






Uncertainty-Based Fingerprinting Model Monitoring for Radio Localization

Maximilian Stahlke , Tobias Feigl , *Member, IEEE*, Sebastian Kram , Bjoern M. Eskofier , and Christopher Mutschler 

Abstract—Indoor radio environments often consist of areas with mixed propagation conditions. In line-of-sight (LoS)-dominated areas, classic time-of-flight (ToF) methods reliably return accurate positions, while in nonline-of-sight (NLoS) dominated areas (AI-based) fingerprinting methods are required. However, fingerprinting methods are only cost-efficient if they are used exclusively in NLoS-dominated areas due to their expensive life cycle management. Systems that are both accurate and cost-efficient in LoS- and NLoS-dominated areas require identification of those areas to select the optimal localization method. To enable a reliable and robust life cycle management of fingerprinting, we must identify altered fingerprints to trigger update processes. In this article, we propose methods for uncertainty estimation of AI-based fingerprinting to determine its spatial boundaries and validity. Our experiments show that we can successfully identify spatial boundaries of the fingerprinting models and detect corrupted areas. In contrast to the state-of-the-art, our approach employs an intrinsic identification through out-of-distribution (OOD) detection, rendering external detection approaches unnecessary.

Index Terms—5G, fingerprinting, hybrid localization, radio localization, ultra-wideband (UWB), uncertainty quantification.

I. INTRODUCTION

INDOOR positioning enables applications such as monitoring of production facilities or robot localization. There are many indoor localization techniques based on LiDAR [1], ultrasonic [2], or cameras [3], that have high localization accuracy but lack robustness due to dynamic environments, ambient noise, and changing light conditions. Therefore, radio-based localization systems are often employed as they provide robust localization results. Here, time-of-flight (ToF) or time-difference-of-arrival (TDoA) radio systems enable centimeter accuracies [4].

To achieve a robust radio frequency (RF)-based ToF or TDoA localization, line-of-sight (LoS) between transmitter and receivers has to be ensured, which is rarely given in realistic indoor

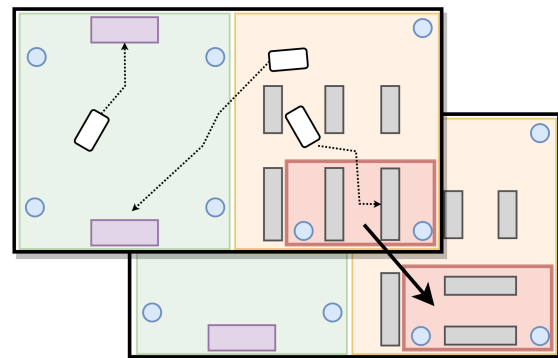


Fig. 1. Schematic top view of a typical (dynamic) industrial environment with a radio localization system.

Manuscript received 2 January 2024; revised 26 March 2024; accepted 22 May 2024. Date of publication 9 May 2024; date of current version 22 May 2024. This work was supported in part by the Federal Ministry of Education and Research of Germany in the programme of “Souverän. Digital. Vernetzt.” joint project 6G-RIC under Grant 16KISK020K and in part by the Bavarian Ministry of Economic Affairs, Regional Development and Energy through the Center for Analytics—Data—Applications (ADA-Center) within the framework of “BAYERN DIGITAL II” under Grant 20-3410-2-9-8. (Corresponding author: Maximilian Stahlke.)

Maximilian Stahlke is with the Fraunhofer Institute for Integrated Circuits IIS, Division Positioning and Networks, Fraunhofer IIS, 90411 Nuremberg, Germany, and also with the Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91052 Erlangen, Germany (e-mail: maximilian.stahlke@iis.fraunhofer.de).

Tobias Feigl, Sebastian Kram, and Christopher Mutschler are with the Fraunhofer IIS, Fraunhofer Institute for Integrated Circuits IIS, Division Positioning and Networks, 90411 Nuremberg, Germany (e-mail: tobias.feigl@iis.fraunhofer.de; sebastian.kram@iis.fraunhofer.de; christopher.mutschler@iis.fraunhofer.de).

Bjoern M. Eskofier is with the Department Artificial Intelligence in Biomedical Engineering, Friedrich-Alexander-Universität Erlangen-Nürnberg, 91052 Erlangen, Germany (e-mail: bjoern.eskofier@fau.de).

Digital Object Identifier 10.1109/JISPIN.2024.3398568

environments. Thus, often error mitigation methods such as nonline-of-sight (NLoS) identification [5], [6], [7], [8] are used to assist RF-based algorithms. While error mitigation methods require at least three transmitters with LoS, fingerprinting methods [9] enable high localization accuracies in NLoS dominated areas, as they do not require LoS conditions. However, the main drawback of fingerprinting is that it requires extensive measurement campaigns with a position reference system with regular updates of the fingerprints when the environment changes [10]. Hence, identification of corrupted fingerprints is crucial to enable reliable localization and trigger update routines [11]. However, as industrial environments may change frequently and the effort of data recording increases significantly with the size of the environment, fingerprinting models may only be employed in NLoS-dominated areas, where it is necessary. Thus, a combination of various localization methods or technologies [12], [13] is useful to ensure a cost-efficient and maintainable localization system.

Fig. 1 shows a typical industrial environment, with machines (purple rectangles) and shelves with goods (gray rectangles). To localize industrial trucks (white rectangles), a radio system with seven receivers (blue circles) is deployed. There are two different sections in this area. For the localization in the machine hall (green area, left), we typically employ classical RF-based positioning approaches, i.e., nondata-driven multilateration or multiangulation methods, as the LoS between the receivers and four transmitters within the area is always guaranteed. However, in the orange area on the right-hand side, the propagation conditions are more challenging for classical localization algorithms as the LoS to the majority of the receivers is blocked. In this area, fingerprinting methods enable high localization accuracy, with the expense of collecting and labeling data. To enable continuous, robust, and cost-efficient localization, a combination of both positioning techniques is necessary. The combination of both methods enables the propagation environment-dependent selection between the classic method in LoS and AI-based fingerprinting in NLoS-dominated areas. However, the selection requires to identify the spatial limitation of the fingerprinting model intrinsically. Environmental changes alter fingerprints, which leads to a degradation in the localization performance [11]. The position of the large shelves in the red area in Fig. 1 changed. This leads to a change in the radio fingerprints as the channel state (LoS/NLoS) and reflections change. So, the fingerprinting model performs badly in this particular area. Hence, we must identify corrupted areas to trigger a retraining process to keep fingerprinting models up-to-date.

To address these challenges, in our previous work [14], we showed how uncertainty estimation of neural networks identifies samples out of the training data area and thus, enables a combination of classic radio-based positioning and fingerprinting methods. We formalize and extend our previous work [14] by examining the spatial limitations of the fingerprint model and the identification of obsolete fingerprints. We call this extension out-of-distribution (OOD). We provide an additional dataset of a 5G radio system with a lower bandwidth compared to ultra-wideband (UWB) to investigate the influence of bandwidth on identification performance. Finally, we evaluate different uncertainty approaches regarding their ability to identify OOD samples. In contrast to state-of-the-art methods, our approach identifies the limitations of the fingerprinting model intrinsically, rendering an additional detection and localization system unnecessary.

The rest of this article is organized as follows. Section II reviews related work. We introduce our novel approach in Section III. Section IV describes the experiments, and Section V discusses the results. Finally, Section VII concludes this article.

II. RELATED WORK

Radio localization achieves high localization accuracies in the centimeter range, even in areas affected by multipath transmission, making it a very good candidate for high-precision localization [15]. However, due to blockage of LoS, classical positioning algorithms are restricted to pure LoS or areas with minor NLoS propagation with the assistance of error mitigation

methods [5]. To enable localization in NLoS-dominated areas, fingerprinting improves the localization accuracy significantly as it benefits from the complexity of the radio channel to create unique radio fingerprints for positioning [16]. There exist both feature-based fingerprinting methods [17], [18] and methods that employ the raw channel state information (CSI) to exploit more information [19], [20]. However, fingerprinting methods require a large training database to cover the entire localization area and high maintenance overhead due to environmental changes [11].

