RadProPoser: Probabilistic Radar Tensor Human Pose Estimation That Knows Its Limits

Jonas Leo Müller[1,2,*], Lukas Engel[3], Eva Dorschky[1], Daniel Krauss[1,2],
Ingrid Ullmann[3], Martin Vossiek[3], Björn M. Eskofier[1,2]
[1]Machine Learning and Data Analytics Lab, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
[2]Munich Center for Machine Learning, Germany
[3]Institute of Microwaves and Photonics, Friedrich-Alexander-Universität Erlangen-Nürnberg, Germany
[*]Corresponding author: jonas.leo.mueller@fau.de

***Abstract*—Radar-based human pose estimation enables privacy-preserving motion tracking for ambient intelligence, yet the noisy nature of radar sensing makes uncertainty quantification essential. We present RadProPoser, an end-to-end probabilistic framework that predicts three-dimensional body joints with per-joint uncertainties from raw radar tensor data. Using a variational encoder-decoder with spectral attention that fuses real and imaginary radar components across temporal frames, we model aleatoric uncertainty through learnable Gaussian and Laplace distributions. Trained on a new benchmark dataset with optical motion-capture ground truth, our method achieves 6.425 cm mean per-joint position error. The model outputs per-joint aleatoric uncertainties, and isotonic recalibration yields calibrated total uncertainty with expected calibration error of 0.027. Since spectral attention operates on individual radar tensor components, extending to multi-radar configurations requires only concatenating additional input streams. On the HuPR benchmark with dual orthogonal radars, this achieves 5.042 cm MPJPE. The framework runs at ∼89 frames per second (FPS) on an NVIDIA RTX 3090, exceeding the 15 Hz radar frame rate.**

***Index Terms*—radar-based human pose estimation, variational inference, probabilistic modeling, uncertainty quantification, smart environments**

I. Introduction

Ambient intelligence envisions environments that adapt to human presence, providing context-aware services for healthcare applications such as rehabilitation monitoring, elderly care, and fall detection [1], [2]. Radar-based sensing offers unique advantages for this domain. Unlike cameras, radar operates independently of lighting, functions through occlusions, and preserves privacy by capturing motion signatures rather than identifiable imagery [3], [4]. However, radar measurements are inherently stochastic, affected by thermal noise, phase instabilities [5], and multipath propagation [6], making uncertainty quantification essential. Knowing *when* a prediction can be trusted is as important as the prediction itself [7].

Prior radar-based human pose estimation (HPE) has pursued two approaches, point cloud methods using Constant False Alarm Rate (CFAR) detection [8], [9], and radar tensor approaches (RTA) that learn end-to-end from raw measurements [3], [4], [10]. However, existing RTA benchmarks rely on RGB-based ground truth, introducing uncertainty from prediction errors and depth estimation [11] that confounds radar-specific uncertainty measurement. Furthermore, prior work has focused on keypoint accuracy alone with limited attention to explicit probabilistic uncertainty modeling [12], [13].

We present RadProPoser (Fig. 1), a framework for uncertainty quantification in radar-based pose estimation. We introduce a benchmark dataset pairing raw radar measurements with optical motion capture ground truth enabled by a synchronized hardware trigger, ensuring uncertainties reflect the radar sensing process itself rather than image-based preprocessing artifacts. Building on Ho et al. [3], we extend the backbone with a spectral attention mechanism in complex frequency space that fuses real and imaginary radar components across temporal frames. Since spectral attention operates on individual radar tensor components, extending to multi-radar configurations [4] requires only concatenating additional input streams. Predictive uncertainty comprises two components [14]. Aleatoric uncertainty is the irreducible ambiguity inherent in the data, here radar noise and multipath effects. Epistemic uncertainty is the reducible uncertainty from limited training data or model capacity. Using variational inference, we learn distributions over latent pose representations, where sampling and decoding multiple latent codes yields predictive distributions whose variance captures aleatoric uncertainty, which dominates in radar sensing where measurements are stochastic. We focus on aleatoric uncertainty modeling as it reflects fundamental hardware limitations (thermal noise, phase instabilities, multipath propagation) that persist regardless of dataset size, making these estimates actionable for system design and sensor placement. We extend these estimates to calibrated total uncertainty through isotonic recalibration [15] on held-out data.

Our contributions:

- A systematic uncertainty quantification framework for radar tensor-based HPE with our new benchmark dataset.
- A spectral attention mechanism for fusing complex radar components that extends to multi-radar configurations, achieving 6.425 cm MPJPE (single-radar) and 5.042 cm MPJPE (dual-radar).

This work was partly funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1483 – Project-ID 442419336, *EmpkinS*.

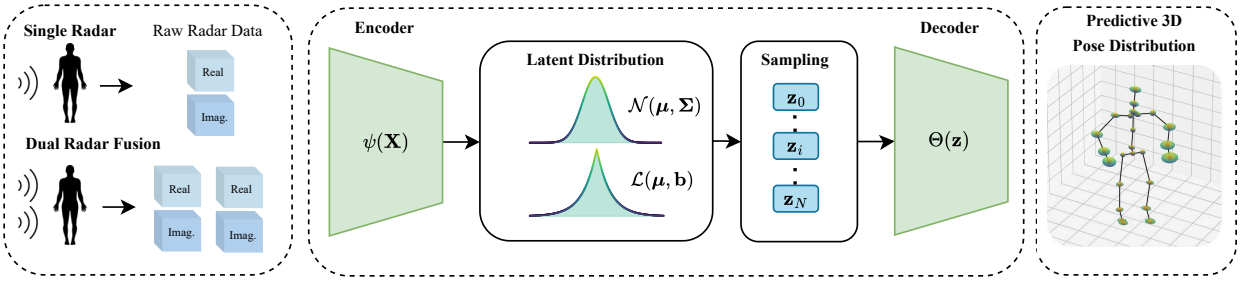


Fig. 1: Overview of our approach. An efficient 3D convolutional encoder $\psi(\cdot)$ maps raw radar data $\mathbf{X}$ to a distribution over latent codes. Samples from this distribution are decoded in parallel by $\Theta(\cdot)$ to yield a predictive distribution over 3D poses. We evaluate both Gaussian and Laplace latent priors. Right: predicted pose means with uncertainty visualized as hulls spanning one standard deviation along each axis.

- Calibrated per-joint uncertainty estimates (ECE of 0.027).

