Delay Estimation in Dense Multipath Environments using Time Series Segmentation

Sebastian Kram[†*], Christopher Kraus[*], Maximilian Stahlke[*], Tobias Feigl[*], Jörn Thielecke[†], and Christopher Mutschler[*]

{ sebastian.kram, krauscr, maximilian.stahlke, tobias.feigl, christopher.mutschler }@iis.fraunhofer.de
{ sebastian.k.kram, joern.thielecke }@fau.de

[*]Fraunhofer IIS,
Fraunhofer Institute for Integrated Circuits IIS,
Division Positioning and Networks

[†]Friedrich-Alexander University Erlangen-Nürnberg (FAU)
Institute of Information Technology
Erlangen, Germany

***Abstract*—Channel measurements at sufficiently high bandwidth in multipath-rich environments include a variety of delay information, which, if accurately extracted, can be exploited for accurate positioning. While previous methods are limited in practice as they rely on iteratively extracting a fixed number of delays, we instead formulate the delay extraction problem as a time series segmentation task. For this, we propose a pipeline built upon the U-Net convolutional neural network architecture. Unlike the state of the art our pipeline extracts an arbitrary number of delays without prior knowledge, includes a threshold for weighting between detection rate and false alarms, and does not rely on computationally demanding operations such as eigenvalue decomposition. We evaluate the presented method with synthetic data of different noise configurations and signal bandwidths and a publicly available dataset, achieving considerable performance gains w.r.t. detection performance and tracking accuracy. Furthermore, we show that the proposed method is far less computationally demanding in inference.**

***Index Terms*—Delay estimation, U-Net, deep learning.**

I. Introduction

In terms of radio frequency (RF) signals, the indoor positioning community has focused on received signal strength (RSS) based signaling obtained via communication standards that are widely available in consumer devices, such as Wi-Fi or BLE. Also widely applied are Time-of-arrival (ToA) or Time-Difference-of-arrival (TDoA) based systems of different bandwidths relying on multilateration or similar geometrical concepts assuming a line-of-sight (LOS) connection [1], [2].

However, the transmission channels in multi-path-rich environments contain more exploitable spatial information in the form of multipath components (MPCs), which are ignored by classical LOS approaches and strongly compressed to a single value in RSS signaling. MPCs contain a variety of spatial information [3], especially being present in the ultra-wideband (UWB) radio channel [4] due to their high spatial resolution. However, to use this information explicitly in multipath-assisted positioning [5], [6] requires the extraction of the signal delays of distinct propagation paths from the received channel information. Thus, the extracted delays are associated with objects in the environment or enclosing reflecting surfaces to generate hypotheses for virtual anchors that can be exploited for positioning.

For robust and computationally efficient tracking it is hence desirable to accurately extract these delays. Unlike classical ToA extraction, each received signal may contain an arbitrary number of (overlapping) relevant propagation paths in the time domain. Hence, the extraction is a challenging multi-step task including (1) the estimation of the number of MPCs and (2) the extraction of the associated signal delays (often iteratively [7]). The signals-of-interest are superimposed with each other with diffuse MPCs through the limited signal bandwidth [8].

In contrast to the state of the art, which estimates the MPCs by peak detection, we formulate the MPC delay extraction as a time series segmentation task that we solve using deep learning. We apply a U-Net convolutional neural network (CNN) architecture [9] for segmentation to achieve a more precise and robust MPC extraction. For a fair benchmark against state-of-the-art approaches we use a simulation pipeline that generates close-to real-world channel state information (CSI) with corresponding reference MPC annotations. We also adapt and apply our method to a publicly available dataset used for previous studies. Our methods achieve a significant gain on the state of the art [7] in detection and extraction performance with significantly smaller computational complexity.

The rest of this paper is organized as follows. Sec. II discusses related work before Sec. III formalizes the problem and reviews common modeling for diffuse multipath channels. We describe the details of our approach in Sec. IV and present experimental setup in Sec. V followed by the results of the evaluation in Sec. VI. Sec. VII concludes our findings.

II. Related Work

While the spatial information contained in RF channel information has long been used implicitly for RSS fingerprinting methods, the high bandwidth of UWB systems allows for the explicit use and extraction of environment-related propagation

paths [4]. This spatial information is beneficial for positioning performance [2], [3], [5], [6], [10]–[12].

