Real-time Outdoor Localization Using Radio Maps: A Deep Learning Approach

Çağkan Yapar, *Graduate Student Member, IEEE,* Ron Levie,

Gitta Kutyniok, *Senior Member, IEEE,* and Giuseppe Caire, *Fellow, IEEE*

Abstract

Global Navigation Satellite Systems typically perform poorly in urban environments, where the likelihood of line-of-sight conditions between devices and satellites is low. Therefore, alternative location methods are required to achieve good accuracy. We present LocUNet: A convolutional, end-to-end trained neural network (NN) for the localization task, which is able to estimate the position of a user from the received signal strength (RSS) of a small number of Base Stations (BS). Using estimations of pathloss radio maps of the BSs and the RSS measurements of the users to be localized, LocUNet can localize users with state-of-the-art accuracy and enjoys high robustness to inaccuracies in the estimations of radio maps. The proposed method does not require generating RSS fingerprints of each specific area where the localization task is performed and is suitable for real-time applications. Moreover, two novel datasets that allow for numerical evaluations of RSS and ToA methods in realistic urban environments are presented and made publicly available for the research community. By using these datasets, we also provide a fair comparison of state-of-the-art RSS and ToA-based methods in the dense urban scenario and show numerically that LocUNet outperforms all the compared methods.

Index Terms

Wireless localization, radio maps, pathloss, deep learning, dataset.

Çağkan Yapar is with the Institute of Telecommunication Systems, TU Berlin, 10623 Berlin, Germany (e-mail: cagkan.yapar@tu-berlin.de).

Ron Levie is with the Faculty of Mathematics, Technion - Israel Institute of Technology, 3200003 Haifa, Israel (e-mail: levieron@technion.ac.il).

Gitta Kutyniok is with the Department of Mathematics, LMU Munich, 80331 München, Germany, and also with the Department of Physics and Technology, University of Tromsø, 9019 Tromsø, Norway (e-mail: kutyniok@math.lmu.de).

Giuseppe Caire is with the Institute of Telecommunication Systems, TU Berlin, 10623 Berlin, Germany (e-mail: caire@tu-berlin.de).

I. INTRODUCTION

The location information of a User Equipment (UE) is essential for many current and envisioned applications that range from emergency 911 services [2], autonomous driving [3], intelligent transportation systems [4], proof of witness presence [5], 5G networks [6], to social networks, asset tracking and advertising [7], just to name a few.

In urban environments, Global Navigation Satellite Systems (GNSS) alone may fail to provide a reliable localization estimate due to the lack of line-of-sight conditions between the UE and the GNSS satellites [8]. In addition, the continuous reception and detection of GNSS signals is one of the dominating factors in battery consumption for hand-held devices. It is thus necessary to resort to other complementary means to achieve localization with the desired high accuracy.

In cellular networks, the position of UE can be estimated by using various statistics that the UE may report or the network can infer. The most prominent localization methods in the literature are based on Time of Arrival (ToA) [9], [10], Time Difference of Arrival (TDoA) [11], Angle of Arrival (AoA) [12] and Received Signal Strength (RSS) [13] measurements.

A. Received Signal Strength (RSS)

Received signal strength quantifies the received power, averaged over a certain time interval and frequency band, of the beacon slots broadcast periodically by base stations (BSs) [14]–[16]. For such measurements coherent reception is not necessary, and small scale fading fluctuations are alleviated due to the time-frequency averaging. Since the transmit power of the BS on its beacon signal slots is fixed and known, the received signal strength is a function of the pathloss of the propagation between the BS and the UE. In fact, received signal strength measurements of beacon signals are routinely performed by UEs and reported to the system as "received signal strength indicators" (RSSI). Reporting RSS information is a standard feature in most current wireless protocols. For example, this is used to trigger handovers and to enable UE-BS association for load balancing purposes. Therefore, exploiting RSS information for localization is attractive since it is a feature already built-in in the wireless protocols and does not require any further specific hardware at the UE, whereas the time-based (ToA and TDoA) and angle-based (AoA) methods require high precision clocks and antenna arrays, respectively [17], [18].