Our code and dataset are available at https://github.com/jonasmueler/RadProPoser_IJCNN and https://zenodo.org/records/14738837.

II. Background and Related Work

A. Radar-based Human Pose Estimation

Radar-based HPE approaches vary in preprocessing complexity, from raw analog-to-digital converter (ADC) data [16] to magnitude-based Fast Fourier Transform (FFT) slicing [4] to compressed point clouds [17], with recent work integrating learnable signal processing [18], [19]. Methods also differ by supervision. Lee et al. [4] use image-space ground truth with 2D-to-3D lifting [20], achieving 6.82 cm MPJPE, while Ho et al. [3] use LiDAR/RGB-derived 3D ground truth with root joint classification and offset regression, achieving 9.91 cm. Most methods favor direct 3D regression from compressed tensors [21], [22]. We adopt end-to-end regression from FFT-processed complex-valued tensors (Table I), ensuring uncertainty estimates reflect data characteristics rather than upstream processing artifacts.

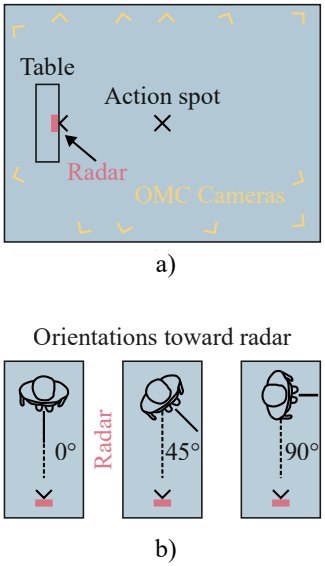
B. Uncertainty Modeling in HPE

Vision-based HPE research has progressed along two directions. *Uncertainty-aware methods* use uncertainty during training [23], [24] or generative techniques [25], [26]. In the radar domain, Chiang et al. [27] apply uncertainty-aware training to point cloud-based pose estimation. *Uncertainty quantification methods*, by contrast, explicitly model predictive uncertainties at inference. The latter includes normalizing flows [28], [29], angular distributions [30], and evidential regression [31]. Direct aleatoric uncertainty regression via Gaussian [14] or Laplacian [28] distributions has proven effective. Despite this progress, uncertainty quantification for RTA-based HPE remains underexplored.

III. Methods

A. Datasets

Our dataset, recorded using the setup of Engel et al. [32], contains radar recordings time-synchronized with optical motion capture (OMC) from 12 participants performing nine



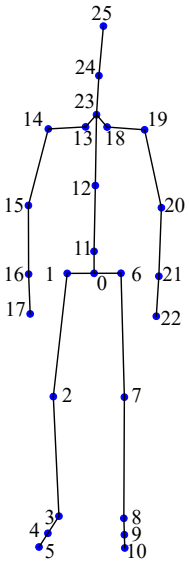


Fig. 2: Left: Recording space. Right: Skeleton keypoint structure predicted by RadProPoser.

exercises at three angles (0°, 45°, 90°). The exercises include left/right/bilateral upper limb extension, bicep curls, front arm rotation, trunk forward bending, left/right front lunge, and squats. We use a Texas Instruments IWR6843AOPEVM 60 GHz radar (Table II) with ground truth from an Optitrack Flex 13 system.

RadProPoser predicts 26 3D keypoints (Fig. 2). Data is split by participant with eight for training, one for hyperparameter tuning (p3), and three for testing (p1, p2, p12). After hyperparameter tuning, models are retrained on all nine non-test participants. Following standard recalibration procedure [15], we fit isotonic recalibration on a separate test participant (p1) and evaluate on the remaining two (p2, p12), avoiding data leakage by keeping hyperparameter tuning (p3) and recalibration fitting disjoint.

To demonstrate multi-radar fusion, we evaluate on

HuPR [4], which uses two orthogonal Texas Instruments IWR1443 radars with only 2D image-space annotations. The original HuPR method predicts 2D keypoints from radar, then lifts to 3D using VideoPose3D [33]. Since no code is provided for 3D ground truth or prediction, we apply MotionBERT [20], a 2D-to-3D lifting model achieving 1 cm lower MPJPE than VideoPose3D on Human3.6M, to both the 2D annotations and HuPR's released 2D predictions, ensuring identical 3D ground truth for fair comparison. We convert HuPR's 14 keypoints to MotionBERT's 17-joint Human3.6M format via direct mapping and interpolation of pelvis, spine, and thorax. We apply our complex-valued preprocessing (Section III-B1) and train RadProPoser end-to-end using the original train/val/test split.

B. RadProPoser

The **RAD**ar **PRO**babilistic **POSER** model (RPP) architecture (Fig. 3) is a function $\phi(\mathbf{I})$ that maps the raw complex-valued radar cube $\mathbf{I} \in \mathbb{C}^{N_{\text{Frames}} \times \text{ADC} \times \text{chirps} \times \text{azimuth} \times \text{elevation}}$ into a distribution over K 3D pose coordinates $\tilde{\boldsymbol{y}} \in \mathbb{R}^{K \times 3}$ using a deterministic encoder and probabilistic decoder, where $K = 26$ for our dataset and $K = 17$ for HuPR [4]. This design balances computational efficiency with probabilistic modeling.

1) Preprocessing Block: The raw radar cube $\mathbf{I}$ contains complex-valued samples organized by ADC samples, chirps, and spatial antenna axes (Table II). Static clutter is removed by centering the complex radar cube in the chirp dimension, followed by a 4D FFT. This yields $\mathbf{X} \in \mathbb{R}^{2 \cdot N_{\text{Frames}} \times \text{Doppler} \times \text{azimuth} \times \text{elevation} \times \text{ADC}}$, where real and imaginary components are stacked along the frame dimension. We use $N_{\text{Frames}} = 8$ (Table VII). For multi-radar configurations (e.g., HuPR [4]), each radar produces its own tensor $\mathbf{X}_i$, stacked to form $\mathbf{X} \in \mathbb{R}^{2 \cdot N_{\text{Radars}} \cdot N_{\text{Frames}} \times \text{Doppler} \times \text{azimuth} \times \text{elevation} \times \text{ADC}}$ for parallel encoding.