State-of-the-art delay estimation algorithms for wideband positioning systems rely on the extraction of MPCs, which has been studied for many years [13]–[16]. In general, the task is viewed as a two-step process. First, the number of MPCs (or *sources*) is estimated from the autocorrelation matrix of the channel transfer function. Classically, this can be achieved by approaches based on eigenvalue decomposition, e.g., minimum description length (MDL) [15], [17]. However, recent approaches based on model-assisted deep neural networks have shown to outperform these classical approaches [18]. Second, the delay of each MPC is estimated. However, prominent approaches such as multiple signal classification (MUSIC) [15], [17] or ESPRIT [19] are both computationally expensive and inaccurate in dense multipath scenarios. To improve the accuracy of these snapshot-based algorithms, Jost et al. [20] proposed a framework to track MPCs over time, outperforming classical snapshot algorithms. However, the problem remains that such classical approaches heavily deteriorate in performance in dense multipath environments [13]. Thus, Kulmer et al. [7] recently presented an algorithm for delay extraction that provides a performance gain over classical approaches in dense multipath situations. However, as with other approaches for MPC extraction, their approach requires an initial estimate of the number of MPCs (e.g., estimated through the MDL algorithm) and its computational complexity increases significantly with a higher initial estimate for the number of MPCs. Apart from the two-step approaches, He et al. [21] studied different unsupervised clustering methods for the separation of MPC clusters rather than the delays associated with distinct MPCs. In contrast, we consider MPC delay extraction as a supervised time series segmentation problem. Hence, our method does not require an a priori estimate of the number of MPCs and still produces valid results. For this, we investigate the U-Net deep neural network architecture, originally developed by Ronneberger et al. [9]. Our approach adapts U-Net for MPC delay extraction by peak extraction and therefore implicitly estimates the source number. It relies on the segmentation of single-channel ultra-wideband measurements in the time domain and an adapted U-Net architecture.

III. Problem Description

We explain the spatial information contained in wideband signals using a simplified 2-dimensional example with a transmitter $\boldsymbol{p}_T$ and a receiver $\boldsymbol{p}_R$ as depicted in Fig. 1. Both nodes use omnidirectional antennas. The transmitter sends a signal $s(t)$ of pulse length T_p in the UWB baseband to the receiver. The received signal $r(t)$ can then be modeled as [3]:

$$r(t) = s(t) * h(t) + w(t), \tag{1}$$

where the first part describes the spatial component (a convolution of the channel impulse response (CIR) $h(t)$ with $s(t)$) and $w(t)$ describes non-spatial noise (such as sensor noise or limitations of the transmission and signal processing chain) that can be modeled as a temporally uncorrelated stationary

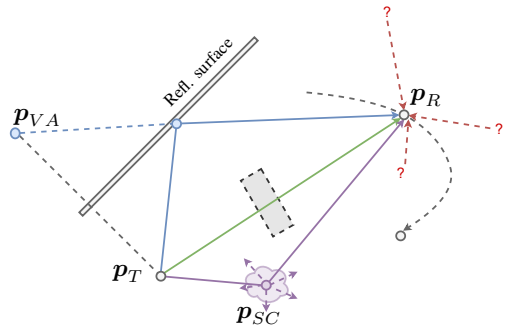

Fig. 1. Schematic illustration of the spatial information contained in radio channel measurements. Green arrow: LOS component with a potential blockage indicated by the dashed box; blue line: specular reflection at a reflecting surface with the location of the associated virtual anchor at $\boldsymbol{p}_{VA}$; purple arrow: strong point-like scatterer; red dashed lines: diffuse multipath ($\nu(t)$).

random process $w(t) \sim \mathcal{N}(0, \sigma_w^2)$ with additive white Gaussian noise and a double-sided power-spectral density of $N_0/2$.

We can decompose the CIR $h(t)$ into spatially deterministic components and diffuse multipath components. The first consists of $M \in \mathbb{N}_0$ distinct propagation components, each characterized by a corresponding delay τ_m and a complex amplitude coefficient α_m; $m \in [0; M]$:

$$h(t) = \sum_{m=1}^{M} \alpha_m \delta(t - \tau_m) + \nu(t), \tag{2}$$

In line with related work [7], we address the LOS component as a MPC for the extraction task in the following. $\nu(t)$ contains additional diffuse MPCs caused by environment interaction such as diffraction and scattering, and that interferes with the spatially deterministic components. A stochastic model for $\nu(t)$ is as a zero-mean random process characterized by a power delay profile $S_\nu(\tau)$, so that $S_\nu(\tau)\delta(\tau - u) = \mathbb{E}\{\nu(t) * \nu(u)^*\}$. Hence, while $\nu(t)$ is random, it is quasi-stationary w.r.t. space, i.e., the statistical behavior of $\nu(t)$ stays similar within the closer environment [3]. Depending on the environment, $S_\nu(\tau)$ may vary significantly. Positioning approaches exploit deterministic components by either extracting

