.4
SD (σ)	2.2	**1.8**	3.2	**2.8**	**5.5**	5.7

(a) Training–Testing: Lab–Lab

	MAE Roll ↓		MAE Pitch ↓		MAE Yaw ↓	
Participants	FCN	P+FCN	FCN	P+FCN	FCN	P+FCN
P-1	10.1	9.7	13.7	11.4	14.1	14.1
P-2	6.0	6.0	6.9	6.7	9.9	9.0
P-3	6.9	6.8	8.8	9.7	22.7	19.1
P-4	8.9	10.3	7.5	7.7	9.2	9.6
P-5	6.2	6.1	6.4	6.5	17.7	16.2
P-6	4.6	4.8	6.6	5.4	11.1	10.4
P-7	7.3	6.0	9.6	9.7	23.8	22.0
P-8	7.2	8.0	11.5	11.4	12.3	12.6
P-9	7.1	7.4	12.2	11.6	16.1	13.5
P-10	7.4	7.1	8.6	8.1	9.9	8.2
Mean (μ)	7.2	7.2	9.2	**8.7**	14.7	**13.5**
SD (σ)	**1.4**	1.6	2.4	**2.1**	5.0	**4.3**

(b) Training–Testing: Office–Office

Bold text highlights the best values.

This demonstrates our model's capacity to accurately estimate head pose rotations in subject-environment-specific tasks.

B. Environment Generalization

We demonstrate how our method effectively achieves environment generalization for head pose estimation. Specifically, the proposed model is trained in laboratory scenes and evaluated in office environments. The training–testing split remains at 50% and 50%, and the configuration of the antenna array geometry and training procedures remains unchanged from the previous experiment.

In the case of the lab–office (training–testing) scene, the average MAEs for roll, pitch, and yaw are 10.8°, 11.7°, and 17.7° for ReDA-PCK + RA-CNN, respectively. On the other hand, the office–lab (training–testing) scene yields average MAEs of 8.7°, 13.6°, and 18.8° for roll, pitch, and yaw, respectively. Specifically, P-6 exhibits the lowest average MAE for the lab–office scene, at 8.3°, while P-10 demonstrates the lowest average MAE for the office–lab scene, at 9.9°. Regarding the highest error, P-1 and P-4 display average MAEs of 20.6° and 16.7°, respectively,