The fusion of classical approaches with data-driven [deep learning (DL)]-based fingerprinting methods helps to achieve a stable and reliable positioning system that is cost efficient. Some approaches parallelize fingerprinting and multilateration by fusing individual Kalman filters with their state covariance matrix, which represents the uncertainty of the position estimates [21]. He et al. [22] employed TDoA localization to identify fingerprinting areas in an LTE network. Bite et al. [23] employed other sensors for RSS fingerprinting area selection, such as pyroelectric infrared sensors. However, the quality of these area selection techniques is restricted to the coarse localization accuracy of the alternative algorithm and system. Instead, we complement fingerprinting with classic localization using the same radio system. In pure LoS environments, we classic localization reliably identifies the fingerprint area, else fingerprinting intrinsically identifies (NLoS) areas.

Techniques that build on DL-based fingerprinting employ uncertainty-based methods, that provide an intrinsic reliability estimation of the predicted position. Foliadis et al. [24] proposed a late fusion approach, that averages fingerprinting results w.r.t. their uncertainty. Russel et al. [25] proposed a direct integration of the predictive uncertainty into a Kalman filter. However, their work focuses on modeling the heteroscedastic uncertainty within the domain, while our work focuses on the detection of OOD samples.

There are several approaches to model the uncertainty of neural networks, such as Monte Carlo (MC) dropout [26], stochastic weight averaging Gaussian (SWAG) [27], or Laplace approximations [28] applied to the full network [29], subnetworks [30], or just the last layer [31]. Especially ensemble-based methods, e.g., based on multiple SWAG models [32] (MultiSWAG) or deterministic neural networks [33], modeling the epistemic uncertainty (model uncertainty) well due to their versatile modeling of the loss landscape [32].

III. METHODOLOGY

We give an overview of uncertainty estimation for neural networks and describe how to detect OOD data samples based on uncertainty estimation in the following.

A. Uncertainty Estimation

In radio-based fingerprint localization, there are several sources of uncertainty, that are not reflected in the estimation of our deterministic neural network and lead to high prediction errors [11]. To also model the uncertainty of our model, we are interested in the predictive distribution [34], which can be

modeled as

$$p(y^* | x^*, \mathcal{D}) = \int \underbrace{p(y^* | x^*, \theta)}_{\text{Data}} \underbrace{p(\theta | \mathcal{D})}_{\text{Model}} d\theta \quad (1)$$

while θ are the parameters of our neural network, y^* is our prediction given an input x^* from a distinct dataset $\mathcal{D} = \{x_k, y_k\}_{k=1}^N$. The predictive distribution $p(y^* | x^*, \mathcal{D})$ is composed of two different types of uncertainties, the aleatoric uncertainty, i.e., data uncertainty, and the epistemic uncertainty, i.e., model uncertainty. The aleatoric uncertainty reflects sensor noise, synchronization errors, or erroneous ToA estimates [35] and is typically constant within space. Hence, it can be modeled as zero-mean Gaussian noise [36]. Thus, in contrast to the epistemic uncertainty, the aleatoric uncertainty is independent of the training data distribution. Fingerprinting models learn a one-by-one mapping of radio signals to positions, with a restricted predictive horizon to its spatial neighborhood. However, fingerprinting fails if radio signals from different areas are fed to the algorithm along with environmental changes altering the radio fingerprints. In both cases, the samples cause high localization errors [11] and are considered as OOD. They are reflected by high epistemic uncertainty and depend highly on the training data distribution.

There are various ways to model the uncertainty. Bayesian and ensemble-based methods are the most promising [34]. We compare different nonensemble-based approaches (single network approaches), SWAG [27] and subnetwork Laplace approximation [30], and ensemble-based approaches such as MC dropout [26], MultiSWAG [32], and deep ensembles [33].

1) *Single Network Approaches*: Single network Bayesian approaches approximate the posterior distribution by modeling the weights of the neural network as Gaussian distributions and are therefore Bayesian. SWAG [27] optimizes the weights θ of the neural network until it converges to a local minimum of the loss function. As training continues, the weights change slightly due to the optimizer's gradient steps and traverse the local minima in the weight space. The weights are used to estimate parameters of Gaussian distributions by maximum likelihood (ML) estimation

$$p(\theta | \mathcal{D}) \approx \mathcal{N}(\bar{\theta}, \Sigma) \quad (2)$$

with

$$\bar{\theta} \approx \frac{1}{N} \sum_{i=1}^N \theta_i \quad (3)$$

where θ_i are the weights after N additional gradient steps, $\bar{\theta}$ is the mean of the weights and

$$\Sigma \approx \frac{1}{N-1} \sum_{i=1}^N (\theta_i - \bar{\theta})(\theta_i - \bar{\theta})^\top \quad (4)$$

the covariance matrix. Estimating the full covariance matrix Σ is often intractable and can be approximated by the assumption of statistical independence between the weight distributions, i.e., $\tilde{\Sigma} = \text{diag}(\Sigma)$. The Laplace approximation [30] works similar. Here, also a converged neural network is employed with the estimated weights $\hat{\theta}$, while the Gaussian distributions are

estimated by the Taylor series expansion of the log posterior

$$\log p(\theta | \mathcal{D}) \approx \log p(\hat{\theta} | \mathcal{D}) - \frac{1}{2}(\theta - \hat{\theta})^\top \mathcal{H}(\theta - \hat{\theta}) \quad (5)$$

where the first-order term vanishes at the minimum of the loss function and $\mathcal{H} = \nabla^2 \log p(\theta | \mathcal{D})$ is the Hessian. The Laplace approximation can then be formulated as

$$p(\theta | \mathcal{D}) \sim \mathcal{N}(\hat{\theta}, \mathcal{H}^{-1}). \quad (6)$$

The burden of the Laplace approximation is to calculate the Hessian \mathcal{H} , which is expensive and intractable for large architectures. Consequently, several approximations are investigated, i.e., a diagonal covariance matrix or a Kronecker factorization [29]. The approximation can also be applied only to the last layer to enhance the efficiency [31]. Most recently, Daxberger et al. [30] showed that a Laplace approximation, with a full covariance matrix, on a subnetwork with the most contributing weights, i.e., those with the largest magnitudes, yields a very good tradeoff between tractability and expressiveness. So, we employ Laplace approximation with a full covariance matrix on 1.000 weights selected by its magnitude.

For inference, SWAG approximates the intractable integral (1) by Monte Carlo estimation, by means of N sampled deterministic networks. Thus, we return a position estimate

$$\hat{y}_k = \frac{1}{N} \sum_{i=1}^N f_{\theta_i}(x_k) \quad (7)$$

while we employ the variance of the Euclidean distance

$$\sigma_k^2 = \frac{1}{N-1} \sum_{i=1}^N (\|f_{\theta_i}(x_k)\|_2 - \|\hat{y}_k\|_2)^2 \quad (8)$$

as a scalar uncertainty estimate, where $\|\cdot\|_2$ is the L2-norm. The Laplace approximation directly estimates the covariance matrix of our 2-D position, where we use the determinant of the covariance matrix as scalar uncertainty estimate.

2) *Ensemble Approaches*: Single network approaches model the distribution of weights only locally around the MAP estimate. Hence, the loss landscape may not be sufficiently represented in the posterior. Instead, ensemble-based approaches have a better representation on the loss landscape, as every neural network will find a different, similar good, local minima in the loss landscape [32]. We employ different ensemble-based approaches, such as deep ensembles [33] and MultiSWAG [32]. We also consider MC dropout [26] as an ensemble-based approach, despite the fact that only a single network is trained. In MC dropout, dropout layers are integrated into the architecture and stay activated during training and evaluation. Thus, during every forward pass, a different subnetwork is created for the prediction. MC dropout estimates the uncertainty similar to SWAG, as given in (8). The other ensemble-based approaches, i.e., MultiSWAG and deep ensembles, train various models with random weight initialization. In MultiSWAG, several SWAG models are trained, while their samples are used to estimate a Gaussian posterior. The ensemble is then modeled as a Gaussian mixture model, while the variance is used as uncertainty estimate. Deep

ensembles learn the variance of the input samples directly during the training process by means of a negative likelihood loss with a head for the mean and the variance. Also here, the ensemble is modeled as Gaussian mixture model. However, in contrast to MultiSWAG, no sampling is required as the models are deterministic.