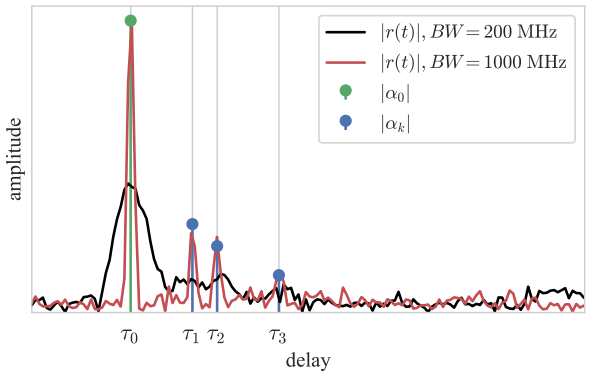

Fig. 2. Examples of the absolute value for received signals $r(t)$ for two different bandwidths with delays τ_m and amplitudes α_m, $m \in \{0, 1, 2, 3\}$

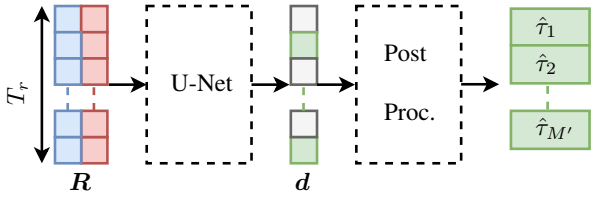

Fig. 3. Time series segmentation based approach to delay estimation.

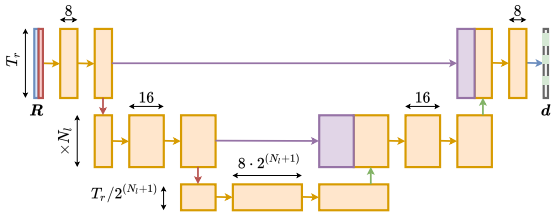

Fig. 4. Architecture of the proposed U-Net. The number on top of the box describes the number of channels. The yellow arrows describe an 1D convolution with ReLU activation function and varying kernel sizes. The red arrows describe a max pooling by the factor 2. The green arrows describe an up-convolution with kernel size 2, which also reduces the number of channels by half. The purple arrows and adjacent blocks indicate a copy operation. The blue arrow describes an 1D convolution with sigmoid activation function and kernel size 1.

the LOS component to use it for lateration or for multipath-assisted tracking (MAT) [2] or by using all of the extracted MPCs for, e.g., channel simultaneous localization and mapping (CSLAM) [5], [6]. In MAT and CSLAM the extracted MPC delays are used for tracking by associating them with virtual anchors that are implied by reflecting surfaces or point scatterers, see also the virtual anchor $\boldsymbol{p}_{VA}$ that is associated to $\boldsymbol{p}_T$ and the reflecting surface in Fig. 1. For a correct association and tracking we must obtain accurate estimates of all the relevant signal delays.

Furthermore, it is desirable to detect all relevant components in $r(t)$ while limiting the number of *false alarms*, i.e., delays associated with wrongly detected components. Hereby, the main challenge is signal overlap caused by diffuse multipath $\nu(t)$ and the limited bandwidth of $s(t)$ limiting temporal resolution of all components of $h(t)$ [4]. The bandwidth influence is indicated in the exemplary channel measurement shown in Fig. 2: at a lower bandwidth of 200 MHz we cannot differentiate τ_1 from τ_2. M can even vary greatly within an application due to difference in the propagation environment.

Hence, the number of MPCs M has to be estimated a priori or set reasonably high, which obviously may cause false positive detections. We therefore formulate delay extraction as a time-series segmentation task to implicitly and jointly estimate the number and the corresponding delays with a U-Net convolutional neural network to exploit features that are hard to model analytically and hidden under the complexity of the channel influence.

IV. Delay Estimation via Time Series Segmenation

We consider the extraction of delays from $r(t)$ as a time-series segmentation problem, as depicted in Fig. 3: The $T_r \times 2$ concatenated input signal $\boldsymbol{R} = [\Re\{\boldsymbol{r}\}, \Im\{\boldsymbol{r}\}]$, where $\boldsymbol{r}$ is the vector representation of $r(t)$, is segmented by the U-Net, resulting in a $T_r \times 1$ vector $\boldsymbol{d}$ indicating a probability of MPC presence, , i.e., $d_i \in \{0, 1\}\ \forall\ d_i \in \boldsymbol{d}$. $\boldsymbol{d}$ is then processed to extract the M' delay estimates $\hat{\tau}_\mu\ \mu \in \{1, ..., M'\}$.