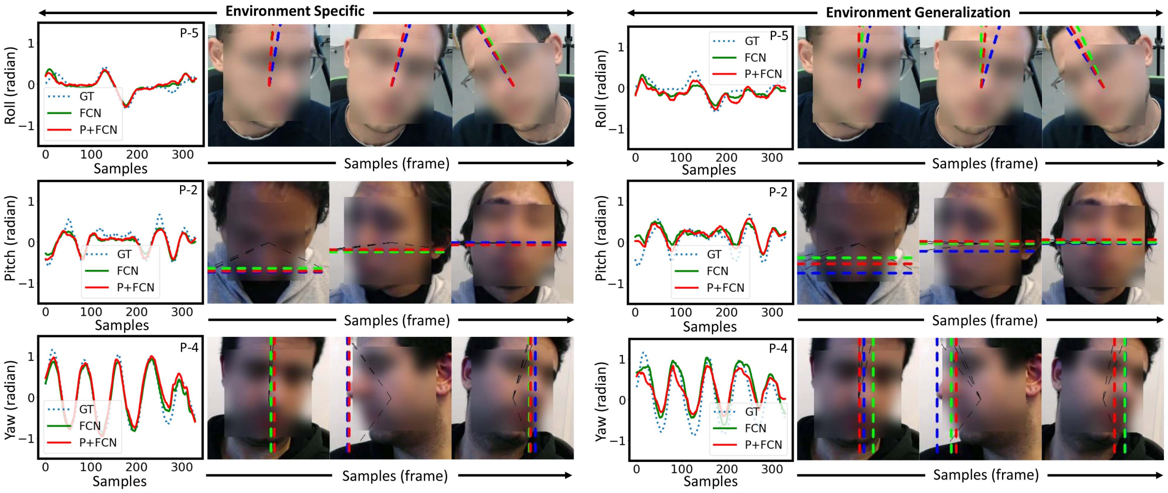
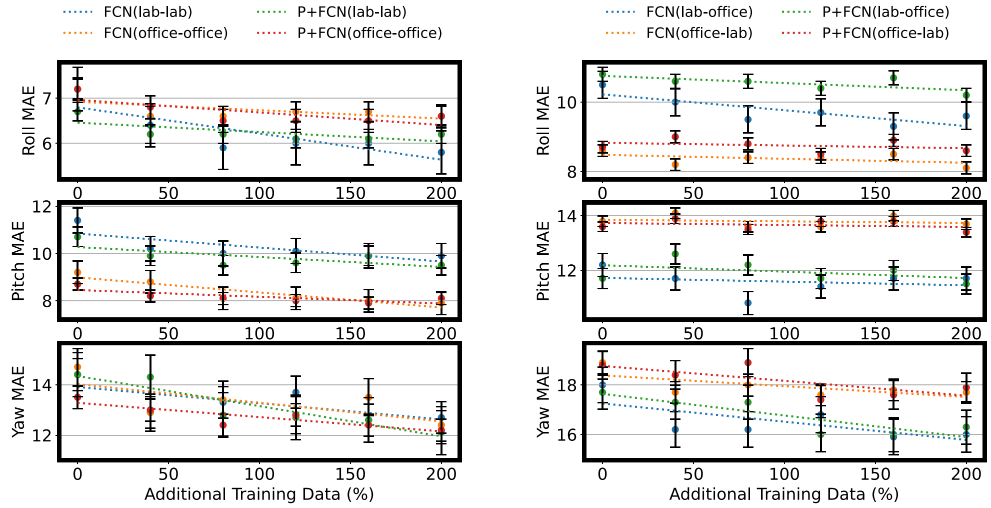


Fig. 7. Qualitative results of our proposed method are depicted for both scenarios: on the left for subject- or environment-specific cases, and on the right, for environment generalization. The figures present a comparison among the ground truth (blue), RA-CNN (green), and ReDA-PKC + RA-CNN (red). Our proposed technique can closely predict head poses, aligning well with the ground truth. [*Note*: FCN = RA-CNN, P+FCN = ReDA-PKC + RA-CNN, and GT = Ground Truth].



(a)

(b)

Fig. 8. Graph illustrates the MAEs) in degrees against the increase in training data size for both (a) environment-specific cases and (b) environment generalization scenarios.

for the lab–office and office–lab scene. We observed a decrease in performance regarding environment generalization ability compared to subject-environment-specific experiments, falling within the range of 2.0°–4.2° However, this degradation is considered to be minor, indicating that the model is capable of environment generalization, as shown in Table III.

C. Effect of the Training Size

In this experiment, we aim to investigate the impact of training data size on the performance of our model. To explore this, we augment the training size by incorporating additional data equivalent to 40%, 80%, 120%, 160%, and 200% of the base training data size. However, we do not add identical movement data sequences (roll, pitch, and yaw) to introduce more variability into the training samples. Instead, we request volunteers to perform movements covering only a partial range of angles within each motion category. For instance, during the yaw movement, volunteers are instructed to execute two phases: first, moving the head from the center to the left, and second, moving it from the center to the right. This approach allows for a broader range of head movements, contributing to a more diverse training dataset. As depicted in Fig. 8, including additional data results in performance improvements, particularly for yaw angle prediction. Here, the percentages represent the additional training data; for instance, 200% indicates that the sample size doubles compared to regular training.