Remark 1: *For ease of exposition, throughout the paper we consider a cellular network scenario with BSs. Essentially, any wireless signal source with known location, e.g., WiFi-Hotspots, can be utilized.*

B. Ranging-Based Methods

In ranging-based techniques, the distances between the UE and the BSs are used to esti-mate the position of UE by lateration. Here, the distances are estimated by using RSS or ToA measurements, based on a statistical signal attenuation or time delay model to estimate the distance between the UE and the BS. However, using such models is not appropriate in urban settings, since in practice the signal undergoes diverse propagation phenomena such as penetrations, reflections, diffractions, and wave guiding effects, due to the presence of obstacles in the environment. This leads to very large errors in the distance estimation (See e.g. [19]). We call this phenomenon *range estimation mismatch*. Several works, e.g. [9], [10], [20]–[26] proposed methods to mitigate the effects of obstructions in the range estimation mismatch. Nevertheless, these methods do not directly use a complete model of the propagation phenomena, and only partially alleviate the aforementioned problem.

In particular, ranging techniques exploiting RSS are based on standard statistical models where the pathloss is modeled as a deterministic function of the distance between the UE and the BS (radial function), with some (typically log-normal) random fluctuation. These techniques are particularly problematic in an urban setting [19] where, for a specific propagation environment, the pathloss function is not at all a radial function (see for example Fig. 3). The usage of standard statistical models for RSS-based ranging has somehow generated the belief that RSS is an unreliable statistic. The matter of the fact is, that this is not a limitation of the RSS itself, but of the statistical models that relate pathloss with distance. It is therefore evident that in order to exploit RSS for positioning, a much more refined estimation of the pahtloss function for any specific propagation environment is needed.

C. Fingerprint-Based Methods

Fingerprint-based methods, as opposed to ranging-based methods, do not impose any modeling assumption on the signal strength propagation. Instead, they rely on offline generated extensive databases of the measured "radio signatures" of the signal at different locations. Given a reported signal fingerprint, these methods infer the location of the UE by "looking up" a location with a similar fingerprint from the database (radio map, cf. Section II-A). Many fingerprint types, such as visual, radio wave (e.g. RSS or the baseband transfer function with complex channel coefficients, i.e., the so-called Channel State Information (CSI)), or motion fingerprints, can be used for localization purpose [27]. In this paper we focus on RSS signatures which, as

already argued before, are more easier to acquire (requiring time-frequency averaged non-coherent power measurements), more robust than complex baseband CSI to small-scale changes in the propagation environment, and much lower dimensional than CSI signatures. Fingerprint-based methods are well-known for outperforming the ranging-based techniques in complex urban environments [28]. Many machine learning (ML) techniques based on RSS fingerprints were applied for positioning in past work, e.g. [29], [30].

D. Physical Simulations of Radio Maps and RadioUNet

A major drawback of fingerprinting is the difficulty in generating and updating the radio signature database for the desired location. Fingerprinting requires a labor intensive, time consuming and expensive site surveying, which needs to be repeated as the environment changes. While this is possible for well-defined and somehow restricted indoor locations such as an airport hall, or a shopping mall, it is clearly difficult if not impossible to produce such data for an extended outdoor area such as a whole city. Furthermore, the outdoor propagation medium changes rapidly due to moving vehicles. Therefore, radio signatures such as baseband CSI (sensitive to small scale effects such as multipath propagation and phase rotations) are extremely fragile.

When dealing only with RSS, a more feasible alternative for generating the radio map is using physical simulation methods such as ray tracing [31]–[33], thereby bypassing the extensive measurement campaigns. Based on an approximate model of the physical signal propagation phenomena, such simulations yield very accurate predictions, especially when an accurate geometrical description of the urban environment is available, see e.g. [34], [35]. Many previous works proposed using simulated radio maps for localization, e.g. [29], [36]–[43].

However, such simulations are computationally demanding and are thus not suitable for real time applications. Recently, an efficient deep learning-based method termed *RadioUNet* [44] (see Subsection II-C) was proposed by the authors of this paper. This algorithm enables the generation of high accuracy pathloss radio maps in much shorter time, and is thus adopted as an important building block of the current work.