A. U-Net architecture.

The U-Net architecture [9], [22], [23] employs a symmetrical architecture consisting of a U-shaped encoder-decoder scheme (see also Fig. 4): The encoder (*contracting path* on the left side) reduces the dimensionality in each layer while the decoder (*expansive path* on the right side) increases it. The key idea of U-Net is to keep high-level features in early layers of the decoder. For this purpose, long skip connections are used to localize the segmentation.

We adapted U-Net for our multipath delay estimation problem in the following way, as depicted in Fig. 4: The input tensor $\boldsymbol{R}$ is interpreted by the U-Net as 2 input channels. First, it is followed by 2 1D convolutions with 8 channels, both of kernel size N_k. In both the encoder and decoder, we use a total of $N_l + 2$ of the sets of 2 convolutional layers with max pooling by kernel size 2 in the encoder (halving the layer size) and up-convolution layers with kernel size 2 in the decoder, restoring the original size. For each consecutive layer, we double the number of channels and subtract the convolutional kernel sizes by 2 (until a minimum of 3) accompanying the halving of the layer size. The layer at the bottom between encoder and decoder forms the bottleneck of the system. A skip connection connects the last convolutional layers to the corresponding decoder layers, effectively copying over the last layer of the set. For training, we use a cross entropy loss that compares the output $\boldsymbol{d}$ to the ground truth vector of the same size. To find an optimal set of hyperparameters, we conduct a gridsearch over $N_l \in \{1, 2, 3\}$ and $N_k \in \{7, 11, 17\}$ for the datasets that are explained in detail in Sec. V.

B. Post-processing

We finally need to extract the actual delay estimates $\hat{\tau}_\mu$ from $\boldsymbol{d}$. This is depicted for an exemplary $|r(t)|$ in Fig. 5: The U-Net output generates distinguishable peaks in the time domain that are easily detectable. Hence, we can apply standard peak extraction based on simple value comparison to obtain a set of distinct peaks from the segmented time series. For clarity, this kind of peak extraction is significantly less computationally demanding than state-of-the-art peak extraction methods applied directly on $r(t)$ such as [7]. We also apply a threshold θ to the magnitudes of the extracted peaks, allowing for a trade-off between the detection of peaks and the robustness against false alarms. This is possible due to the sigmoid activation function assuring that all values are in the range of $\boldsymbol{d} \in [0; 1]$. The delays corresponding to the peaks above the threshold then constitute the M' delay estimates $\hat{\tau}_\mu$ of our proposed method.

V. Used Datasets

A. Simulation Setup

For simulation, we employ a statistical model based on the distinction of deterministic signal components from other diffuse MPCs (i.e., $\nu(t)$ in (2)), inspired by the implementation provided by Kulmer et al. in [7]. This distinction is not clearly defined and depends on the application. Letinger et al. [6] define specular reflections caused by enclosing walls as usable, while Gentner et al. [5] also include MPCs produced by point-like scatterers (such as pillars or lamp posts).

To distinguish between (α_k, τ_k) and $\nu(t)$, we use power ratios to describe the quality of channel measurements and define limits such as the Cramer-Rao lower bound: The signal-to-interference ratio SIR_k is the power ratio of a distinct deterministic component (α_k, τ_k) to diffuse spatial components, represented by the outcomes of $S_\nu(\tau_k)$, scaled by the effective pulse duration T_p [3]. The signal-to-noise ratio (SNR) describes the power ratio of (α_k, τ_k) to non-spatial noise (which is distributed equally over time) $\mathrm{SNR}_k = \frac{|\alpha_k|^2}{N_0}$. For theoretical considerations, both are usually combined into the signal-to-interference-plus-noise ratio [3]:

$$\mathrm{SINR}_k = \frac{|\alpha_k|^2}{N_0 + T_p S_\nu(\tau_k)}. \tag{3}$$

These ratios, however, are only applicable for theoretical considerations of single MPCs as they require a clear distinction between the signal components that is not available in recorded scenarios without the estimation of the power delay profile of the diffuse components. Also, for each τ_k, α_k and $S_\nu(\tau_k)$ varies significantly, resulting in different power ratios. Hence, for simulating and characterizing our channel measurements we use the SIR and SNR as scaling parameters for the spatial and non-spatial noise components in (2):

$$r(t) = r_{det}(t) + 10^{\frac{-\mathrm{SIRN}}{20}} \nu(t) + 10^{-\frac{\mathrm{SNRN}}{20}} n(t), \tag{4}$$

therefore assigning normalized SIR and SNR values to $r(t)$ instead of each separate deterministic component within. We draw $n(t)$ and $\nu(t)$ from standard normal distributions and process them through the signal processing chain, see Eq. (2). For our considerations we define the normalized SIR SIRN and normalized SNR SNRN so that 0 dB corresponds to the power of the (theoretical) LOS peak of the signal based on the geometric distance between RX and TX.