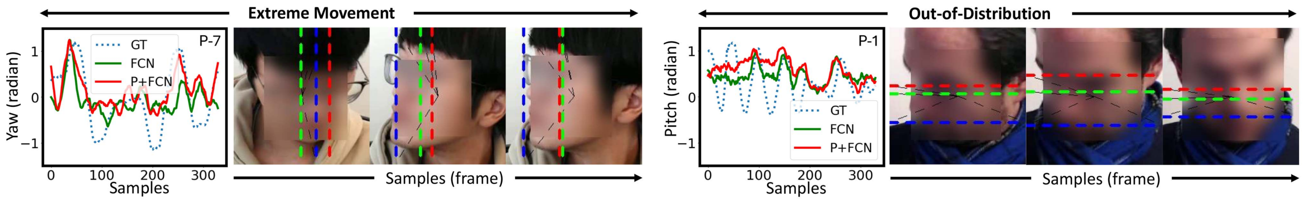


Fig. 9. Example of qualitative results illustrating the worst-case scenarios prediction: on the left, an example due to extreme movement of the head (i.e., movement involving not just the head but also distinctively the body), and on the right, an example due to OOD cases (i.e., cases where the training data does not cover all angles present in the testing data) [*Note*: FCN = RA-CNN, P+FCN = ReDA-PKC + RA-CNN, GT = Ground Truth].

TABLE III
Head Pose Estimation Results for Each Participant (Environment Generalization, i.e., Train in Lab and Test in Office and Vice Versa)

Note: FCN = RA-CNN, P+FCN = ReDA-PKC + RA-CNN

	MAE Roll ↓		MAE Pitch ↓		MAE Yaw ↓	
Participants	FCN	P+FCN	FCN	P+FCN	FCN	P+FCN
P-1	20.0	21.2	16.2	17.5	23.8	23.2
P-2	13.4	13.0	9.2	7.8	13.3	10.9
P-3	7.2	7.8	11.6	11.5	13.5	13.8
P-4	14.4	14.1	16.3	18.2	15.5	14.7
P-5	8.6	9.8	9.9	8.2	19.4	20.4
P-6	6.3	4.5	8.9	7.9	13.7	12.3
P-7	5.5	6.3	11.5	10.6	26.8	27.2
P-8	10.9	11.3	16.1	14.9	18.4	17.0
P-9	9.2	9.3	12.7	12.0	23.1	25.1
P-10	9.7	10.4	9.8	8.8	12.6	12.7
Mean (μ)	**10.5**	10.8	12.2	**11.7**	18.0	**17.7**
SD (σ)	**4.2**	4.4	**2.8**	3.7	**4.9**	5.5

(a) Training–Testing: Lab–Office

	MAE Roll ↓		MAE Pitch ↓		MAE Yaw ↓	
Participants	FCN	P+FCN	FCN	P+FCN	FCN	P+FCN
P-1	12.3	12.3	20.2	19.1	18.3	16.5
P-2	11.1	10.4	13.8	12.1	10.9	10.7
P-3	6.6	5.8	12.4	12.5	16.1	17.8
P-4	9.2	10.6	14.6	14.3	22.1	25.3
P-5	6.7	7.2	9.2	9.8	24.2	21.6
P-6	5.2	4.0	11.6	11.3	14.9	15.5
P-7	12.8	11.9	16.1	15.1	24.5	22.7
P-8	8.8	9.6	14.9	16.9	21.9	21.2
P-9	6.8	7.3	13.9	14.6	25.0	25.9
P-10	6.4	7.9	10.8	10.6	11.4	11.3
Mean (μ)	**8.6**	8.7	13.8	**13.6**	18.9	**18.8**
SD (σ)	2.6	2.6	2.9	**2.8**	5.1	5.1

(b) Training–Testing: Office–Lab

Bold text highlights the best values.