E. Machine Learning For Localization

Lately, several ML-based localization approaches have been proposed, e.g. [29], [36]–[43], [45]–[48], see the recent surveys [30], [49]. To the best of our knowledge, none of the previous works considered applying convolutional neural networks on pathloss radio maps. We note that the radio maps may serve as a mean to assess the likelihood of a location for the UE with

respect to a BS. However, in order to compute such likelihood in a classical sense (i.e., not using a deep-learning training-based method), one would need to characterize the a posteriori probability of the location given the radio map and the RSS measurement. This requires in turns a statistical characterization that goes well beyond the simplistic pathloss models used in wireless communications.

As opposed to the previous work, LocUNet does not require new signature generation (i.e., new RSS/CSI samples) for re-training when deployed in a new environment. Moreover, LocUNet efficiently utilizes radio map estimations (See Section III) thanks to its convolutional design (See Remark 3).

F. State-of-the-Art Urban Localization in Current Commercial Use

The industry state-of-the-art on radio localization as implemented by systems such as Google Maps and Apple Maps are based on a multitude of sensor data, including GNSS, "radio signatures" of all sort (e.g., WiFi SSIDs, cellular channel quality, detailed channel frequency response of the fading channels over the OFDM subcarriers, and so on), and these data are fused via some machine learning scheme. Since these practical implementations are ad-hoc industry intellectual property and not revealed in the open literature at a level of details sufficient to run accurate performance comparisons, it is virtually impossible to assess the performance of our proposed method against the "real-world" state-of-the-art.

In contrast, it is important to notice here that our method is based uniquely on RSS measurements from known BSs/WiFi-Hotspots, which are collected by default by standard user devices since each device continuously measures the received signal strength from BS beacon signals. Furthermore, these measurements are very robust since BS beacon signals are easy to detect and do not require any calibration or accurate timing and frequency synchronization, as they are simply wideband power measurements. Our method based on such a simple setting, still achieves state-of-the-art results with respect to the previous works and could be incorporated into some industry scheme by appropriate data fusion. This direction, however, goes beyond the scope of this paper and it is left for future work.

G. Our Contribution

Our contribution can be summarized as follows. 1) We propose an accurate and computationally efficient localization method, merely based on RSS measurements, which does not necessitate additional signal processing or hardware (e.g., calibrated antenna arrays) at the user devices. 2) Relying on the recent developments in deep learning approximations of ray tracing (RadioUNet

[44]) to generate pathloss radio map estimations, our method enjoys low run-time, which would allow real-time applications. 3) Using such radio map estimations and measured RSS values, the proposed LocUNet yields very accurate localization results, which are demonstrated by numerical experiments. The proposed method achieves about 5m of accuracy in mean absolute error, when trustworthy radio map estimates are available, which outperforms current state-of-the art. 4) The proposed method demonstrates very high robustness to the inaccuracies in the radio map estimates in a rigorously studied realistic setting. Moreover, the robustness of the method to out-of-distribution measurements, and to the inaccuracies in the location labels of the training set, are shown. 5) Extensive numerical comparisons with ToA ranging-based methods are presented, where LocUNet outperforms the state-of-the-art in different settings. 6) We justify our neural network design by a comprehensive ablation study. 7) Two synthetic datasets are introduced and set publicly available for the research community.

The rest of the paper is organized as follows. We present the preliminaries in Section II. The proposed method is explained in Section III. Numerical results are presented in Section IV, and conclusions are drawn in Section V.

II. PRELIMINARIES

In this section, we present the preliminaries that serve as the building blocks of LocUNet.