B. Synthetic Dataset

Our simulation setup geometry consists of uniformly distributed transmitter and receiver nodes and specular reflection points within a $15\ \mathrm{m} \times 15\ \mathrm{m}$ square. The number of the reflection points is sampled from a Poisson distribution with mean $\lambda = 3$ to capture a realistic setup within a rectangular room with reflecting walls and partial obstruction. The reflection coefficients are sampled from a uniform distribution between $[0.3, 0.8]$, which reflects realistic material properties [24] at the chosen center frequency of 4 GHz, which is typically used in UWB signaling. 50% of the generated signals include a LOS component (i.e., $\epsilon_0 = 1$), the remaining do not (i.e., $\epsilon_0 = 0$). Diffuse multipath components are modeled with different SIRNs $\in \{20, 30, 40, 50\}$ dB. The signals are simulated with a raised cosine pulse at different bandwidths $\in \{200\ \mathrm{MHz}, 500\ \mathrm{MHz}, 1\ \mathrm{GHz}\}$. The non-spatial noise components are simulated using SNRNs $\in \{20, 30, 40, 50\}$ dB. For each combination of bandwidth, SIRN, and SNRN we created $100{,}000$ datapoints, separated into 70% training, 10% validation, and 20% test data. Additionally, we evaluate a dataset of $100{,}000$ mixed $\mathrm{SNR}/\mathrm{SNR}$ with the same split.

C. Real-world Dataset

To evaluate the validity of our model with real data, we applied it on the dataset published by Kulmer et al. [25], recorded in a room with 4 walls and an available LOS connection. The dataset consists of 420 channel measurements and label sets. As this dataset is too small to train the neural network alone, we train different U-Nets with simulated $100{,}000$ test and $10{,}000$ validation datapoints of the same sampling frequency (6.95 GHz) and bandwidth (416.7 MHz) as the real-world dataset, with $\mathrm{SIRN} \in [30, 36]$ dB and $\mathrm{SNRN} = 45$ dB. We then adapt the pre-trained U-Nets to the real-world dataset via retraining them (on synthetic) with data from the real-world dataset. We split the real-world dataset into 200 for training, 50 for validation, and 170 test datapoints for evaluation.

VI. Results

A. Performance metrics

To evaluate the accuracy of all delay estimates of a received signal $r(t)$, we must consider the fit of M' estimated delays $\hat{\tau}_\mu\ \forall\ \mu \in \{1, ..., M'\}$ with the M actual delays $\tau_m\ \forall\ m \in \{1, ..., M\}$. Thus, we apply the Munkres algorithm [7] to derive aligned pairs of the likeliest fits $(\hat{\tau}_\mu, \tau_m)$. Fig. 5 shows an example of results of the proposed delay estimation algorithm that includes all relevant errors: The example consists of four correctly detected MPCs $(\hat{\tau}_1, \tau_1)$, $(\hat{\tau}_2, \tau_2)$, $(\hat{\tau}_3, \tau_3)$, $(\hat{\tau}_4, \tau_5)$, one false alarm $\hat{\tau}_5$ and one undetected MPC τ_4.

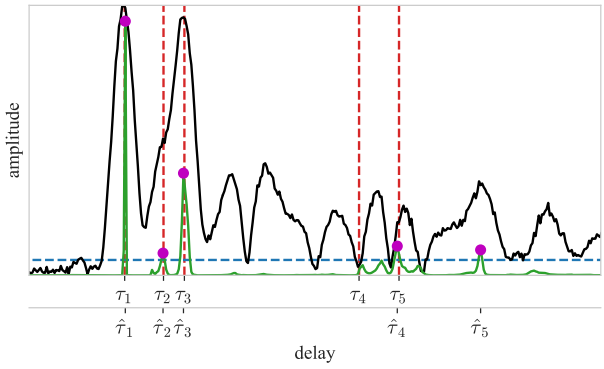

Fig. 5. Example normalized output of the U-Net (green line) with estimated delays (magenta points, $\hat{\tau}_\mu$) for given received signal magnitude $|r(t)|$ (black line) and real delays (red dashed lines, τ_m). Estimated delays are the local maxima of the output signal which are above a threshold value θ (blue dashed line).