D. Comparison With Alternative Models

This section compares our proposed model with alternative designs to highlight its performance for human head pose estimation. As, to the best of our knowledge, no other model is available in the open literature for a direct, like-for-like radar-based comparison, we implement the following four models inspired from relevant literature. All these models perform frame-by-frame predictions for a fair comparison with our proposed approach, specifically, the following:

1) FCN with a regression layer for range-angle map superresolution inspired by the work by J. Smith et al. [30];
2) Multiview-3DCNN for human pose inference inspired by the work by S. Scholes et al. [31];

TABLE IV
Head Pose Estimation Results: Average MAE for Proposed Models Versus Alternative Models (Environment Generalization, i.e., Train in Lab and Test in Office and Vice Versa)

Note: Super Res- = Super Resolution, Head Ori- = Head Orientation

Methods	Input	Task	Roll ↓	Pitch ↓	Yaw ↓
J. Smith [30]	RA	Super Res-	12.5	23.9	31.9
S. Scholes [31]	RA,RD	Body Pose	11.8	23.8	26.5
J. Jung [32]	RA	Head Ori-	11.2	19.7	25.1
mmPose [13]	Point Cloud	Body Pose	10.6	11.8	19.8
FCN (Ours)	RA	Head Pose	**10.5**	12.2	18.0
P+FCN (Ours)	RA	Head Pose	10.8	**11.7**	**17.7**

(a) Training–Testing: Lab–Office

Methods	Input	Task	Roll ↓	Pitch ↓	Yaw ↓
J. Smith [30]	RA	Super Res-	16.2	22.5	26.3
S. Scholes [31]	RA,RD	Body Pose	14.3	21.0	25.1
J. Jung [32]	RA	Head Ori-	13.5	19.9	20.9
mmPose [13]	Point Cloud	Body Pose	9.7	16.8	18.9
FCN (Ours)	RA	Head Pose	**8.6**	13.8	18.9
P+FCN (Ours)	RA	Head Pose	8.7	**13.6**	**18.8**

(a) Training–Testing: Office–Lab

Bold text highlights the best values.

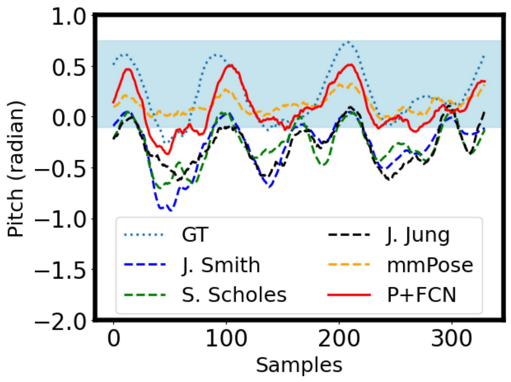


Fig. 10. Example of pitch estimation for each considered model on P-5. It is shown that our approach achieves more accurate predictions in unseen regions (light blue area), i.e., those not fully included in the training data.

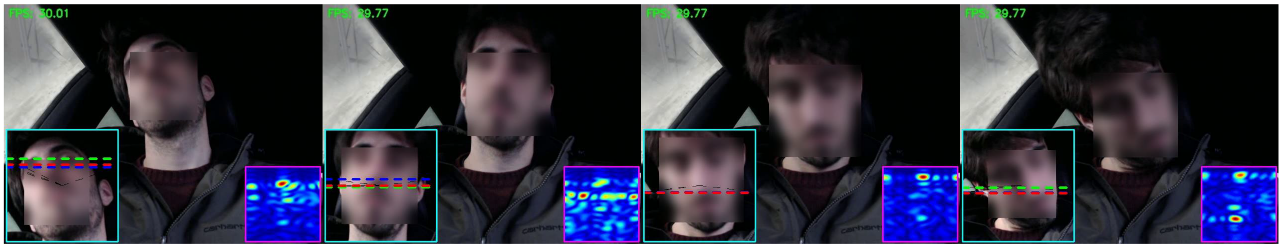


Fig. 11. Example of applying the proposed model in the front seat of a car, where the model was previously trained on data collected in an office scene.