A. Pathloss and Radio Map

The *pathloss* (or *large-scale fading coefficient*), quantifies the loss of wireless signal strength between a transmitter (Tx) and receiver (Rx) due to large scale effects. The signal strength attenuation can be caused by many factors, such as free-space propagation loss, penetration, reflection and diffraction losses by obstacles like buildings and cars in the environment. In dB scale[1], pathloss amounts to $P_L = (P_{Rx})_{dB} - (P_{Tx})_{dB}$, where P_{Tx} and P_{Rx} denote the transmitted and received locally averaged power (RSS) at the Tx and Rx locations, respectively. Notice that "locally averaged" power is defined as the energy per unit time averaged over time intervals of the order of a typical transmission slot in the underlying protocol (e.g., the duration of a Resource Block in 5G NR standard [50]) and over the whole system bandwidth. Hence, the effect of the small-scale frequency selective fading is averaged out and only the frequency-flat pathloss matters. A *radio map* $R(\mathbf{x}_1, \mathbf{x}_2)$ defines the pathloss between any two points $\mathbf{x}_1$ and $\mathbf{x}_2$ in the plane. For fixed Tx position $\mathbf{x}_1 = \mathbf{x}_{tx}$, the radio map is a function of the Rx position

---

[1]$(\cdot)_{dB} := 10 \log_{10}(\cdot)$.

$\mathbf{x}_2 = \mathbf{x}_{rx}$, i.e., it can be represented as a 2-dimensional image where the value of the pathloss $R(\mathbf{x}_{tx}, \mathbf{x}_{rx})$ at each suitably discretized position $\mathbf{x}_{rx}$ corresponds to a "pixel" of the image.

B. Radio Map Simulations

One well-known class of numerical methods for solving Maxwell's equations in far-field propagation conditions is *Ray Tracing* [32]. In ray tracing, the signal is modeled as rays that are cast from the transmitter, travel in straight lines in homogeneous medium (like free space), and undergo rules of reflection, refraction and diffraction when the medium changes (e.g., when hitting an obstacle).

There are two main approaches for finding the ray paths in a given environment. Classical ray tracing searches for the paths between the transmitter and each receiver point, which necessitates very high computational time, with exponential dependency between the number of interactions and the complexity of the geometry (e.g., the number of buildings), and a linear dependency between the computation time and the number of receiver pixels [51]. The second approach, ray launching [52], discretizes the propagation angle at the transmitter and launches rays in all directions. The rays interact with the environment as before, and terminate when a predetermined maximum number of interactions is reached or when the signal strength goes below a minimum value. The received signal strength is then computed as the sum of all rays at all receiver pixels (i.e., discretized locations).

In the following, we briefly explain the simulation models used in this paper, computed using the software WinProp [33].

1) Dominant Path Model (DPM) [34]: DPM is based on the assumption that the dominant propagation path from Tx to Rx must arrive via convex corners of the obstacles to Rx, and thus only diffractions are taken into account, omitting reflections and penetrations altogether. The model assigns the pathloss of the ray with the largest received power to each receiver point. This is equivalent to finding the shortest free space path to each point, and computing the corresponding pathloss. The formula of the pathloss is given by some statistical model, based on the following parameters as inputs: The length of the path, a parameter that depends on the distance and visibility between the Tx and Rx, and a diffraction loss function, which depends on the angle of the diffraction and the number of previous diffractions in the path. The model can also consider waveguiding pathloss effects when calculating the pathlosses of the paths, where reflection loss of the walls along the path as well as their distance to the path influence the value. This simplified model of the signal propagation allows a very efficient implementation

with respect to more traditional ray tracing methods, while demonstrating good accuracy in many real-world urban propagation scenarios [34].

2) Intelligent Ray Tracing (IRT) [51]: This algorithm starts with a pre-processing step for the considered map, where the faces of the obstacles are discretized into regular tiles, their edges into segments, and the visibility relations among the centers of these elements are found. For the visible element pairs, the distance between their centers and the subtended angles are calculated. The computations of the paths, based on the ray launching approach, are then accelerated using the pre-computed visibility model. The maximum number of interactions is set by the user. In our simulations, we set the maximum number of interactions as two (at most one diffraction and one reflection). We set the tile length as 100 m, which results in having only one tile for most of the building walls and for all of the cars (cf. Section III-B for the detailed description of the dataset). High accuracy of this method was also proven by field measurements in several cities, see e.g. [35] and the references therein.