2) Encoder: The encoder $\psi(\cdot)$ maps $\mathbf{X}$ to frame-wise features for real and imaginary components through a shared 3D Convolutional Neural Network (CNN) (Fig. 3), using the 3D HRNet backbone from Ho et al. [3], [34], [35]. Features from all resolution branches are upsampled, concatenated, and compressed via two 3D conv layers with batch normalization and ReLU, producing $\mathbf{F} \in \mathbb{R}^{2 \cdot N_{\text{Frames}} \times \text{Embed. Dim}}$.

3) Spectral Attention: To capture temporal and real/imaginary dependencies, we introduce spectral attention with learnable positional encodings. We implement the attention layer in a standard way with residual connection and MLP. Frame features are projected and positional encodings $\mathbf{P} \in \mathbb{R}^{2 \cdot N_{\text{Frames}} \times d_k}$ are added to queries and keys:

$$\mathbf{Q} = \mathbf{F}\mathbf{W}_Q + \mathbf{P}, \quad \mathbf{K} = \mathbf{F}\mathbf{W}_K + \mathbf{P}, \quad \mathbf{V} = \mathbf{F}\mathbf{W}_V, \quad (1)$$

where $\mathbf{W}_Q$, $\mathbf{W}_K$, $\mathbf{W}_V$ are learnable projections. The learned encodings project features from different time steps and signal components (real/imaginary) into distinct embedding regions. The attention output is computed as

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}}\right)\mathbf{V}, \quad (2)$$

where d_k is the key dimensionality, enabling each frame to aggregate information across all others while integrating real and imaginary channels.

For multi-radar configurations, features from each radar are concatenated along the frame dimension, yielding $\mathbf{F} \in \mathbb{R}^{2 \cdot N_{\text{Radars}} \cdot N_{\text{Frames}} \times \text{Embed. Dim}}$, with positional encodings $\mathbf{P} \in \mathbb{R}^{2 \cdot N_{\text{Radars}} \cdot N_{\text{Frames}} \times d_k}$ that project each radar into separate embedding regions while enabling cross-radar and temporal fusion. The output $\mathbf{f}$ is obtained by flattening the attention layer output and linear projection.

4) Latent Space: We use variational inference to model a distribution over latent variables conditioned on the radar input, evaluating both isotropic Gaussian and Laplace priors. The spectral attention output $\mathbf{f}$ is projected to produce the latent distribution parameters $(\boldsymbol{\mu}_l, \boldsymbol{\sigma}_l^2)$ for Gaussian or $(\boldsymbol{\mu}_l, \boldsymbol{b}_l)$ for Laplace priors, where $\mathbf{z} \in \mathbb{R}^{256}$ for single-radar and $\mathbf{z} \in \mathbb{R}^{512}$ for dual-radar setups.

For the Gaussian prior, we regularize with Kullback-Leibler (KL) divergence to $\mathcal{N}(\mathbf{0}, \mathbf{I})$

$$\text{KL}_{\mathcal{N}}(\boldsymbol{\mu}_l, \boldsymbol{\sigma}_l^2) = -\frac{1}{2}\left(1 + \log \boldsymbol{\sigma}_l^2 - \boldsymbol{\mu}_l^2 - \boldsymbol{\sigma}_l^2\right). \quad (3)$$

A learnable scaling factor α adjusts sampling variance for downstream tasks. Latent vectors are sampled as

$$\boldsymbol{z}_i = \boldsymbol{\mu}_l + \alpha \boldsymbol{\epsilon}_i \boldsymbol{\sigma}_l, \quad \boldsymbol{\epsilon}_i \sim \mathcal{N}(0, \mathbf{I}). \quad (4)$$

For the Laplace latent, we exploit its sharp and well-behaved form for inverse cumulative distribution function sampling without KL regularization [36], sampling as $\boldsymbol{z}_i \sim \mathcal{L}(\boldsymbol{\mu}_l, \boldsymbol{b}_l)$ where $\boldsymbol{b}_l = \text{softplus}(\tilde{\boldsymbol{b}}_l)$ ensures positivity.

5) Decoder: The decoder computes the output distribution

$$\Theta(\boldsymbol{z}) = \tilde{\boldsymbol{y}}, \quad (5)$$

where Θ is a two-layer linear network. We draw $N = 500$ samples from the latent distribution, yielding $\boldsymbol{z} \in$

TABLE I: Raw Tensor-Based Radar Pose Estimation Datasets

Authors	Ground Truth	Radar Config.	Evaluation	2D/3D	Keypoints	Method	Reported Score
Ho et al. [3]	RGB, LiDAR	12 TX, 16 RX	MPJPE	3D	15	3D-HR-layers	9.91 cm
Lee et al. [4]	RGB	2 × 3 TX, 4 RX	MPJPE	2D	14	Multi-stage feature fusion	6.82 cm
Ours	OMC	3 TX, 4 RX	MPJPE	3D	26	Prob. modeling	6.425 cm

MPJPE: Mean per-joint position error; TX: Transmitter; RX: Receiver; OMC: Optical motion capture.