3) MobileNet for eye-gaze and head orientation inspired by the work by J. Jung et al. [32];
4) mmPose [13], a state-of-the-art radar-based pose estimation model.

For a clear and comprehensive performance comparison, this evaluation is performed using an environment generalization scheme as described in Section IV-B. Moreover, we reoptimized some of the models' layers to address differences in input/output sizes and data variance, as detailed in the Appendix.

As shown in Table IV, the proposed models outperform all the considered alternative models, showing the effectiveness of our purposely designed model architectures for the task of head pose estimation with radar. Moreover, a key distinction is that, due to the translation equivariance property, our model performs better in unseen regions than the others (i.e., in regions that were not fully included in the training data), as also illustrated in Fig. 10.

V. Limitations and Discussion

The results presented in this article highlight the potential of the proposed approach for predicting subject-specific head pose rotation from FMCW radar data. Nevertheless, it is essential to explore limitations as well and identify the cases where the model struggles to predict head poses effectively, categorized into two prominent cases: extreme head movement cases, i.e., those performed by also distinctively moving the whole body and not just the head, and out-of-distribution (OOD) instances, as illustrated in Fig. 9. Finally, in the discussion in Section V-C, we explore the potential application of our work in different settings, including utilizing our model in car environments.

A. Extreme Head Movement Cases

The extreme head movement occurs when volunteers move their bodies and heads distinctively. For example, in the case of P-7, this presents an example of incorrect estimation provided by the proposed model due to the volunteer leaning their body toward the radar while in motion. As the reflected radar signal originates from the superposition of two different extended sources, such as the head and body, the model faces difficulty distinguishing between head and body movements, resulting in incorrect predictions.

B. OOD Cases

In these cases, incorrect predictions may occur when the testing movement data differs significantly from the training data. For example, in the case of P-1, the volunteer's head movements were limited to a slight angle while collecting pitch movement data (training data). However, in the subsequent free movement sequences (testing data), the volunteer moved their head at a wider angle in the pitch direction. As a result, the prediction becomes constrained within specific values, unable to extrapolate head poses to unseen larger angles. Despite the network being designed to preserve translation equivariance property, it still struggles to handle such scenarios where there is a considerable discrepancy between the training and testing data.

C. Discussion

This study demonstrates the feasibility of using FMCW radar data for head pose estimation tasks through DL techniques. Specifically, we introduce the combination of ReDA-PKC and RA-CNN networks, which predicts head pose rotation frame-by-frame from range-angle plots. However, we only focus on subject-specific tasks (i.e., training and testing on data from the same participant), leaving subject generalization aspects unexplored in this study. Nevertheless, the results mentioned earlier showed that the model can generalize to data collected in unseen environments, for example, by training with data collected in the lab and testing with data collected in the office, and vice versa. Furthermore, we investigated the capability of the approach to generalize to a new environment by collecting additional data within a car's cabin, an environment with rather complex propagation characteristics due to the confined space and highly reflective surfaces. Preliminary results for this test are presented in Fig. 11. Specifically, by training the DL model on office scenes and testing it with car scene data, we highlight its adaptability beyond laboratory and office settings, showing potential applications in areas such as driving behavior detection and drowsiness detection.

VI. Conclusion

This article proposes the first implementation of subject-specific head pose estimation using FMCW radar data. Specifically, a DL model is formulated based on the combination of ReDA-PKC and RA-CNN networks, and this can estimate head rotations/orientations in terms of roll, pitch, and yaw angles based on a single frame of radar data. Experimental results reveal that our model achieves average MAEs ranging from 6.7° to 7.2° for roll, 8.7° to 10.7° for pitch, and 13.5° to 14.4° for yaw in environment-specific scenarios. In environment generalization scenarios where the model is trained with data in one location and tested with data in another, the MAEs range from 8.7° to 10.8° for roll, 11.7° to 13.6° for pitch, and 17.7° to 18.8° for yaw. Regarding model comparison, the performance of ReDA-PKC + RA-CNN shows improvements over RA-CNN ranging from 5.75% to 8.89% in environment-specific cases and from 0.53% to 6.21% in environment generalization cases. Notably, these errors are relatively small compared to the actual head rotation, indicating the model's effectiveness in accurately estimating head pose rotation from FMCW radar data for each individual.