B. Out of Distribution Identification

We have to identify samples that are either outside of the fingerprinting area or corrupted due to environmental changes to enable a proper fingerprint life cycle management. In our previous work [14], we found that OOD samples are linearly separable from in domain samples. So, we have to define a single threshold for the uncertainty estimates to classify OOD samples. Statistical methods such as interquartile ranges approximate a threshold that may not fit perfectly and require fine-tuning.

Thus, to identify out of area samples, we propose a classification scheme that is based on logistic regression to have a more reliable threshold estimation. Data samples are recorded inside and outside the fingerprinting area and a logistic regression classifier [37] is trained based on the uncertainty of the samples. As we solve a binary classification problem, we label the samples as inside and outside of the area, and so, do not require a reference system.

C. Tracking Filter

In our evaluation, we show the fusion of UWB ToF measurements with a fingerprinting model. For the fusion of fingerprinting models and the ToF measurements we use a 2-D extended Kalman filter with a constant velocity model. As we directly estimate the position of the moving object with our fingerprinting models, the observation model is an identity matrix w.r.t. the x and y positions. Our observation model for the ToF measurements from the UWB system is based on Feng et al. [38]. For the fingerprinting, we assume that the position estimation errors of our DL models follow a homoscedastic unbiased Gaussian distribution within the training area. We assume this as we record a high density of data within the training area, and thus, the noise is only affected by the noisy ToA estimation, which is typically considered as unbiased and Gaussian distributed. This allows to estimate a constant variance of the position for the measurement noise covariance matrix using the training data distribution. Samples that are identified as OOD are not fed into the Kalman filter, as the fingerprinting model cannot extrapolate to unknown areas, as shown in Section V-A. For the UWB-ToF measurements, we assume a constant variance of the ToF errors as the error does not depend on the environment in LoS areas.

IV. EXPERIMENTAL DESIGN

A. UWB Data

For data acquisition we employed a small robot platform that carries a Decawave DW1000 UWB radio transceiver module. For the radio system, we configured a two-way ranging setup with a bandwidth of 499.2 MHz at a center frequency of 4 GHz with six stationary transceivers. We recorded data at 4 Hz and

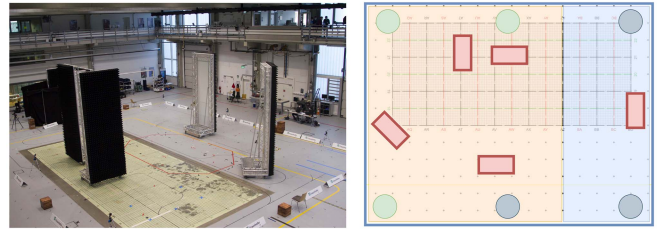


Fig. 2. Real world image (left) and top view (right) of the UWB environment. The anchors are placed at the edges of the environment (green and blue dots). The environment contains (red) walls that block the signals on the outside (black surface) and reflect the signals in the inside. The orange area indicates subspace for fingerprinting, whereas the blue area is for classical ToF-based positioning.

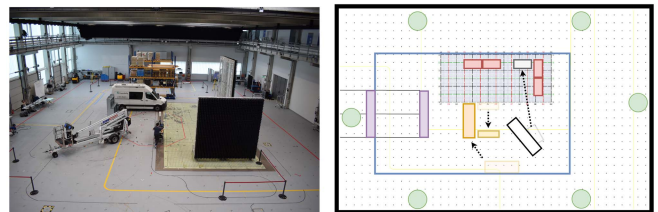


Fig. 3. Real world image (left) and top view (right) of the 5G environment. The anchors (green dots) are placed at the edges of the environment. The environment contains blocking walls (red), small shelves (blue), a fork lift (grey), a van (orange), and a work platform (white). The blue rectangle indicates the recording area.

reference positions with a millimeter-accurate camera-based motion tracking system (Qualysis). We constructed a complex NLoS dominated environment including walls, that reflect radio signals on the inner side (iron surface) and absorb them on the outside (black surface). The real world environment is shown in Fig. 2 on the left-hand side, while the schematic top view is shown on the right. The transceivers, indicated as dots, are placed at the edges of the recording area. The reflective walls, indicated in red, are placed to mimic a blockage of the LoS between the anchors and the robot platform. This causes ranging errors in the UWB radio system and leads to high localization errors with classical positioning approaches. For training and testing, we employ different recordings to avoid overfitting to the training database.

B. 5G Data

To obtain 5G data, we employed a proof-of-concept 5G downlink-TDoA positioning setup software-defined radio (SDR) system with six transmitters. The carrier frequency is at 3.75 GHz, with a bandwidth of 100 MHz, and a recording frequency of 100 Hz. We utilized the Nikon iGPS, a submillimeter optical reference system, to obtain position ground truth information. The environment of the 5G dataset, shown in Fig. 3, reflects a typical industrial environment. The real environment is shown on the left, while a schematic top view is shown on the right. It includes a forklift (grey), a work platform (white), a van (orange), absorber walls (red), a small shelf (yellow), and large metal shelves (purple). Again, the receivers, indicated as green

dots, are placed at the edges of the environment. We recorded two states of the environment. In the first state (Industrial 1), the obstacles are placed at the semitransparent positions, while we moved the obstacles to the nontransparent positions (Industrial 2). For training and testing, we employ different recordings to avoid overfitting the training database.

C. Data Preprocessing

For data preprocessing, we follow the idea of [11] to exploit the raw CIR in the time domain and the corresponding ToF / TDoA using a CNN. We generate a 3-D tensor with dimension $[2, N_A, L_w]$, with N_A Anchors, L_w time steps in the temporal resolution of the CIR and their real and imaginary parts (2). We pad the CIRs by the corresponding ToF/TDoA within the tensor to model the unique relative shift of the first path of arrival for every position in the area.

D. Model Architectures

Inspired by Stahlke et al. [11], we used a simple six-layer CNN that consists of four convolutional layers and two fully connected layers as our fingerprinting model. For the convolutional layers we employ batch normalization for training stabilization and the rectified linear units (ReLU) for all layers except for the last layer, where we employ the identity function. We did not use local pooling. Instead, we used a global average pooling over all channels before the dense layers to keep the temporal resolution of CIRs. The sizes of the kernels are increased in the consecutive convolutional layers to ensure a growing receptive field with the depth of the model. For the UWB model, the first convolutional layer has a kernel size of $[1 \times 3]$, the second $[2 \times 5]$, the third $[4 \times 15]$, and the last $[6 \times 30]$. The 5G model utilizes a kernel size of $[3 \times 3]$ for the first, $[5 \times 5]$ for the second, $[7 \times 7]$ for the third and $[10 \times 10]$ for the last layer. Instead, for UWB, we selected larger kernels widths as the temporal resolution of the UWB CIRs is higher compared to the 5G CIRs. The MC dropout models have an additionally dropout layer, with a dropout rate of 30%, before the last layer. We removed the batch normalization layers for the SWAG models as they contradict the SWAG logic as each sample of the weights requires the batch normalization to be adjusted by the training data [27]. This would lead to a high overhead, hindering real-time inference. For the models within the deep ensemble, we added two fully connected layers with a single output as a second head for the variance. Table I summarizes the architectures. Note that we did not optimize the architecture for performance. Instead, we focused on simplicity and inference speed to provide an uncertainty-based OOD identification approach that does not suffer from complex architectures.