TABLE I
MEAN AND 90TH PERCENTILE OF ADE IN cm.

(SNRN, SIRN)	real-world		(30, 20)		(40, 30)		(50, 40)		mixed	
	mean	P90	mean	P90	mean	P90	mean	P90	mean	P90
K-MDL	-	-	13.2	44.6	7.6	14.8	4.8	9.7	8.9	19.2
K-COR	12.8	34.9	19.1	63.8	8.8	18.2	4.9	9.6	9.2	20.4
K-FIX 10	13.9	37.5	24.9	72.5	12.9	42.6	6.3	12.3	13.0	40.7
K-FIX 15	12.9	32.3	25.8	71.3	14.3	47.9	6.9	13.8	14.4	46.7
U-Net-S-0.01	11.4	27.0	22.7	64.5	13.1	43.5	6.8	13.0	13.7	44.3
U-Net-S-0.1	14.1	39.6	6.7	12.0	4.9	9.6	3.2	6.3	4.5	9.4
U-Net-F-0.01	7.1	16.5	-	-	-	-	-	-	-	-
U-Net-F-0.1	7.0	15.6	-	-	-	-	-	-	-	-

Absolute distance error (ADE). For each aligned pair $(\hat{\tau}_\mu, \tau_m)$, the absolute difference $d_{\mu,m} = c|\hat{\tau}_\mu - \tau_m|$ is calculated over the complete set of delays of all $r(t)$. However, ADE only considers assigned delays, so additional performance measures have to be considered to include detection properties.

Detection rate (DR). If $d_{\mu,m} > 1$ m, the MPC is considered *undetected*. We therefore include the ratio of delay estimates correctly assigned to true MPCs as DR (true positive rate).

Mean number of false alarms (MNFA). MNFA is the mean number of $\hat{\tau}_\mu$, that are not assigned to a corresponding τ_m, per datapoint (false positives).

Thresholding (θ). Adapting θ allows for a trade-off between DR and MNFA. We visualize this trade-off, similar to a receiver operating characteristic (ROC).

Computational runtime. For a fair comparison, we determine the inference time of the baseline and U-Net (batch size = 1) on a single Intel(R) Xeon(R) Gold 5120 CPU at 2.20 GHz. We calculated the mean computation time for 10,000 random samples of the mixed dataset.

B. Baseline method

As a baseline, we use the state-of-the-art approach by Kulmer et al. [7]. Since it requires the number of MPCs M' to extract their delays, we use the MDL algorithm (K-MDL) [15] to estimate it. We also provide the baseline with the known correct number of MPCs (K-COR) and of MPCs (K-FIX) $\in (10, 15)$ large enough to assure that it is smaller than M. For clarity, since our approach is based on time-series segmentation, it does not require a previous estimate of M.

C. Synthetic Data

For our simulated datasets, we evaluated a variety of SNRN/SIRN combinations as detailed in Sec. V. Overall, for all methods, a higher signal bandwidth results in higher overall performance in both detection and delay estimation. Thus, we only show the results of the best-performing methods for a bandwidth of 500 MHz in Fig. 6. With decreasing noise, the results expectably improve for all methods. For every combination, the U-Net outperforms all baseline methods w.r.t. detection performance. Regarding the estimation error, U-Net achieves similar results as K-FIX 10 and K-FIX 15 as in Table I. However, with a larger θ of 0.1, our methods outperforms all baselines by a considerable margin. The comparatively low ADEs for K-MDL have to be seen in connection with the low detection rate: MDL only selects the most significant peaks, resulting in the increased accuracy. An U-Net architecture with $N_l = 3, N_k = 7$ yields the best results on the mixed dataset. Instead, for the datasets with specific SIRN/SNRN combinations, different $N_l \in \{1, 2, 3\}$ and $N_k \in \{11, 17\}$ performed best. Our method also outperforms K-FIX significantly.

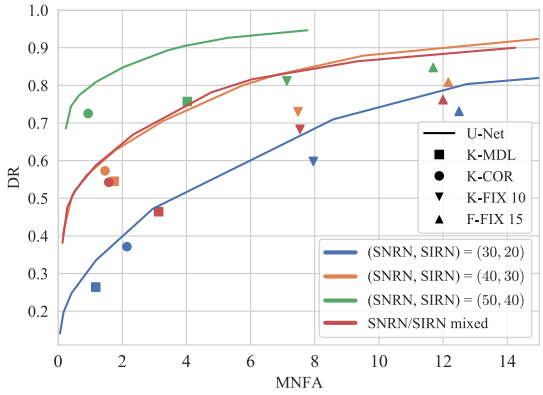

Fig. 6. MNFA vs. DR for the synthetic dataset, bandwidth = 500 MHz. The best-performing U-Net architecture for each combination was used (θ varies between 0.006 and 0.3).