APPENDIX

A. Modifications of Alternative Models

To achieve optimal training for each model and ensure they are compatible with our input and output sizes, we reoptimized certain layers in the models as follows.

1) *J. Smith. et al. [30]:* This FCN-based model cannot output a single node for roll, pitch, and yaw. To address this, we simply added a single node output and adjusted the first convolution layer to match the input filter size with the number of chirps.
2) *S. Scholes. et al. [31]:* This multiview-3DCNN has two branches: one for the range-angle map and another for the range-Doppler map. We modified the input filter size of the first convolution layer in the first branch to match the number of chirps, and the second branch to match the number of virtual antennas.
3) *J. Jung et al. [32]:* This MobileNet-based model was modified to match the input filter size with the number of chirps, and a single node output was added for roll, pitch, and yaw estimation.
4) *mmPose [13]:* We modified only the final layer to output a single node for roll, pitch, and yaw estimates.

Finally, the output estimations of each model undergo the same postprocessing as in our model for a fair comparison.

ACKNOWLEDGMENT

The authors would like to express their gratitude to all participants of the data collection at TU Delft, including I. Tokucoglu, R. Güneri, N. Kruse, A. Pappas, I. Roldan, S. Zhu, D. Wang, and S. Yuan. Furthermore, the authors would especially like to thank D. Wang, S. Yuan, and J. Wang for the valuable technical discussions on radar signal processing and N. B. Onat for lending his car for this experiment.

REFERENCES

[1] C. Li et al., "A review on recent progress of portable short-range noncontact microwave radar systems," *IEEE Trans. Microw. Theory Techn.*, vol. 65, no. 5, pp. 1692–1706, May 2017.

[2] S. Pisa, E. Pittella, and E. Piuzzi, "A survey of radar systems for medical applications," *IEEE Aerosp. Electron. Syst. Mag.*, vol. 31, no. 11, pp. 64–81, Nov. 2016.

[3] C. Waldschmidt, J. Hasch, and W. Menzel, "Automotive radar—From first efforts to future systems," *IEEE J. Microw.*, vol. 1, no. 1, pp. 135–148, Jan. 2021.

[4] S. Lim, J. Jung, E. Lee, J. Choi, and S.-C. Kim, "In-vehicle passenger occupancy detection using 60-GHz FMCW radar sensor," *IEEE Internet Things J.*, vol. 11, no. 4, pp. 7002–7012, Feb. 2024.

[5] Y. Jiang, A. Sadeqi, E. L. Miller, and S. Sonkusale, "Head motion classification using thread-based sensor and machine learning algorithm," *Sci. Rep.*, vol. 11, no. 1, 2021, Art. no. 2646.

[6] Z. Zheng et al., "Recovering human pose and shape from through-the-wall radar images," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5112015.

[7] X. Li, Y. He, F. Fioranelli, X. Jing, A. Yarovoy, and Y. Yang, "Human motion recognition with limited radar micro-doppler signatures," *IEEE Trans. Geosci. Remote Sens.*, vol. 59, no. 8, pp. 6586–6599, Aug. 2021.

[8] S. Yang et al., "The human activity radar challenge: Benchmarking based on the 'radar signatures of human activities' dataset from glasgow university," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 4, pp. 1813–1824, Apr. 2023.

[9] M. Zhao et al., "RF-based 3D skeletons," in *Proc. Conf. ACM Special Int. Group Data Commun.*, 2018, pp. 267–281.

[10] A. Vaswani et al., "Attention is all you need," in *Proc. 31st Int. Conf. Neural Inf. Process. Syst.* Red Hook, NY, USA: Curran Associates Inc., 2017, pp. 6000–6010.