E. Performance Metrics

We use the uncertainty estimations of the neural network to identify the spatial limitations or corrupted areas of the fingerprinting model. To evaluate the performance, we formulate the problem as a classification problem considering samples within the training area or below a certain localization error as positive

TABLE I
PARAMETERS OF THE FINGERPRINTING ARCHITECTURES FOR THE UWB AND 5G RADIO SETUPS

Layer type	Channels		Kernel size		Activation
	UWB	5G	UWB	5G	
Conv.	8	32	1×3	3×3	ReLU
Conv.	8	32	2×5	5×5	ReLU
Conv.	8	32	4×7	7×7	ReLU
Conv.	16	64	6×10	10×10	ReLU
Avg. Pool.	6×200	6×49	—	—	—
Fully Con.	300	250	—	—	ReLU
Fully Con.	2	2	—	—	—

samples and samples above as negative samples. Since we need to define a continuous uncertainty value threshold to classify OOD samples, we use the receiver operating characteristic (ROC) curve to evaluate the performance of classification for all thresholds. The ROC curve compares the false positive rate (FPR), i.e., false positives divided by the number of negatives, to the true positive rate (TPR), i.e., true positives divided by the number of positives, for all thresholds in a binary classification problem. To quantify the performance of the classifier, we exploit the area under the ROC curve (AUROC), which quantifies the overall performance and threshold stability. The AUROC ranges from $[1, 0]$, while 1 is a perfect classifier. To quantify the performance at a certain threshold, we use the balanced accuracy, as we often have imbalanced datasets w.r.t. in-domain and out-of-domain samples. The balanced accuracy is the mean of TPR and true negative rate (TNR), i.e., $TNR = 1 - FPR$. As localization error, we employ the Euclidean distance from the estimated position to the reference and evaluate the circular error at the 90% percentile (CE90) and the mean absolute error (MAE).

V. EVALUATION

First, we evaluate the uncertainty estimation for OOD (out of area identification and the identification of environmental changes) of fingerprinting models. Second, we investigate the effect of combining ToF positioning and CIR fingerprinting on the positioning accuracy.

A. Out of Distribution Identification

OOD samples cause large errors in fingerprinting models as they cannot extrapolate into unknown areas or fail if the fingerprint is altered due to environmental changes [11]. Thus, we investigate the capability of the out of area identification and the detection of outdated fingerprints. For the investigation of the out of area identification, we created several subspaces in the environments of the 5G and UWB datasets. First, we investigate an extrapolation scenario, where we vertically split the environment and so, create two different nonoverlapping areas. As fingerprinting models cannot extrapolate outside of the training area, an uncertainty-based identification may be rather simple. In a second, more challenging investigation, we remove an areal within the training area. The fingerprinting model may be able to interpolate, as the “left out” area is surrounded by the training

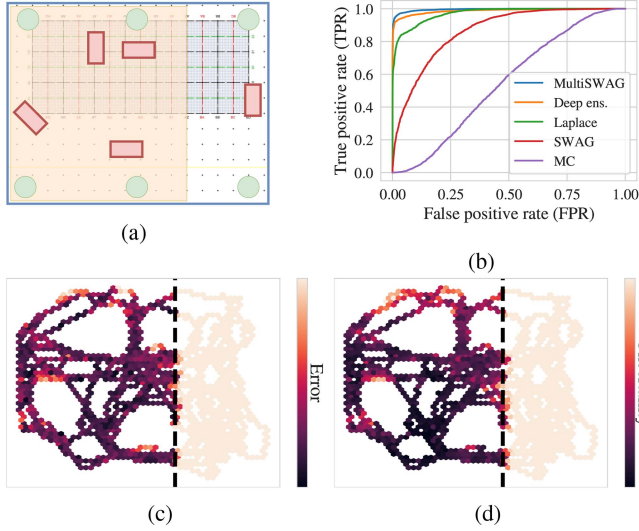


Fig. 4. Uncertainty estimation for out of area identification with a vertical split for the UWB environment. (a) Training subspace (orange rectangle) in the recording area is shown as blue rectangle. (b) ROC-curves of the out of area identification for the uncertainty models. (c) Heat map of the positional error. (d) Heat map of the predictive uncertainty of MultiSWAG.

data. So, also, the uncertainty-based OOD identification is more challenging as the data samples are more similar to the training data compared to the extrapolation case. As a final evaluation, we examine whether we identify environmental changes in the 5G environment.

1) *Vertical Split*: First, we investigate the abilities of the uncertainty approaches to identify spatially separated areas. We created a fingerprinting area, which has no overlap with the unseen area, hence, the fingerprinting model needs to extrapolate. Both environments, the UWB environment and the INDUSTRIAL 1 environment of the 5G dataset shown in Figs. 4(a) and 5(a) are split vertically. For both figures, the orange rectangles indicate the training areas, whereas the blue rectangle shows the total area for prediction. The errors of the fingerprinting models are shown in Fig. 4(c) for the UWB and in Fig. 5(c) for the 5G environment. The black dashed line shows the transition from the training area (left) to the unseen area (right). The heatmaps show the limited extrapolation capabilities of the neural network, as none of the networks can extrapolate into the right area, causing large errors. The uncertainty, estimated by the best uncertainty approaches, i.e., MultiSWAG for the UWB and deep ensembles for the 5G environment, are shown in Fig. 4(d) for the UWB and in Fig. 5(d) for the 5G model. Both uncertainty approaches can identify the spatial limitations of the model very well. Here, the MultiSWAG model from the UWB dataset achieves an AUROC of 0.99. This indicates that the model has a low sensitivity to the uncertainty threshold, while achieving a high balanced accuracy of 0.97 for the optimal threshold. We see similar results for the 5G dataset, as the AUROC is at 0.97 for the deep ensembles, with a balanced accuracy of 0.91 for the best threshold. The balanced accuracy is lower compared to the UWB database, as samples close to the transition boundary are falsely classified shown in Fig. 5(d). Samples close to the transition line on the right still have a low

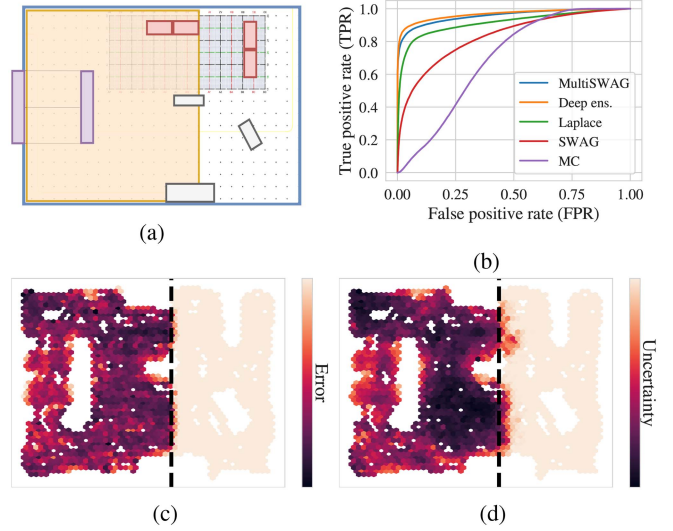


Fig. 5. Uncertainty estimation for out of area identification with a vertical split for the 5G environment. (a) Fingerprinting training subspace (orange rectangle) in the recording area (blue rectangle). (b) ROC-curves of the out of area identification for the uncertainty models. (c) Heat map of the positional error. (d) Heat map of the predictive uncertainty of Deep ens.