C. RadioUNet

RadioUNet is a UNet-based [53] pathloss radio map estimation method introduced in [44], [54]. In this paper we use the so called RadioUNet$_C$. The trained model takes as input the map of the city and the location of the BS (both encoded as B/W images) and returns an estimation of the corresponding radio map represented as a gray-scale image. RadioUNet demonstrates high accuracy, with root-mean-square-error of order of 1dB in various scenarios, and a run-time order of milliseconds on NVIDIA Quadro GP100 [44]. RadioUNet is trained by supervised learning to match simulations of radio maps, using the RadioMapSeer Dataset [55].

III. Pathloss Fingerprint-Based Localization with Deep Learning

In this section we present LocUNet – a deep learning localization method based on estimations of pathloss radio maps.

A. Proposed Method

Suppose that a UE with location coordinates $\mathbf{x}_U = (x_U, y_U)$ reports the strength of beacon signals (non-interfering identification signals), transmitted from a set of BSs B_j, $j = 1, \ldots, J$, with known coordinates $\mathbf{x}_{B_j} = (x_{B_j}, y_{B_j})$. We refer to such BSs as "anchor BSs" since they belong to the network of the wireless operator and, as such, their position is known. Based on the relation between transmit/receive powers and the pathloss $(P_{Rx})_{dB} = P_L + (P_{Tx})_{dB}$, the pathloss between the device and the BSs, p_j, $j = 1, \ldots, J$, can be calculated, where we assume

the small-scale fading effects are eliminated by averaging over time and system bandwidth. In our approach, the position of the UE is estimated based on the following information.

1) The pathloss values p_j, $j = 1, \ldots, J$,

2) The estimations of the radio maps for each anchor BS $R_j(x, y) := R(\mathbf{x}_{B_j}, (x, y))$, $j = 1, \ldots, J$, computed via RadioUNet,[2]

3) The map of the urban environment, the locations of the anchor BSs (which are fixed and known).

The processing is typically performed in a centralized manne in the cloud, based on the RSS information reported by the UE. Alternatively, localized processing at the UE may also be considered, but in this case the UE must be aware of the BSs location (beyond their identity) and of the corresponding radio maps. At this point we do not distinguish between these two options, which are equivalent in terms of system performance (although may not be equivalent in terms of protocol overhead, computation complexity at the UE, and privacy of the UE location with respect to the network control).

Our method, called LocUNet, computes an estimation of the UE location using a UNet variant. In order to input the above information (1)–(3) to the UNet, this is first represented as a set of 2D images.

1) The RSS values $\{p_j\}$ are converted to gray-level. Each pathloss p_j is represented as a 2D image $P_j(x, y)$ of a constant value p_j, i.e., for each j this is an all-gray uniform image, but the level of gray differs for different indices j.

2) Each radio map $R_j(x, y)$ is represented as a 2D image, with pixel values in gray-level. Radio maps are obtained by using RadioUNet, which takes the map of the urban environment, and the locations of BSs as input features [44].

3) The map of the urban environment is represented as a binary black and white image, where the interior of the buildings are white (pixel value $= 255$), and the exterior is black (pixel value $= 0$).

4) The location of each BS B_j is represented as a one-hot binary image, where the pixel at location (x_{B_j}, y_{B_j}) is white, and the rest is black.

Inputs (3) and (4) are optional for LocUNet as input features (Note that they are still needed to generate the input (2) via RadioUNet), and through numerical experiments we observe their

---

[2]From now on it is convenient to represent the locations as pixel coordinates (x, y) in the 2-dimensional plane.

TABLE I: Architecture of the first part of LocUNet (w/o final CoM layer). *Resolution* is the number of pixels of the image in each feature channel along the x, y axis. *Filter size* is the number of pixels of each filter kernel along the x, y axis. The input layer is concatenated in the last two layers before the CoM layer. $in = 10/11/15/16$, cf. Table X.