[11] C. Xie et al., "Accurate human pose estimation using RF signals," in *Proc. IEEE 24th Int. Workshop Multimedia Signal Process.*, 2022, pp. 1–6.

[12] S.-P. Lee, N. P. Kini, W.-H. Peng, C.-W. Ma, and J.-N. Hwang, "Hupr: A benchmark for human pose estimation using millimeter wave radar," in *Proc. IEEE/CVF Winter Conf. Appl. Comput. Vis.*, 2023, pp. 5704–5713.

[13] A. Sengupta, F. Jin, R. Zhang, and S. Cao, "mm-Pose: Real-time human skeletal posture estimation using mmwave radars and CNNs," *IEEE Sensors J.*, vol. 20, no. 17, pp. 10032–10044, Sep. 2020.

[14] A. Sengupta and S. Cao, "mmPose-NLP: A natural language processing approach to precise skeletal pose estimation using mmwave radars," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 34, no. 11, pp. 8418–8429, Nov. 2023.

[15] S. Hu, S. Cao, N. Toosizadeh, J. Barton, M. G. Hector, and M. J. Fain, "mmPose-FK: A forward kinematics approach to dynamic skeletal pose estimation using mmwave radars," *IEEE Sensors J.*, vol. 24, no. 5, pp. 6469–6481, Mar. 2024.

[16] C. Shi, L. Lu, J. Liu, Y. Wang, Y. Chen, and J. Yu, "mPose: Environment- and subject-agnostic 3 D skeleton posture reconstruction leveraging a single mmwave device," *Smart Health*, vol. 23, 2022, Art. no. 100228.

[17] H. Xue et al., *Towards Generalized Mmwave-Based Human Pose Estimation Through Signal Augmentation*. New York, NY, USA: Association for Computing Machinery, 2023.

[18] A. Gharamohammadi, A. Khajepour, and G. Shaker, "In-vehicle monitoring by radar: A review," *IEEE Sensors J.*, vol. 23, no. 21, pp. 25650–25672, Nov. 2023.

[19] C. Ding et al., "Inattentive driving behavior detection based on portable FMCW radar," *IEEE Trans. Microw. Theory Techn.*, vol. 67, no. 10, pp. 4031–4041, Oct. 2019.

[20] R. Chae, A. Wang, and C. Li, "Fmcw radar driver head motion monitoring based on doppler spectrogram and range-doppler evolution," in *Proc. IEEE Topical Conf. Wireless Sensors Sensor Netw.*, 2019, pp. 1–4.

[21] D. G. Bresnahan and Y. Li, "Classification of driver head motions using a mm-wave FMCW radar and deep convolutional neural network," *IEEE Access*, vol. 9, pp. 100472–100479, 2021.

[22] J. Jung, S. Lim, B.-K. Kim, and S. Lee, "CNN-based driver monitoring using millimeter-wave radar sensor," *IEEE Sens. Lett.*, vol. 5, no. 3, pp. 1–4, Mar. 2021.

[23] C. Xu, X. Zheng, Z. Ren, L. Liu, and H. Ma, "Uhead: Driver attention monitoring system using UWB radar," *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, vol. 8, no. 1, pp. 1–28, Mar. 2024.

[24] N. Kumchaiseemak et al., "Toward ant-sized moving object localization using deep learning in FMCW radar: A pilot study," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, 2022, Art. no. 5112510.

[25] L. Zhang et al., "PeakConv: Learning peak receptive field for radar semantic segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, 2023, pp. 17577–17586.

[26] G. L. Charvat, *Small and Short-Range Radar Systems*. Boca Raton, FL, USA: CRC Press, 2014.

[27] S. Yuan, F. Fioranelli, and A. G. Yarovoy, "3DRUDAT: 3D robust unambiguous doppler beam sharpening using adaptive threshold for forward-looking region," *IEEE Trans. Radar Syst.*, vol. 2, pp. 138–153, 2024.