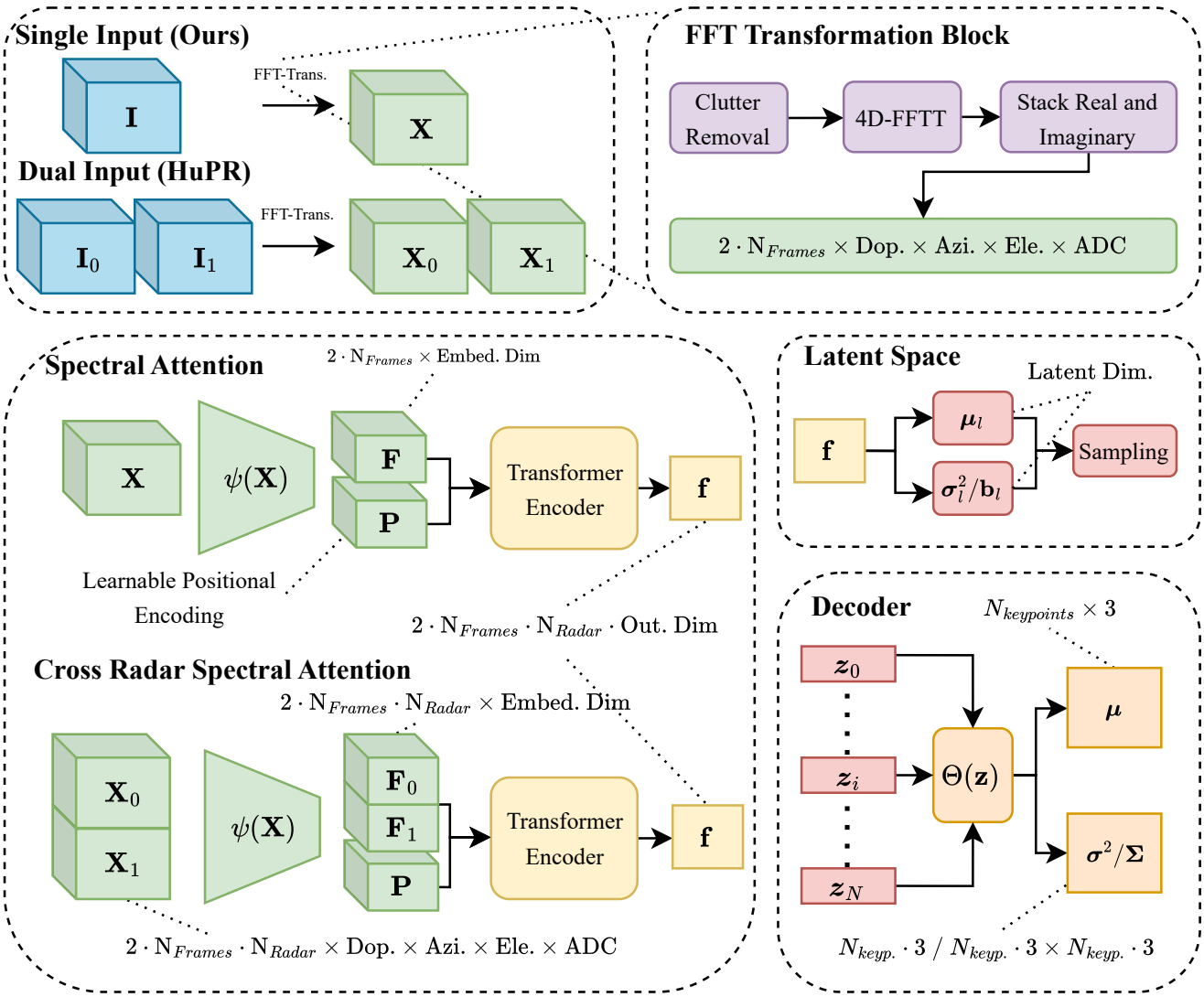


Fig. 3: Overview of the RadProPoser architecture. Top left: input handling for single-radar (ours) and dual-radar (HuPR) configurations. Top right: FFT transformation block with clutter removal. Middle left: spectral attention mechanism with learnable positional encodings. Bottom left: cross-radar spectral attention for multi-sensor fusion. Right: latent space parameterization and decoder producing pose predictions with uncertainty estimates.

TABLE II: Radar System Parameter Settings

Parameter	Value
Frequency	60 GHz
Radio-frequency bandwidth	≈ 1.02 GHz
Frame rate	15 Hz
Chirp duration	≈ 17 μs
Samples per chirp	64
ADC sampling frequency	3.8 MHz
Chirps per frame	128
Number of transmitter antennas	3
Number of receiver antennas	4
TDM-MIMO virtual data channels	12

ADC: analog-to-digital converter; TDM: time-division multiplexing; MIMO: multiple-input multiple-output.

$\mathbb{R}^{N\times\text{Latent Dim.}}$, which the decoder maps to $\tilde{\boldsymbol{y}} \in \mathbb{R}^{N\times K\cdot 3}$. From these we compute mean $\boldsymbol{\mu}$ and variance $\boldsymbol{\sigma}^2 \in \mathbb{R}^{K\cdot 3}$ (or covariance $\boldsymbol{\Sigma}$ when modeling correlations). This keeps the encoder deterministic while enabling parallel sampling via the batch dimension through the probabilistic decoder.

C. Aleatoric Loss Formulations

We model heteroscedastic aleatoric uncertainty [14] by minimizing negative log-likelihood (NLL), comparing predicted mean $\boldsymbol{\mu}$ against ground truth $\boldsymbol{y} \in \mathbb{R}^d$, weighted by predicted uncertainty.

The full training loss combines the NLL with a KL divergence term from the variational latent space:

$$\mathcal{L} = \text{NLL}(\boldsymbol{y}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) + \beta\,\text{KL}(\boldsymbol{\mu}_l, \boldsymbol{\sigma}_l^2), \tag{6}$$

TABLE III: Average mean per-joint position error ↓ of three test-set participants for the three recorded angles in cm.

Model	MPJPE (0°)			MPJPE (45°)			MPJPE (90°)			Avg. MPJPE			Overall Avg.
	P12	P2	P1	P12	P2	P1	P12	P2	P1	P12	P2	P1	
Baselines													
Ho et al. [3]	10.881	8.700	5.946	6.476	6.208	7.438	6.481	6.544	7.945	7.946	7.151	7.110	7.402
Engel et al. [32]†	11.527	8.318	6.643	6.108	6.829	7.662	6.733	7.503	9.695	8.186	7.609	8.031	7.946
Evidential Regression	13.522	12.732	11.285	9.632	11.396	11.086	9.923	11.562	11.396	11.026	11.897	11.256	11.393
RPP Variants (ours)													
RPP-Normalizing-Flows	**7.417**	7.632	7.097	5.788	6.085	7.023	6.639	6.941	7.196	6.947	6.886	7.072	7.022
RPP-Gauss.-Gauss.-Cov.	9.369	**7.228**	**5.027**	**4.922**	5.849	**6.264**	5.635	6.307	7.220	**6.642**	**6.461**	**6.170**	**6.425**
RPP-Gauss.-Gauss.	9.970	8.028	5.127	5.556	**5.710**	6.374	**5.581**	**6.001**	**7.022**	7.036	6.580	6.174	6.597
RPP-Laplace-Laplace	9.988	9.061	7.783	5.591	6.665	7.098	5.683	6.949	7.321	7.087	7.558	7.401	7.349
RPP-Laplace-Gauss.	9.063	7.641	5.203	5.202	5.975	6.583	5.763	6.568	7.166	6.676	6.728	6.317	6.688
RPP-Gauss.-Laplace	10.777	9.491	7.014	5.813	7.015	7.320	6.007	7.163	8.126	7.532	7.890	7.487	7.592

The Models are presented as RPP-[Latent Prior]-[Log-likelihood], with -Cov. indicating full covariance output. †Pointcloud-based approach.