D. Real-world Data

To evaluate the performance of the methods on the Kulmer et al. dataset [25], we (1) train U-Net-S on synthetic data and test its performance on the Kulmer dataset and (2) fine-tune U-Net-S on a small portion of the Kulmer dataset (see Sec. V) to test it (U-Net-F) on the remaining data. Fig. 7 shows the detection results for the baseline methods and the best U-Nets ($N_l = 2, N_k = 17$): for higher $\theta \gtrapprox 0.02$, U-Net-S exhibits a worse detection performance than K-COR and K-FIX 10. However, for lower θ the trade-off is similar to K-FIX 15.

In contrast, U-Net-F outperforms the baseline even if it is provided the correct number of sources (K-COR). Again, the threshold allows for a weighting and so it offers higher flexibility. U-Net-S shows similar delay estimation accuracy as K-FIX, see Table I. Unlike the synthetic dataset, MDL provided unrealistically high estimates M' and therefore was excluded from the evaluation. In terms of estimation accuracy, the 90th percentile of the ADE for our method is below 17 cm for all thresholds, which shows a significant improvement against the baseline, even for K-FIX 15.

E. Runtime

Since U-Net only applies matrix multiplication and the baseline methods apply computationally demanding Fourier transforms and eigenvalue decompositions. Table II shows the

TABLE II
MEAN RUNTIME ON THE MIXED DATASET.

	time (s)	U-Net	
K-MDL	0.4513	$N_k = 11$	time (s)
K-COR	0.1108	$N_l = 1$	0.00247
K-FIX 10	0.2191	$N_l = 2$	0.00358
K-FIX 15	0.3021	$N_l = 3$	0.00475

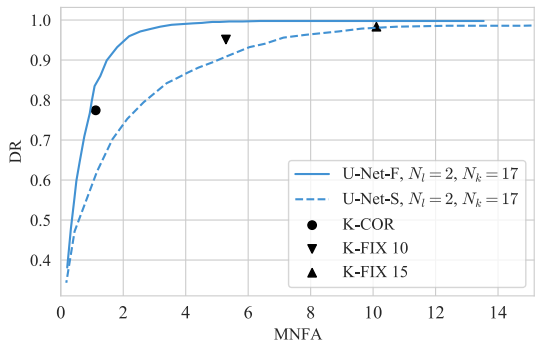

Fig. 7. MNFA vs. DR for recorded data (θ varies between 0.001 and 0.3).

mean runtime for a single inference: U-Net has a significantly lower inference time than the baseline algorithms, as it runs up to 183 times (95 for the largest U-Net architecture) faster than K-MDL and up to 45 (23 for the largest U-Net architecture) time faster than K-COR. The difference is most significant for a previous estimation of M' (K-MDL), as our approach does not require this additional and computationally demanding step. The runtime for the U-Net expectably increases with N_l and N_k.

VII. Conclusion

We present a novel method to extract multipath delays from ultra-wideband channel measurements. We formulate the process of delay extraction as a time series segmentation task using a U-Net CNN and post-processing steps. Unlike the state of the art, our approach does not require an a-priori estimate of the number of multipath components and can be tuned for higher detection ratio or a lower number of false alarms using a single parameter. We present results for a realistic synthetic dataset of different bandwidths and noise characteristics, outperforming the state of the art in both detection and estimation accuracy. Training our U-Net on realistic synthetic data and fine-tuning it on real-world data [25] yields a high detection performance and estimation accuracy and outperforms state-of-the-art methods with a mean distance estimation error of up to 7 cm. After a computationally intensive offline training, our U-Net provides MPC delay estimates about up to 183 times faster than state-of-the-art methods.

Acknowledgment

This work was supported by the Bavarian Ministry for Economic Affairs, Infrastructure, Transport and Technology through the Center for Analytics Data Applications (ADA-Center) within the framework of "BAYERN DIGITAL II". We thank our colleague Phuong Duong for her help in implementation.

References

[1] D. Dardari, P. Closas, and P. Djuric, "Indoor tracking: Theory, methods, and technologies," *IEEE Trans. Vehicular Technology,*, vol. 64, pp. 1263–1278, 04 2015.

[2] S. Aditya, A. F. Molisch, and H. M. Behairy, "A survey on the impact of multipath on wideband time-of-arrival based localization," *Proceedings of the IEEE*, vol. 106, no. 7, 2018.