TABLE II
AUROC RESULTS FOR BOTH RADIO SYSTEMS (UWB, 5G) ON THE OUT OF AREA IDENTIFICATION FOR THE VERTICAL SPLIT (SPLIT) AND THE PATCH LEFT OUT (PATCH), AND THE ENVIRONMENTAL CHANGE OF THE 5G DATASET (CHANGED)

Sys.	Scen.	M.SWAG	D. ens.	Laplace	SWAG	MC
UWB	Split	0.99	0.99	0.97	0.86	0.55
	Patch	0.94	0.88	0.87	0.59	0.39
5G	Split	0.96	0.97	0.92	0.83	0.69
	Patch	0.91	0.92	0.73	0.65	0.45
	Changed	0.86	0.84	0.58	0.56	0.55

The bold values indicate the best accuracies for system and scenario.

uncertainty and are considered as inside. This effect is smaller at the UWB dataset shown in Fig. 4(d). We think this is due to the higher aleatoric uncertainty that is caused by the lower bandwidth of the 5G radio system. The standard deviation of the training error, only including in domain samples, is 0.29 m for the 5G system, while the UWB system has only a standard deviation of 0.16 m, and thus, also a lower aleatoric uncertainty. As the neural network estimates a combined uncertainty of aleatoric and epistemic uncertainty, the samples need a higher distance to the training area as the epistemic uncertainty has to dominate the overall uncertainty to enable a reliable identification of OOD samples.

We evaluated different uncertainty approaches to identify the spatial limitation of the model. The ROC curves are shown in Fig. 4(b) for the UWB and in Fig. 5(b) for the 5G dataset and the AUROC values are presented in Table II. In both datasets, the ensemble-based methods, deep ensembles and MultiSWAG, achieve the highest AUROC values of 0.99 for the UWB and 0.96 and 0.97 for the 5G dataset. Laplace approximations is lower than the ensemble-based methods, but still achieve an AUROC value of 0.97 for the UWB and 0.92 for the 5G dataset. SWAG

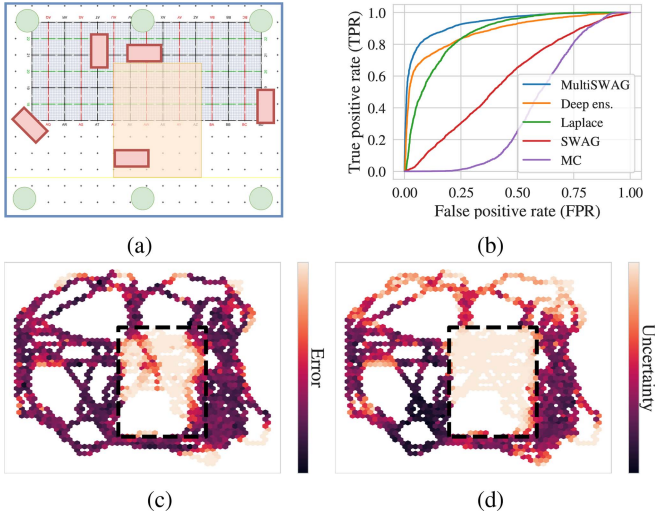


Fig. 6. Uncertainty estimation for out of area identification with a patch left out for the UWB environment. (a) Training subspace (orange rectangle) in the recording area shown as a blue rectangle. (b) ROC-curves of the out of area identification for the uncertainty models. (c) Heat map of the positional error. (d) Heat map of the predictive uncertainty of MultiSWAG.

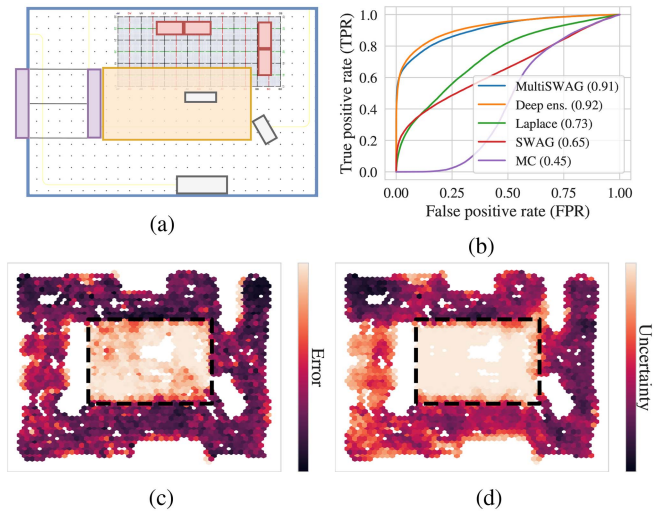


Fig. 7. Uncertainty estimation for out of area identification with a patch left out for the 5G environment. (a) Training subspace (orange rectangle) in the recording area shown as a blue rectangle. (b) ROC-curves of the out of area identification for the uncertainty models. (c) Heat map of the positional error. (d) Heat map of the predictive uncertainty of Deep ens.

achieves a lower AUROC of only 0.86 for the UWB and 0.83 for the 5G dataset. MC fails with only an AUROC of 0.55 for the UWB and 0.69 for the 5G dataset.

2) *Patch Subspace*: In a second evaluation, we remove a rectangular area from the training datasets. In contrast to the vertical split, the unseen area is surrounded by the training area and thus, we assume that the network interpolates positions to a certain extent, rendering the uncertainty-based area identification potentially more challenging. The left out areas are shown in Fig. 6(a) for the UWB and in Fig. 7(a) for the 5G dataset,

whereas the blue areas indicate the total area and the orange rectangles the left-out patches. The errors of localization are shown in Fig. 6(c) for the UWB and in Fig. 7(c) for the 5G dataset. In both scenarios, the neural networks interpolates in some sections, while the fingerprinting models still fail in the most sections of left-out areas. The results of the uncertainty estimation are shown in Fig. 6(d) for the UWB and in Fig. 7(d) for the 5G dataset. The uncertainty approaches identify the spatial limitations of the training area well, with an AUROC of 0.94 for the UWB with the MultiSWAG approach and 0.92 for the 5G dataset with the deep ensembles model. Here, the AUROC is similar to the left subspace evaluation, lower for the 5G dataset, which is due to the same reason as in the vertical split evaluation. As the 5G system has a higher aleatoric uncertainty, a distinction between aleatoric and epistemic uncertainty is more difficult, which results in a lower reliability close to the borders of the training area.

In general, the AUROC is lower for all uncertainty approaches compared to the vertical split evaluation of the UWB dataset shown in Fig. 6(b) and the 5G dataset shown in Fig. 7(b). The AUROC values are presented in Table II. Again, the ensemble-based approaches, i.e., MultiSWAG and deep ensembles, achieve the best results with an AUROC of 0.94 and 0.88 for the UWB and 0.86 and 0.84 for the 5G dataset. The Laplace approximation achieves a higher AUROC of 0.87, close to the deep ensembles, for the UWB dataset but fails in the 5G dataset with an AUROC value of only 0.73. Also, SWAG and MC only achieve low AUROC values of below 0.6 in both datasets and thus, cannot identify the spatial limitations of the fingerprinting models reliably.

3) *Environmental Change*: Stahlke et al. [11] showed that environmental changes affect the accuracy of the fingerprinting model, mostly locally at the area of the environmental change. In this evaluation, we investigate whether uncertainty approaches identify outdated fingerprints to provide a reliability measure or trigger a retraining process to update the fingerprinting model. We consider samples with an error that is larger than three standard deviations of the training dataset as corrupted, 0.9 m for the 5G dataset, and thus, OOD. The samples with accuracies below 0.9 m are considered to be valid. Fig. 8(a) shows the training environment with the objects at the semitransparent positions and the test environment at the nontransparent locations. Fig. 8(c) shows the errors of the fingerprinting model. We see that the errors are especially high close to the environmental changes, indicated as black dashed rectangles. This makes it easy to identify these areas and update and correct the fingerprint model. The uncertainty estimated by the best model, MultiSWAG, is shown in Fig. 8(d). While the uncertainty identifies the corrupted areas, the performance is worse compared to the other evaluation with an AUROC of only 0.86. While the corrupted area left to the new position of the van (most left rectangle), reflects the error quite well, the uncertainty of other areas is underestimated, especially at the transition between the valid and corrupted areas. We assume that this is the same effect as in the other evaluations, in which the transition areas are often underestimated due to the uncertainty. As the affected areas are relatively small, the transition areas dominate the corrupted area, which leads to