TABLE IV: Procrustes aligned mean per-joint position error ↓ of three test-set participants for the three recorded angles in cm.

Model	P-MPJPE (0°)			P-MPJPE (45°)			P-MPJPE (90°)			Avg. P-MPJPE			Overall Avg.
	P12	P2	P1	P12	P2	P1	P12	P2	P1	P12	P2	P1	
Baselines													
Ho et al. [3]	5.939	6.048	4.524	4.249	4.540	5.559	5.097	5.427	6.315	5.095	5.339	5.466	5.300
Engel et al. [32]†	**5.122**	**5.384**	4.521	4.132	4.638	5.625	5.156	5.797	7.762	4.803	5.273	5.969	5.348
Evidential Regression	10.417	12.130	9.576	9.346	11.718	10.394	9.989	12.118	10.709	9.917	11.989	10.226	10.711
RPP Variants (ours)													
RPP-Normalizing-Flows	6.264	7.028	5.168	4.363	5.503	6.424	5.368	6.082	6.725	5.665	6.204	6.106	6.325
RPP-Gauss.-Gauss.	5.939	6.389	4.764	3.948	4.608	5.290	4.900	5.423	**5.793**	4.929	5.473	5.282	5.228
RPP-Gauss.-Gauss.-Cov.	5.363	5.516	**3.970**	**3.738**	**4.128**	**5.074**	**4.654**	**4.797**	5.918	**4.585**	**4.813**	**4.987**	**4.795**
RPP-Laplace-Laplace	6.601	7.396	5.116	4.374	5.152	5.807	4.870	5.938	6.287	5.282	6.162	5.737	5.727
RPP-Laplace-Gauss.	5.772	6.050	4.237	3.980	4.250	5.416	4.865	5.377	5.942	4.872	5.226	5.198	5.099
RPP-Gauss.-Laplace	6.927	7.950	5.131	4.692	5.680	6.155	5.238	6.186	6.841	5.619	6.605	6.042	6.200

The Models are presented as RPP-[Latent Prior]-[Log-likelihood], with -Cov. indicating full covariance output. †Pointcloud-based approach.

where β weights the KL regularization.

a) Gaussian.: Following Kendall and Gal [14], we assume independence across output dimensions:

$$\mathrm{NLL}_{\mathrm{Gauss.}}(\boldsymbol{y}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) = \sum_{i=1}^{d} \left[\frac{(\boldsymbol{y}_i - \boldsymbol{\mu}_i)^2}{\boldsymbol{\sigma}_i^2} + \gamma \log \boldsymbol{\sigma}_i^2 \right]. \quad (7)$$

b) Full Covariance Gaussian.: To capture output correlations, we use full covariance following Gundavarapu et al. [37]:

$$\mathrm{NLL}_{\mathrm{Cov.}}(\boldsymbol{y}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (\boldsymbol{y} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\boldsymbol{y} - \boldsymbol{\mu}) + \gamma \log |\boldsymbol{\Sigma}| . \quad (8)$$

c) Laplace.: As an alternative to Gaussian assumptions, we consider independent Laplace likelihoods with scale parameters $\boldsymbol{b} \in \mathbb{R}^d$:

$$\mathrm{NLL}_{\mathrm{Lap.}}(\boldsymbol{y}, \boldsymbol{\mu}, \boldsymbol{b}) = \sum_{i=1}^{d} \left[\frac{|\boldsymbol{y}_i - \boldsymbol{\mu}_i|}{\boldsymbol{b}_i} + \gamma \log \boldsymbol{b}_i \right]. \quad (9)$$

This may better capture heavy-tailed errors. In all formulations, γ scales the log-variance term ($\gamma = 1$ recovers standard NLL). All hyperparameters are provided in the code repository.

D. Calibration and Recalibration

Probabilistic models often produce overconfident or underconfident predictions [15], [31]. A distributional predictor is calibrated if predicted confidence levels match empirical frequencies. We quantify miscalibration using expected calibration error (ECE) [15]

$$\mathrm{ECE} = \frac{1}{C} \sum_{j=1}^{C} \left| \hat{P}(p_j) - p_j \right|, \quad (10)$$

where C is the number of confidence levels and $\hat{P}(p_j)$ is the empirical frequency of samples below target coverage level p_j. We also evaluate sharpness, the mean predicted variance

$$\mathrm{Sharpness} = \frac{1}{NK} \sum_{k=1}^{K} \sum_{i=1}^{N} \sigma_{i,k}^2, \quad (11)$$

since well-calibrated but overly wide intervals provide little actionable information. To correct miscalibration, we apply isotonic regression to learn a recalibration mapping $R : [0,1] \to [0,1]$ between the model's predicted cumulative distribution function (CDF) values and empirical coverage probabilities [15]. This non-parametric method enforces monotonicity while fitting the empirical calibration curve. The learned mapping transforms the predicted distributions such that recalibrated confidence levels accurately reflect empirical coverage rates.

IV. Experiments

A. Experimental Setup

We evaluate six RadProPoser (RPP) variants combining two latent spaces (Gaussian, Laplace) with three output likelihoods

TABLE V: Pose Estimation Performance on HuPR Test Set (cm) ↓

Model	GT Source	Elbow		Wrist		Knee		Ankle		Total	
		MPJPE	P-MPJPE	MPJPE	P-MPJPE	MPJPE	P-MPJPE	MPJPE	P-MPJPE	MPJPE	P-MPJPE
mmMesh†	VideoPose3D	11.29	–	21.82	–	–	–	–	–	7.13	–
HuPR†‡	VideoPose3D	8.53	–	15.64	–	–	–	–	–	6.82	–
HuPR‡	MotionBERT	16.883	12.755	31.292	25.696	4.979	6.354	7.226	9.189	9.610	9.233
RPP-Gauss.-Gauss.-Cov.	MotionBERT	**8.168**	**6.065**	**14.612**	**11.886**	**3.244**	**3.056**	**5.538**	**5.262**	**5.042**	**4.741**

†Results taken from original publication. ‡Uses 2D-to-3D lifting at inference. GT Source indicates the lifting model used to generate 3D pseudo-ground-truth from HuPR's 2D annotations; RPP predicts 3D poses directly. Per-joint values are averaged over left and right keypoints.