[3] E. Leitinger, P. Meissner, C. Rüdisser, G. Dumphart, and K. Witrisal, "Evaluation of position-related information in multipath components for indoor positioning," *IEEE J. Selected Areas in Communications*, 2015.

[4] A. Molisch, "Ultra-wide-band propagation channels," *Proceedings of the IEEE*, vol. 97, 03 2009.

[5] C. Gentner, T. Jost, W. Wang, S. Zhang, A. Dammann, and U.-C. Fiebig, "Multipath assisted positioning with simultaneous localization and mapping," *IEEE Trans. Wireless Communications*, vol. 15, 09 2016.

[6] E. Leitinger, F. Meyer, F. Hlawatsch, K. Witrisal, F. Tufvesson, and M. Win, "A belief propagation algorithm for multipath-based slam," *IEEE Trans. Wireless Communications*, vol. 18, 12 2019.

[7] J. Kulmer, S. Grebien, E. Leitinger, and K. Witrisal, "Delay estimation in presence of dense multipath," *IEEE Wireless Communications Letters*, 2019.

[8] T. Feigl, E. Eberlein, S. Kram, and C. Mutschler, "Robust ToA-estimation using convolutional neural networks on randomized channel models," in *Proc. 11th Intl. Conf. Indoor Positioning and Indoor Navigation*, Lloret de Mar, Spain, 2021.

[9] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *LNCS*, vol. 9351, 10 2015.

[10] S. Kram, M. Stahlke, T. Feigl, J. Seitz, and J. Thielecke, "Uwb channel impulse responses for positioning in complex environments: A detailed feature analysis," *Sensors*, vol. 19, 2019.

[11] M. Stahlke, S. Kram, F. Ott, T. Feigl, and C. Mutschler, "Estimating toa reliability with variational autoencoders," *IEEE Sensors Journal*, 2021.

[12] M. Stahlke, S. Kram, C. Mutschler, and T. Mahr, "Nlos detection using uwb channel impulse responses and convolutional neural networks," in *Intl. Conf. on Localization and GNSS*, Gothenburg, Sweden, 2020.

[13] A. Richter, *On the Estimation of Radio Channel Parameters: Models and Algorithms (RIMAX)*. PhD thesis, 05 2005.

[14] C. Falsi, D. Dardari, L. Mucchi, and M. Win, "Time of arrival estimation for uwb localizers in realistic environments," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, 12 2006.

[15] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[16] J. Ehrenberg, T. Ewart, and R. Morris, "Signal-processing techniques for resolving individual pulses in a multipath signal," *Journal of the Acoustical Society of America*, vol. 63, 1978.

[17] L. Zhou, G. Li, Z. Zheng, and X. Yang, "Toa estimation with cross correlation-based music algorithm for wireless location," in *Proc. 4th Intl. Conf. Communication Systems and Network Technologies*, 2014.

[18] Y. Yang, F. Gao, C. Qian, and G. Liao, "Model-aided deep neural network for source number detection," *IEEE Signal Processing Letters*, vol. 27, 2020.

[19] R. Roy, A. Paulraj, and T. Kailath, "Esprit-a subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 34, 1986.

[20] T. Jost, W. Wang, U.-C. Fiebig, and F. Perez Fontan, "Detection and tracking of mobile propagation channel paths," *IEEE Trans. Antennas and Propagation*, vol. 60, 10 2012.

[21] R. He, Q. Li, B. Ai, Y. L.-A. Geng, A. F. Molisch, V. Kristem, Z. Zhong, and J. Yu, "A kernel-power-density-based algorithm for channel multipath components clustering," *IEEE Trans. Wireless Communications*, vol. 16, no. 11, pp. 7138–7151, 2017.

[22] M. Perslev, M. Jensen, S. Darkner, P. J. r. Jennum, and C. Igel, "U-time: A fully convolutional network for time series segmentation applied to sleep staging," in *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[23] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, 2018.

[24] O. Landron, M. Feuerstein, and T. Rappaport, "A comparison of theoretical and empirical reflection coefficients for typical exterior wall surfaces in a mobile radio environment," *IEEE Trans. Antennas and Propagation*, vol. 44, no. 3, 1996.

[25] J. Kulmer, S. Hinteregger, B. Großwindhager, M. Rath, M. S. Bakr, E. Leitinger, and K. Witrisal, "Using decawave uwb transceivers for high-accuracy multipath-assisted indoor positioning," in *2017 IEEE Intl. Conf. Communications Workshops*, 2017.