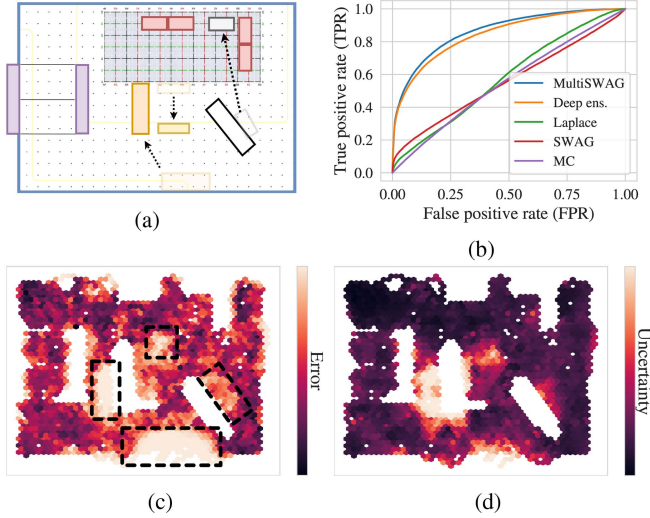


Fig. 8. Uncertainty estimation for the identification of environmental changes for the 5G environment. (a) Training scenario with the semi-transparent and test scenario with the nontransparent elements. (b) ROC-curves of the identification for the uncertainty models. (c) Heat map of the positional error. (d) Heat map of the predictive uncertainty of MultiSWAG.

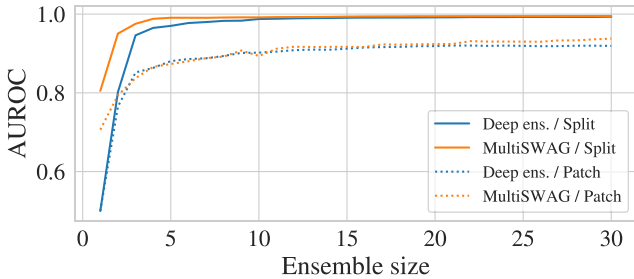


Fig. 9. AUROC for increasing ensemble sizes for the UWB dataset. Blue indicates the deep ensemble model, while orange indicates MultiSWAG. The split scenario is indicated as solid line, while the patch scenario is indicated as dotted line.

worse results. Fig. 8(b) shows that only the ensemble-based methods, i.e., deep ensembles and MultiSWAG, achieve a high AUROC of 0.84 and 0.86. The other approaches perform poorly with AUROC values only up to 0.58, rendering a identification of corrupted areas impossible.

B. Ensemble Size

Our evaluations have shown that ensemble-based methods, i.e., deep ensembles and MultiSWAG, achieve the highest accuracies in OOD identification. However, compared to single network approaches, such as SWAG and Laplace approximations, ensembles, beside MC dropout, have a high overhead during training and inference, as multiple networks have to be trained and evaluated during runtime. Thus, we analyze the required size of the ensemble to reliably identify OOD samples. Fig. 9 shows the AUROC values for various ensemble sizes for the UWB radio evaluations. Blue indicates deep ensembles, while orange indicates MultiSWAG. The scenarios, split and patch,

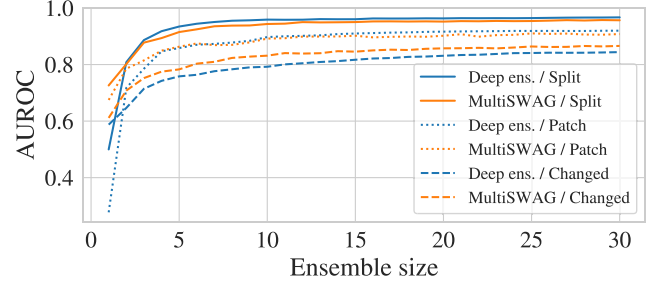


Fig. 10. AUROC for increasing ensemble sizes for the 5G datasets. Blue indicates the deep ensemble model, while orange indicates MultiSWAG. The split scenario is indicated as solid line, the patch scenario is indicated as dotted line and the changed scenario as dashed line.

are indicated as solid and dotted line. In the split scenario, MultiSWAG already achieves the highest results at an ensemble size of four models, while the deep ensembles require around ten networks to achieve similar results. In the patch scenario, both models are almost similar, and still improve the performance for ensemble sizes larger than ten. However, the improvement is only small and generates a high overhead during training and evaluation. We see similar results for the 5G radio system, shown in Fig. 10. Deep ensembles and MultiSWAG show similar results, for various ensemble sizes, while the easier scenarios, i.e., split and patch, require around ten networks to achieve the highest results. In the environmental change (Changed) scenario, the ensembles still benefit from a larger number of networks. The results show that a number of around ten networks have the best trade of between training and evaluation complexity and OOD identification performance.

C. Threshold Stability

A crucial step to identify OOD samples is to define a threshold, at which uncertainty the samples are within or out of the distribution. While this can easily be done with a few labeled samples or interquartile ranges, as described in Section III-B, estimating a new threshold for every new model or environmental change is time consuming or not feasible. We thus, evaluate the robustness of thresholds within an environment for the several states of the environment, i.e., split and patch subspace and the environmental change. For a state, the threshold with the highest balanced accuracy is selected and applied to the other states of the environment, to benchmark the sensitivity of the threshold across the environmental states. When we apply the ideal threshold within the same environment, i.e., split and patch, for the UWB dataset, we achieve a balanced accuracy of 0.97 and 0.87, while we only have a degradation of 0.02 if we apply the threshold from the patch environment in the split environment. When we apply the threshold from the split environment to the patch environment, we have a degradation of 0.04 and achieve a balanced accuracy of 0.83. The results show that the performance is relatively stable for different thresholds. For the 5G datasets the threshold stability is even better (see Table III). The diagonal elements show the balanced accuracy with the ideal threshold for the environment and the off-diagonal elements show the results with the threshold of the other environment state. The rows

TABLE III
BALANCED ACCURACIES FOR THE ENVIRONMENTAL STATES OF THE 5G SYSTEM

	Split	Patch	Changed
Split	0.91	0.88	0.90
Patch	0.82	0.82	0.82
Changed	0.77	0.77	0.78

The optimal threshold within the environmental condition is used in the diagonal elements. In the off-diagonal elements, the optimal thresholds from the other environmental states are applied to the row's environmental state.

show the results within the environment, while the columns indicate the environment from which the threshold is derived. The highest degradation in the accuracy, by 0.03, is the threshold applied from the patch environment to the split environment. For other combinations, the thresholds are stable across the environmental states. Interestingly, the thresholds estimated by the spatially limited models, i.e., split and patch, reliably estimate corrupted areas in the changed environment state. This indicates that a threshold estimated once, for a certain environment, is stable also across environmental changes and spatial model compositions.

D. Positioning

In this section, we investigate the fusion of classical ToF positioning and radio fingerprinting in a Kalman filter in the UWB environment. We only trained the fingerprinting models in a subspace to show the combination of classical ToF positioning with data-driven fingerprinting models using the proposed out-of-area identification. For the ToF positioning baseline we only used the ToF measurements in the Kalman filter. Environments often have areas with different propagation conditions, such as the environment shown in Fig. 2 on the right-hand side. On the right-hand side of the environment, some transceivers, indicated as gray dots, exhibit reliable LoS conditions, while on the left hand side there is no clear subspace with reliable LoS conditions to a subset of transceivers. We therefore train a fingerprinting model on the left hand side, indicated as orange area, and use classical ToF positioning on the right hand side, indicated as blue area, only considering the gray transceivers. We use the classical ToF positioning as fall-back solution, which takes over if the fingerprinting model has a high uncertainty. The logistic regression classifier for the threshold estimation is trained on a different dataset within the same environment. Fig. 11 shows the cumulative distribution functions (CDFs) of the localization error (Euclidean distance) using only ToF positioning (UWB ToF), only fingerprinting trained on the left hand side (UWB Fp.) and the combination of classical ToF positioning and fingerprinting (Combined). The results show that the combined approach achieves the highest accuracy with an MAE of 0.21 m and a CE90 of 0.32 m. The classical approaches achieve reliable positioning on the right hand side, while the fingerprinting model takes over on the left hand side to enable a high accuracy. By only using the classical ToF positioning the localization accuracy is lower with an MAE of 0.56 m and a CE90 of 1.39 m.