(Gaussian diagonal, Gaussian full covariance, Laplace). The Laplace latent space is motivated by Li et al. [28], who showed heavier-tailed distributions improve pose estimation. We additionally evaluate a normalizing flow latent space [38] using RealNVP [39], which captures total uncertainty directly without aleatoric decomposition.

As probabilistic baseline, evidential regression [40] provides closed-form aleatoric uncertainty without sampling. As non-probabilistic baselines, we adapt the single-frame HRNet encoder from Ho et al. [3] and retrain the point cloud model from Engel et al. [32] on our split. Lee et al. [4] are excluded from direct comparison due to incompatible dual-radar 2D predictions, but we evaluate on HuPR separately.

Following Zheng et al. [41], we measure pose accuracy using Mean Per-Joint Position Error (MPJPE) and Procrustes-aligned MPJPE (P-MPJPE), which isolates structural accuracy via rigid transformation. For probabilistic models, predictive uncertainty per keypoint is $\boldsymbol{u}_k = \sum_{d=1}^{3} \boldsymbol{\sigma}_{k,d}^2$, with calibration assessed using ECE and sharpness [15].

V. Results and Discussion

We evaluate RadProPoser across four dimensions. First, we compare all model variants on our dataset for pose estimation accuracy (Tables III and IV), uncertainty quantification (Table VIII), and inference efficiency (Table XI). We then evaluate the best-performing configuration (RPP-Gauss.-Gauss.-Cov.) for multi-radar sensor fusion on the HuPR benchmark.

A. Pose Estimation Accuracy

Our best-performing model, RPP-Gauss.-Gauss.-Cov., achieves 6.425 cm MPJPE, representing a 13.2% improvement over the adapted baseline from Ho et al. [3] (7.402 cm) and a 19.1% improvement over the point cloud method by Engel et al. [32] (7.946 cm). Under Procrustes alignment, which isolates structural pose accuracy from global positioning, our model achieves 4.795 cm P-MPJPE compared to 5.300 cm for Ho et al. and 5.348 cm for Engel et al., improvements of 9.5% and 10.3% respectively.

The full-covariance Gaussian formulation outperforms the diagonal variant (6.425 cm vs. 6.597 cm) by modeling correlations between joint positions. The Normalizing Flow model achieves 7.022 cm, while evidential regression performs worse (11.393 cm). All multi-frame RadProPoser variants outperform the single-frame baseline from Ho et al. [3].

Table VI shows spectral attention improves performance on both single and dual radar configurations.

TABLE VI: Spectral Attention Ablation

Configuration	Spec. Att.	MPJPE	P-MPJPE
Our dataset (single radar)	×	7.992	5.733
Our dataset (single radar)	✓	6.425	4.795
HuPR (dual radar)	×	5.999	5.833
HuPR (dual radar)	✓	5.042	4.741

MPJPE and P-MPJPE in cm.

Table VII shows the impact of frame count and latent dimensionality, where 8 frames and 256-dimensional latent space yield optimal results.

TABLE VII: Impact of Frame Count and Latent Dimensionality

	$d = 256$			$t = 8$	
t / d	2	4	8	512	1024
MPJPE (cm)	7.357	7.215	6.425	6.961	7.332

$t = N_{\text{Frames}}$, d = Latent Dim.

B. Multi-Radar Sensor Fusion on HuPR

To demonstrate generalization to different radar configurations, we train RPP-Gauss.-Gauss.-Cov. on the HuPR benchmark [4] with dual orthogonally-mounted radars (Table V). Our model achieves 5.042 cm MPJPE compared to HuPR's 9.610 cm with identical MotionBERT [20] lifting and using the given model weights for 2D prediction, a 47.5% improvement. This gain reflects our spectral attention mechanism effectively fusing multi-radar inputs and our preprocessing preserving full complex-valued signal content. The improvement is particularly pronounced for challenging keypoints with wrist error decreasing from 31.292 cm to 14.612 cm (53.3% reduction) and elbow error from 16.883 cm to 8.168 cm (51.6% reduction), validating our end-to-end approach.

C. Uncertainty Quantification and Calibration

Table VIII summarizes calibration performance. Prior to recalibration, models exhibit moderate miscalibration with ECE ranging from 0.078 to 0.185. After isotonic recalibration,

TABLE VIII: Calibration and Sharpness on Our Test Set

Model	ECE	ECE (Cal.)	Sharpness	Sharpness (Cal.)
RPP-Gauss.-Gauss.-Cov.	0.140	0.027	6.290	40.884
RPP-Gauss.-Gauss.	0.163	0.031	2.146	16.871
RPP-Laplace-Gauss.	0.171	0.022	1.938	12.964
RPP-Gauss.-Laplace	0.185	0.021	1.532	14.771
RPP-Laplace-Laplace	0.154	0.026	8.970	58.372
RPP-Normalizing-Flows	0.078	0.044	28.623	57.061
Evidential Regression	0.180	0.042	9054.15	4906.734

RPP models are presented as RPP-[Latent Prior]-[Log-likelihood], with -Cov. indicating full covariance output. Sharpness in cm^2. Cal.: calibrated.

all models achieve ECE below 0.05, with the best result of ECE = 0.021 (RPP-Gauss.-Laplace). RPP-Gauss.-Gauss.-Cov. achieves ECE = 0.027 while maintaining the highest pose accuracy, offering flexibility between raw aleatoric estimates and calibrated total uncertainty for downstream decision-making. The sharpness values reflect this correction. RPP-Gauss.-Gauss.-Cov. increases from 6.290 to 40.884 cm² as overconfident predictions are relaxed to achieve accurate coverage. Per-

TABLE IX: MPJPE and Uncertainty by Body Region

	Center (n=6)	Left Leg (n=5)	Right Leg (n=5)	Left Arm (n=5)	Right Arm (n=5)
MPJPE	5.649	5.069	5.379	9.115	8.077
Uncal.	4.825	4.065	3.861	10.149	8.845
Cal.	28.526	25.606	20.847	72.318	59.596

MPJPE in cm, uncertainty (Uncal./Cal.) in cm². Values averaged over keypoints per region. Results for RPP-Gauss.-Gauss.-Cov.

keypoint analysis (Table IX) shows that learned uncertainties reflect physical signal properties. Extremities with smaller reflective surfaces exhibit higher errors and proportionally higher uncertainties, while torso keypoints show the lowest of both. This approach generalizes to HuPR, where RPP-Gauss.-Gauss.-Cov. improves from ECE 0.103 to 0.023 after recalibration (Table X). While sharpness increases from 5.991 to 22.371 cm² after recalibration, the uncertainties remain informative. The lower overall sharpness compared to our single-radar setup likely reflects the additional information from dual-radar fusion. Wrist predictions show the highest sharpness (94.554 cm²), consistent with known limitations of 2D-to-3D lifting models [20] for arm estimation. The calibrated uncertainties thus capture errors from both radar sensing and the lifting model used to generate ground truth.