[28] T. Martyniuk, O. Kupyn, Y. Kurlyak, I. Krashenyi, J. Matas, and V. Sharmanska, "DAD-3DHeads: A large-scale dense, accurate and diverse dataset for 3D head alignment from a single image," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2022, pp. 20942–20952.

[29] S. Garrido-Jurado, R. Muñoz-Salinas, F. Madrid-Cuevas, and M. Marín-Jiménez, "Automatic generation and detection of highly reliable fiducial markers under occlusion," *Pattern Recognit.*, vol. 47, no. 6, pp. 2280–2292, 2014.

[30] J. W. Smith, O. Furxhi, and M. Torlak, "An FCNN-based super-resolution mmwave radar framework for contactless musical instrument interface," *IEEE Trans. Multimedia*, vol. 24, pp. 2315–2328, 2022.

[31] S. Scholes, A. Ruget, F. Zhu, and J. Leach, "Human pose inference using an elevated mmwave FMCW radar," *IEEE Access*, vol. 12, pp. 115605–115614, 2024.

[32] J. Jung, J. Kim, S.-C. Kim, and S. Lim, "Eye-gaze tracking based on head orientation estimation using FMCW radar sensor," *IEEE Trans. Instrum. Meas.*, vol. 73, 2024, Art. no. 8508510.

Nakorn Kumchaiseemak (Graduate Student Member, IEEE) received the M.S. degree in physics from Kasetsart University, Bangkok, Thailand, in 2017. He is currently working toward the Ph.D. degree in information science and technology with the School of Information Science and Technology (IST), Vidyasirimedhi Institute of Science and Technology (VISTEC), Rayong, Thailand.

In 2023, he served as a Visiting Scholar with the Microwave Sensing, Signals and Systems (MS3) group, Delft University of Technology, Delft, The Netherlands. Throughout his M.S. and Ph.D. studies, he has authored and coauthored more than 15 journal and conference papers. His research interests include radar signal processing, human–computer interaction, and deep learning.

Francesco Fioranelli (Senior Member, IEEE) received the Ph.D. degree in electronic engineering with Durham University, Durham, U.K., in 2014.

He was a Research Associate with the University College London in 2014–2016, and an Assistant Professor with the University of Glasgow, in 2016–2019. He is currently an Associate Professor with the Delft University of Technology, Delft, The Netherlands. He has authored more than 190 peer-reviewed publications, and edited the books on *Micro-Doppler Radar and Its Applications* and *Radar Countermeasures for Unmanned Aerial Vehicles* (IET-Scitech). His research interests include radar systems and automatic classification for human signatures analysis in healthcare and security, unmanned aerial vehicles (UAVs) detection and classification, and automotive radar.

Dr. Fioranelli was the recipient of four best paper awards and the IEEE AESS Fred Nathanson Memorial Radar Award 2024.

Theerawit Wilaiprasitporn (Senior Member, IEEE) received the Ph.D. degree in engineering from the Graduate School of Information Science and Engineering, Tokyo Institute of Technology, Tokyo, Japan, in 2017.

He specializes in medical artificial intelligence (AI) and is a dedicated advocate for deep tech startups. He founded Interfaces, an AI-driven health research team with the Vidyasirimedhi Institute of Science and Technology (VISTEC), and SensAI, a platform focused on AI-driven anomaly sensing for improved health outcomes.

Dr. Wilaiprasitporn was the recipient of the prestigious Young Scientist Award 2024 by the Foundation for the Promotion of Science and Technology under the Patronage of His Majesty the King, recognizing him as one of Thailand's leading young professionals. His contributions have been instrumental in developing remote health monitoring systems, which supported more than 30 000 individuals during the COVID-19 pandemic, which led to his nomination for the 2022 IEEE R10 Humanitarian Technology Activities Outstanding Volunteer Award.