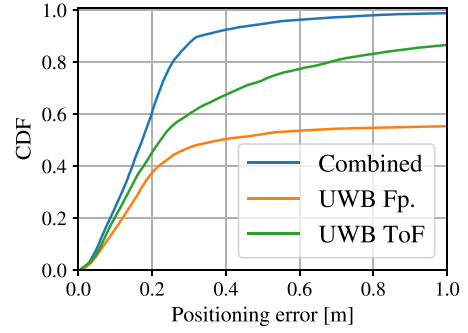


Fig. 11. Cumulative distribution function of the positioning error for the split environment.

NLoS links cause errors in the ToF measurements, which leads to errors in the localization accuracy on the left hand side of the area. Localization only using the fingerprinting model fails with an MAE of 2.72 m and a CE90 at 8.46 m. The model is only trained on the left hand side, which leads to significant errors as the neural network cannot extrapolate into unknown areas as shown in Section V-A1. To identify the limitations of the neural network is therefore crucial to enable a reliable positioning.

VI. DISCUSSION

In this section, we discuss the key findings of our evaluations along with recommendations for robust model monitoring.

A. OOD Identification is Challenging at the Boundaries

Our evaluations show that uncertainty-based OOD identification works, for various radio systems with different bandwidths. We found that samples, that are spatially far apart, are reliably identified as OOD, while samples close to the border line from in and out-of-domain areas are more challenging to identify. This may be due to the fact that samples closer to the in-domain are more similar as samples far away due to the manifold of the CSI [39], [40]. Thus, samples in the proximity of the border line have a lower epistemic uncertainty compared to samples that are far away.

B. Bandwidth Impacts the OOD Identification Performance

We show in our evaluations in Section V-A, that a radio system with a lower bandwidth has a higher OOD error at the transition line compared to a system with higher bandwidth. This is due to the higher aleatoric caused by the lower bandwidth. In hybrid localization scenarios, when we combine classical TDoA/ToF-based localization with fingerprinting, we thus, recommend to add a padding to the training area overlapping with the areas of classical positioning. So, an unreliable identification within the transition area does not affect the accuracy, as the fingerprinting and the classical methods are valid in this transition area.

C. Ensemble-Based Methods are Superior in OOD Detection

We have shown that ensemble-based methods, i.e., deep ensembles and MultiSWAG, perform superior compared to single network approaches. However, their effort for training and evaluation is significantly higher compared to single network approaches, as multiple networks have to be trained and evaluated to estimate the uncertainty. While the performance of MultiSWAG and deep ensembles is almost on par, MultiSWAG requires an additional sampling process for every single networking, rendering the inference even more expensive. Also, SWAG and MC dropout require multiple forward passes to estimate the uncertainty, rendering them equally expensive as the deterministic deep ensemble, which requires only a single forward pass for every network. Also, the Laplace approximation has an overhead during inference, as the Jacobi matrix of the subnetwork has to be calculated to estimate the uncertainty. However, in contrast to the other single-network approaches, the effort is comparably low. We thus, recommend to employ deep ensembles, which require a higher training effort, but similar effort on the inference compared to single network approaches. However, if inference time is crucial, the Laplace approximation can be deployed for simpler use cases.

D. Thresholds for OOD Identification are Stable Within an Environment

A binary classification of OOD samples requires a threshold for the uncertainty. While the determination of a threshold for a spatial limited model is straightforward as described in Section III-B, defining a threshold, that identifies corrupted areas, is more difficult. We typically do not have several environmental states at training time to derive a threshold by the proposed approach. However, our analysis of the threshold stability described in Section V-C shows that, within an environment, a threshold is stable across various environmental states, i.e., fingerprinting models trained in different areas of the environment and at environmental changes. Thus, we recommend deriving a uncertainty threshold via out-of-area samples and monitoring the model for outdated fingerprints.

E. Generalization Across Radio Technologies

Our experiments show that our approach works across UWB and 5G ToA/ToF-based localization systems. It may also be applicable to WIFI-based [41] systems and other multiantenna architectures such as MIMO [42], as long as they provide a clear match of channel measurements and position. However, for MIMO systems, the bandwidth may not be the limiting factor of OOD identification due to their high spatial diversity.

VII. CONCLUSION

In this article, we show that our uncertainty-based approach successfully identifies spatial constraints of fingerprint models and corrupted fingerprints. In contrast to the state-of-the-art, we identify OOD samples of the fingerprint model intrinsically, so additional positioning systems are not required for identification.

In our ablation study, we find that ensemble-based uncertainty methods are superior in identifying OOD samples compared to single-network approaches on UWB and 5G radio systems data. Our method enables the combination of classic radio localization and fingerprinting while simultaneously providing a reliability measurement of the fingerprint model. Thus, our method drastically reduces the cost of data collection and labeling process in LoS areas and enables more robust positioning in complex NLoS environments. On top of that, our system guarantees an efficient maintenance process.

ACKNOWLEDGMENT

The authors would like to thank M. Kasparek, A. Eidloth, J. Niklas Bauer, and M. Soliman for implementing the 5G uplink-TDoA positioning setup and processing pipeline.

Author contributions: M. Stahlke wrote and edited the paper, reviewed related work, wrote the code, performed experiments, and conceptualized the methodology. M. Stahlke, T. Feigl, and S. Kram interpreted and discussed the results. M. Stahlke designed the experiments and ablation studies. T. Feigl, S. Kram, B. M. Eskofier, and C. Mutschler reviewed and edited the paper. B. M. Eskofier and C. Mutschler supervised the work.

REFERENCES

- [1] R. W. Wolcott and R. M. Eustice, "Fast LiDAR localization using multiresolution Gaussian mixture maps," in *Proc. IEEE Int. Conf. Robot. Autom.*, 2015, pp. 2814–2821.
- [2] X. Luo, H. Wang, S. Yan, J. Liu, Y. Zhong, and R. Lan, "Ultrasonic localization method based on receiver array optimization schemes," *Int. J. Distrib. Sensor Netw.*, vol. 14, no. 11, 2018, Art. no. 1550147718812017.
- [3] A. Kendall, M. Grimes, and R. Cipolla, "PoseNet: A convolutional network for real-time 6-DOF camera relocalization," in *Proc. Int. Conf. Comput. Vis.*, 2015, pp. 2938–2946.
- [4] B. Silva and G. P. Hancke, "IR-UWB-based non-line-of-sight identification in harsh environments: Principles and challenges," *IEEE Trans. Ind. Inform.*, vol. 12, no. 3, pp. 1188–1195, Jun. 2016.
- [5] M. Stahlke, S. Kram, C. Mutschler, and T. Mahr, "NLOS detection using UWB channel impulse responses and convolutional neural networks," in *Proc. IEEE Int. Conf. Localization GNSS*, 2020, pp. 1–6.
- [6] M. Stahlke, S. Kram, F. Ott, T. Feigl, and C. Mutschler, "Estimating TOA reliability with variational autoencoders," *IEEE Sensors J.*, vol. 22, no. 6, pp. 5133–5140, Mar. 2022.
- [7] V. Barral, C. J. Escudero, J. A. García-Naya, and R. Maneiro-Catoira, "NLOS identification and mitigation using low-cost UWB devices," *Sensors*, vol. 19, no. 16, 2019, Art. no. 3464.
- [8] S. Marano, W. M. Gifford, H. Wymeersch, and M. Z. Win, "NLOS identification and mitigation for localization based on UWB experimental data," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 7, pp. 1026–1035, Sep. 2010.
- [9] Y. Zhu et al., "SNWPM: A siamese network based wireless positioning model resilient to partial base stations unavailable," *IEEE China Commun.*, vol. 20, no. 9, pp. 20–33, Sep. 2023.
- [10] M. Widmaier, M. Arnold, S. Dörner, S. Cammerer, and S. T. Brink, "Towards practical indoor positioning based on massive MIMO systems," in *Proc. IEEE 90th Veh. Technol. Conf.*, 2019, pp. 1–9.
- [11] M. Stahlke, T. Feigl, M. H. C. García, R. A. Stirling-Gallacher, J. Seitz, and C. Mutschler, "Transfer learning to adapt 5G AI-based fingerprint localization across environments," in *Proc. IEEE 95th Veh. Technol. Conf.*, 2022, pp. 1–5.
- [12] F. Furfari et al., "Discovering location based services: A unified approach for heterogeneous indoor localization systems," *Internet Things*, vol. 13, 2021, Art. no. 100334.
- [13] F. Furfari, A. Crivello, P. Barsocchi, F. Palumbo, and F. Potorti, "What is next for indoor localisation? Taxonomy, protocols, and patterns for advanced location based services," in *Proc. IEEE Int. Conf. Indoor Positioning Indoor Navigation*, 2019, pp. 1–8.