TABLE X: Calibration and Sharpness on HuPR Test Set

Method	ECE ↓	Elbow	Wrist	Knee	Ankle	Total
Uncalibrated	0.103	7.686	19.183	4.172	9.786	5.991
Calibrated	0.023	30.718	94.554	9.607	25.952	22.371

Sharpness in cm². Per-joint values are averaged over left and right keypoints.

TABLE XI: Inference Time Analysis

Model	Params (M)	FLOPs (G)	Time (ms)
RPP-Gauss.-Gauss.-Cov.	9.514	37.841	11.204
RPP-Normalizing-Flows	10.04	38.359	12.152
Evidential Regression	9.019	37.791	11.048
Ho et al. [3]	7.762	2.357	4.286

Single-radar configuration. Batch size = 1, averaged over 50 repetitions, G = giga, M = million.

D. Inference Efficiency

RadProPoser achieves real-time performance at approximately 11 ms per frame (∼89 FPS) on an NVIDIA RTX 3090 GPU, exceeding the 15 Hz radar frame rate (Table XI). The encoder processes input frames in parallel through 3D convolutions. We then draw $N = 500$ samples from the latent distribution, stack them in the batch dimension, and process them through the decoder in one parallel forward pass to obtain the full output distribution. The method of Ho et al. [3] is faster (4.286 ms) due to single-frame processing, but at reduced accuracy.

E. Limitations and Future Directions

Our dataset was recorded in a controlled laboratory environment to ensure reliable uncertainty estimation, while evaluation on HuPR demonstrates generalization across sensor configurations. Future work will explore uncertainty-guided data simulation [42], and employ the calibrated per-joint uncertainties for downstream applications including biomechanical simulations [43] and digital physiotherapy [44].

VI. Conclusion

This work introduces RadProPoser, the first uncertainty quantification framework for radar-based human pose estimation. Through variational inference with full covariance modeling and spectral attention for multi-radar fusion, RadProPoser provides calibrated per-joint uncertainties that reflect physical signal properties. The framework generalizes across sensor configurations and establishes radar as a viable modality for trustworthy ambient intelligence systems that enhance human wellbeing without compromising privacy.

References

[1] F. Jin, A. Sengupta, and S. Cao, "mmfall: Fall detection using 4-d mmwave radar and a hybrid variational rnn autoencoder," *IEEE Transactions on Automation Science and Engineering*, vol. 19, no. 2, pp. 1245–1257, 2020.

[2] D. Krauss, L. Engel, T. Ott, J. Bräunig, R. Richer, M. Gambietz, N. Albrecht, E. M. Hille, I. Ullmann, M. Braun, P. Dabrock, A. Kölpin, A. D. Koelewijn, B. M. Eskofier, and M. Vossiek, "A review and tutorial on machine learning-enabled radar-based biomedical monitoring," *IEEE Open Journal of Engineering in Medicine and Biology*, vol. 5, pp. 680 – 699, 2024.

[3] Y.-H. Ho, J.-H. Cheng, S. Y. Kuan, Z. Jiang, W. Chai, H.-W. Huang, C.-L. Lin, and J.-N. Hwang, "Rt-pose: A 4d radar tensor-based 3d human pose estimation and localization benchmark," in *European Conference on Computer Vision*. Springer, 2024, pp. 107–125.

[4] 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 *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 2023, pp. 5715–5724.

[5] K. Thurn, R. Ebelt, and M. Vossiek, "Noise in homodyne fmcw radar systems and its effects on ranging precision," in *2013 IEEE MTT-S International Microwave Symposium Digest (MTT)*. IEEE, 2013, pp. 1–3.

[6] M. A. Richards, *Fundamentals of radar signal processing*. Mcgraw-hill New York, 2005, vol. 1.

[7] M. Abdar, F. Pourpanah, S. Hussain, D. Rezazadegan, L. Liu, M. Ghavamzadeh, P. Fieguth, X. Cao, A. Khosravi, and U. R. Acharya, "A review of uncertainty quantification in deep learning: Techniques, applications and challenges," *Information fusion*, vol. 76, pp. 243–297, 2021.

[8] M. I. Skolnik, *Introduction to radar systems*. McGraw-hill New York, 1980, vol. 3.

[9] A. Sengupta, F. Jin, and S. Cao, "Nlp based skeletal pose estimation using mmwave radar point-cloud: A simulation approach," in *2020 IEEE Radar Conference (RadarConf20)*. IEEE, 2020, pp. 1–6.

[10] J. Müller, D. Krauß, L. Engel, M. Vossiek, B. Eskofier, and E. Dorschky, "End-to-End Learning for Human Pose Estimation from Raw Millimeter Wave Radar Data," in *Asilomar Conference on Signals, Systems, and Computers*, 2024.

[11] S. Zhang, C. Wang, W. Dong, and B. Fan, "A survey on depth ambiguity of 3d human pose estimation," *Applied Sciences*, vol. 12, no. 20, p. 10591, 2022.

[12] Z. Zheng, J. Pan, Z. Ni, C. Shi, D. Zhang, X. Liu, and G. Fang, "Recovering human pose and shape from through-the-wall radar images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–15, 2022.

[13] J. Fan, J. Yang, Y. Xu, and L. Xie, "Diffusion model is a good pose estimator from 3d rf-vision," in *European Conference on Computer Vision*. Springer, 2025, pp. 1–18.

[14] A. Kendall and Y. Gal, "What uncertainties do we need in bayesian deep learning for computer vision?" *Advances in neural information processing systems*, vol. 30, 2017.

[15] V. Kuleshov, N. Fenner, and S. Ermon, "Accurate uncertainties for deep learning using calibrated regression," in *International conference on machine learning*. PMLR, 2018, pp. 2796–2804.