- [14] M. Stahlke, T. Feigl, S. Kram, B. M. Eskofier, and C. Mutschler, "Uncertainty-based fingerprinting model selection for radio localization," in *Proc. IEEE 13th Int. Conf. Indoor Positioning Indoor Navigation*, 2023, pp. 1–6.
- [15] S. Aditya, A. F. Molisch, and H. M. Behairy, "A survey on the impact of multipath on wideband time-of-arrival based localization," *Proc. IEEE*, vol. 106, no. 7, pp. 1183–1203, Jul. 2018.
- [16] X. Zhu, W. Qu, T. Qiu, L. Zhao, M. Atiqzaman, and D. O. Wu, "Indoor intelligent fingerprint-based localization: Principles, approaches and challenges," *IEEE Commun. Surveys Tuts.*, vol. 22, no. 4, pp. 2634–2657, Fourth Quarter 2020.
- [17] M. Stahlke, S. Kram, T. Mumme, and J. Seitz, "Discrete positioning using UWB channel impulse responses and machine learning," in *Proc. IEEE Int. Conf. Localization GNSS*, 2019, pp. 1–6.
- [18] M. N. de Sousa and R. S. Thomä, "Enhancement of localization systems in NLOS urban scenario with multipath ray tracing fingerprints and machine learning," *Sensors*, vol. 18, no. 11, 2018, Art. no. 4073.
- [19] A. Niitsoo, T. Edelhäußer, E. Eberlein, N. Hadaschik, and C. Mutschler, "A deep learning approach to position estimation from channel impulse responses," *Sensors*, vol. 19, no. 5, 2019, Art. no. 1064.
- [20] A. Foliadis, M. H. C. Garcia, R. A. Stirling-Gallacher, and R. S. Thomä, "CSI-based localization with CNNs exploiting phase information," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2021, pp. 1–6.
- [21] A. C. Eyng, O. K. Rayel, E. Oroski, and J. L. Rebelatto, "Kalman filtering-aided hybrid indoor positioning system with fingerprinting and multilateration," in *Proc. IEEE 91st Veh. Technol. Conf.*, 2020, pp. 1–5.
- [22] J. He and H. C. So, "A hybrid TDOA-fingerprinting-based localization system for LTE network," *IEEE Sensors J.*, vol. 20, no. 22, pp. 13653–13665, Nov. 2020.
- [23] M. A. Bitew, R.-S. Hsiao, H.-P. Lin, and D.-B. Lin, "Hybrid indoor human localization system for addressing the issue of RSS variation in fingerprinting," *Int. J. Distrib. Sensor Netw.*, vol. 11, no. 3, 2015, Art. no. 831423.
- [24] A. Foliadis, M. H. C. Garcia, R. A. Stirling-Gallacher, and R. S. Thomä, "Reliable deep learning based localization with CSI fingerprints and multiple base stations," in *Proc. IEEE Int. Conf. Commun.*, 2022, pp. 3214–3219.
- [25] R. L. Russell and C. Reale, "Multivariate uncertainty in deep learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 12, pp. 7937–7943, Dec. 2022.
- [26] Y. Gal and Z. Ghahramani, "Dropout as a Bayesian approximation: Representing model uncertainty in deep learning," in *Proc. Int. Conf. Mach. Learn.*, 2016, pp. 1050–1059.
- [27] W. J. Maddox, P. Izmailov, T. Garipov, D. P. Vetrov, and A. G. Wilson, "A simple baseline for Bayesian uncertainty in deep learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 13153–13164.
- [28] E. Daxberger, A. Kristiadi, A. Immer, R. Eschenhagen, M. Bauer, and P. Hennig, "Laplace redux - effortless Bayesian deep learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2021, pp. 20089–20103.
- [29] H. Ritter, A. Botev, and D. Barber, "A scalable laplace approximation for neural networks," in *Proc. 6th Int. Conf. Learn. Representations*, 2018, pp. 5463–5477.
- [30] E. Daxberger, E. Nalisnick, J. U. Allingham, J. Antorán, and J. M. Hernández-Lobato, "Bayesian deep learning via subnetwork inference," in *Proc. Int. Conf. Mach. Learn.*, 2021, pp. 2510–2521.
- [31] A. Kristiadi, M. Hein, and P. Hennig, "Being Bayesian, even just a bit, fixes overconfidence in ReLU networks," in *Proc. Int. Conf. Mach. Learn.*, 2020, pp. 5436–5446.
- [32] A. G. Wilson and P. Izmailov, "Bayesian deep learning and a probabilistic perspective of generalization," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 4697–4708.
- [33] B. Lakshminarayanan, A. Pritzel, and C. Blundell, "Simple and scalable predictive uncertainty estimation using deep ensembles," in *Proc. Adv. Neural Inf. Process. Syst.*, 2017, pp. 6405–6416.
- [34] J. Gawlikowski et al., "A survey of uncertainty in deep neural networks," *Artif. Intell. Rev.*, vol. 56, no. Suppl_1, pp. 1513–1589, 2023.
- [35] T. Feigl, E. Eberlein, S. Kram, and C. Mutschler, "Robust ToA-estimation using convolutional neural networks on randomized channel models," in *Proc. IEEE Int. Conf. Indoor Positioning Indoor Navigation*, 2021, pp. 1–8.
- [36] D. Musicki, R. Kaune, and W. Koch, "Mobile emitter geolocation and tracking using TDOA and FDOA measurements," *IEEE Trans. Signal Process.*, vol. 58, no. 3, pp. 1863–1874, Mar. 2009.
- [37] S. Menard, *Applied Logistic Regression Analysis*. Newbury Park, CA, USA: Sage, 2002.
- [38] D. Feng, C. Wang, C. He, Y. Zhuang, and X.-G. Xia, "Kalman-filter-based integration of IMU and UWB for high-accuracy indoor positioning and navigation," *IEEE Internet Things J.*, vol. 7, no. 4, pp. 3133–3146, Apr. 2020.
- [39] A. Schwartz and R. Talmon, "Intrinsic isometric manifold learning with application to localization," *SIAM J. Imag. Sci.*, vol. 12, no. 3, pp. 1347–1391, 2019.
- [40] M. Stahlke, G. Yammine, T. Feigl, B. M. Eskofier, and C. Mutschler, "Indoor localization with robust global channel charting: A time-distance-based approach," *IEEE Trans. Mach. Learn. Commun. Netw.*, vol. 1, pp. 3–17, 2023.
- [41] X. Wang, L. Gao, S. Mao, and S. Pandey, "DeepFi: Deep learning for indoor fingerprinting using channel state information," in *Proc. IEEE Wireless Commun. Netw. Conf.*, 2015, pp. 1666–1671.
- [42] C. Wu et al., "Learning to localize: A 3D CNN approach to user positioning in massive MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 20, no. 7, pp. 4556–4570, Jul. 2021.