[16] P. Zhao, C. X. Lu, B. Wang, N. Trigoni, and A. Markham, "Cubelearn: End-to-end learning for human motion recognition from raw mmwave radar signals," *IEEE Internet of Things Journal*, vol. 10, no. 12, pp. 10 236–10 249, 2023.

[17] A. Sengupta, F. Jin, R. Zhang, and S. Cao, "mm-pose: Real-time human skeletal posture estimation using mmwave radars and cnns," *IEEE Sensors Journal*, vol. 20, no. 17, pp. 10 032–10 044, 2020.

[18] J. Giroux, M. Bouchard, and R. Laganiere, "T-fftradnet: Object detection with swin vision transformers from raw adc radar signals," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 4030–4039.

[19] L. Chen and G. Wang, "Cpformer: End-to-end multi-person human pose estimation from raw radar cubes with transformers," *IEEE Sensors Journal*, 2025.

[20] W. Zhu, X. Ma, Z. Liu, L. Liu, W. Wu, and Y. Wang, "Motionbert: A unified perspective on learning human motion representations," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2023, pp. 15 085–15 099.

[21] Z. Cao, W. Ding, R. Chen, J. Zhang, X. Guo, and G. Wang, "A joint global–local network for human pose estimation with millimeter wave radar," *IEEE Internet of Things Journal*, vol. 10, no. 1, pp. 434–446, 2022.

[22] Z. Cao, J. Zhang, R. Chen, X. Guo, and G. Wang, "Task-specific feature purifying in radar-based human pose estimation," *IEEE Transactions on Aerospace and Electronic Systems*, 2023.

[23] J. Zhang, Y. Chen, and Z. Tu, "Uncertainty-aware 3d human pose estimation from monocular video," in *Proceedings of the 30th ACM International Conference on Multimedia*, 2022, pp. 5102–5113.

[24] J. N. Kundu, S. Seth, P. YM, V. Jampani, A. Chakraborty, and R. V. Babu, "Uncertainty-aware adaptation for self-supervised 3d human pose estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 20 448–20 459.

[25] J. Gong, L. G. Foo, Z. Fan, Q. Ke, H. Rahmani, and J. Liu, "Diffpose: Toward more reliable 3d pose estimation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 13 041–13 051.

[26] R. Feng, Y. Gao, T. H. E. Tse, X. Ma, and H. J. Chang, "Diffpose: Spatiotemporal diffusion model for video-based human pose estimation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 14 861–14 872.

[27] H.-C. Chiang, G.-H. Li, F. Wang, S. Shirmohammadi, and C.-H. Hsu, "Enhancing skeletal pose estimation from mmwave point clouds through uncertainty reduction," in *Proceedings of the 5th International Workshop on Human-centric Multimedia Analysis*, 2024, pp. 45–53.

[28] J. Li, S. Bian, A. Zeng, C. Wang, B. Pang, W. Liu, and C. Lu, "Human pose regression with residual log-likelihood estimation," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 11 025–11 034.

[29] S. K. Dwivedi, C. Schmid, H. Yi, M. J. Black, and D. Tzionas, "Poco: 3d pose and shape estimation with confidence," in *2024 International Conference on 3D Vision (3DV)*. IEEE, 2024, pp. 85–95.

[30] S. Prokudin, P. Gehler, and S. Nowozin, "Deep directional statistics: Pose estimation with uncertainty quantification," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 534–551.

[31] L. Bramlage, M. Karg, and C. Curio, "Plausible uncertainties for human pose regression," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 15 133–15 142.

[32] L. Engel, J. Mueller, E. J. F. Rendon, E. Dorschky, D. Krauss, I. Ullmann, B. M. Eskofier, and M. Vossiek, "Advanced millimeter wave radar-based human pose estimation enabled by a deep learning neural network trained with optical motion capture ground truth data," *IEEE Journal of Microwaves*, pp. 1–15, 2025.

[33] D. Pavllo, C. Feichtenhofer, D. Grangier, and M. Auli, "3d human pose estimation in video with temporal convolutions and semi-supervised training," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 7753–7762.

[34] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, and X. Wang, "Deep high-resolution representation learning for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 43, no. 10, pp. 3349–3364, 2020.

[35] B. Cheng, B. Xiao, J. Wang, H. Shi, T. S. Huang, and L. Zhang, "Higherhrnet: Scale-aware representation learning for bottom-up human pose estimation," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5386–5395.

[36] I. Habernal, "How reparametrization trick broke differentially-private text representation learning," in *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*. Dublin, Ireland: Association for Computational Linguistics, 2022.

[37] N. B. Gundavarapu, D. Srivastava, R. Mitra, A. Sharma, and A. Jain, "Structured aleatoric uncertainty in human pose estimation." in *CVPR Workshops*, vol. 2, 2019, p. 2.

[38] D. Rezende and S. Mohamed, "Variational inference with normalizing flows," in *International conference on machine learning*. PMLR, 2015, pp. 1530–1538.

[39] L. Dinh, J. Sohl-Dickstein, and S. Bengio, "Density estimation using real nvp," in *International Conference on Learning Representations (ICLR)*, 2017.

[40] A. Amini, W. Schwarting, A. Soleimany, and D. Rus, "Deep evidential regression," *Advances in neural information processing systems*, vol. 33, pp. 14 927–14 937, 2020.

[41] C. Zheng, W. Wu, C. Chen, T. Yang, S. Zhu, J. Shen, N. Kehtarnavaz, and M. Shah, "Deep learning-based human pose estimation: A survey," *ACM Computing Surveys*, vol. 56, no. 1, pp. 1–37, 2023.

[42] C. Schüßler, M. Hoffmann, J. Bräunig, I. Ullmann, R. Ebelt, and M. Vossiek, "A realistic radar ray tracing simulator for large mimo-arrays in automotive environments," *IEEE Journal of Microwaves*, vol. 1, no. 4, pp. 962–974, 2021.

[43] J. M. Wakeling, M. Febrer-Nafría, and F. De Groote, "A review of the efforts to develop muscle and musculoskeletal models for biomechanics in the last 50 years," *Journal of biomechanics*, vol. 155, p. 111657, 2023.

[44] J. L. Mueller, A. Weiss, and B. M. Eskofier, "Adaptive biofeedback for digital physiotherapy using sakoe-chiba constrained pose matching." Berlin, Heidelberg: Springer-Verlag, 2025.