LocUNet										
Layer	**In**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Resolution	256	256	128	64	64	32	32	16	16	16
Channel	in	20	50	60	70	90	100	120	120	135
Filter size	3	5	5	5	5	5	5	3	5	5
Layer	**10**	**11**	**12**	**13**	**14**	**15**	**16**	**17**	**18**	**19**
Resolution	8	8	4	4	2	4	4	8	8	16
Channel	150	225	300	400	500	$400+400$	$300+300$	$225+225$	$150+150$	$135+135$
Filter size	5	5	5	5	4	5	4	5	4	5
Layer	**20**	**21**	**22**	**23**	**24**	**25**	**26**	**27**	**28**	**29**
Resolution	16	16	32	32	64	64	128	256	256	256
Channel	$120+120$	$120+120$	$100+100$	$90+90$	$70+70$	$60+60$	$50+50$	$20+20+in$	$20+in$	1
Filter size	3	6	5	6	5	6	6	5	5	-

negligible influence on the performance (See Section IV-H1).

The first part of LocUNet is a UNet [53], with average pooling, upsampling + bilinear interpolation, and Leaky ReLU (with parameter 0.2) as the activation function. We call the output feature map of the UNet, $H(x, y)$ a *quasi-heatmap*, as its value at a point (x, y) in the map quantifies the likelihood of the UE to be located at this point, while it can take negative values, due to LeakyReLU being the activation function of the network.

The final layer of LocUNet computes the center of mass (CoM) (μ_x, μ_y) of $H(x, y)$,

$$\mu_x = \frac{\sum_{x=1}^{256} \sum_{y=1}^{256} x H(x, y)}{\sum_{x=1}^{256} \sum_{y=1}^{256} H(x, y)}, \quad \mu_y = \frac{\sum_{x=1}^{256} \sum_{y=1}^{256} y H(x, y)}{\sum_{x=1}^{256} \sum_{y=1}^{256} H(x, y)},$$

where 256 is the number of pixels along each axis.

The architecture of LocUNet is summarized in Table I.

We measure the accuracy of the proposed method over a given subset $\mathcal{S}$ of the dataset by average Euclidean distance, i.e., mean absolute error (MAE), between the estimated UE location and the true UE location, given by

$$L_{\text{MAE}}(\mathbf{u}, \tilde{\mathbf{u}}) = \frac{1}{|\mathcal{S}|} \sum_{k \in \mathcal{S}} ||\mathbf{u}^k - \tilde{\mathbf{u}^k}||_2, \tag{1}$$

where $\tilde{\mathbf{u}^k} := (\mu_x^k, \mu_y^k)$ and $\mathbf{u}^k := (x_U^k, y_U^k)$ denote the LocUNet estimation and the corresponding true location of the kth instance of the dataset.

We also consider using the average squared Euclidean distance, i.e., mean squared error (MSE),

TABLE II: RadioLocSeer Dataset parameters.

Parameter	Value
Map size	256^2 pixels
Pixel length	1 meter
BS, UE height	1.5 meter
Center carrier frequency	5.9 GHz
Channel bandwidth	10 MHz
Noise power spectral density	-174 dBm/Hz
Transmit power	23 dBm
Noise floor (NF)	-134 dB

as a loss function, which is the average squared 2D Euclidean distance

$$L_{\text{MSE}}(\mathbf{u}, \tilde{\mathbf{u}}) = \frac{1}{|\mathcal{S}|} \sum_{k \in \mathcal{S}} ||\mathbf{u}^k - \tilde{\mathbf{u}^k}||_2^2. \tag{2}$$

When using stochastic gradient descent for training, we take $\mathcal{S} = \mathcal{B}_m$ in one of the cost functions mentioned above, where $\mathcal{B}_m$ is the m-th mini-batch of the training set.

B. Datasets

We introduce two new datasets to allow for comparisons among RSS (pathloss) and ToA ranging-based algorithms. These two new datasets, which we call *RadioLocSeer Dataset* and *RadioToASeer Dataset*, are available at [55] and https://RadioMapSeer.GitHub.io/LocUNet.html.

More details on the datasets can be found in [56].

1) RadioLocSeer Dataset: We introduce a dataset of city maps, BS locations, and the corresponding radio maps simulated by the ray tracing software WinProp [33] and by RadioUNet. The dataset is based on the test set of RadioMapSeer [44], which contains 99 maps and 80 BS per map. We report the parameters of the dataset in Table II (cf. [44] Section III.B for details). The maps are randomly split into 69 training, 15 validation and 15 test maps.

For each map we randomly generate 200 UE positions and randomly pick 5 BS positions out of the 80 available ones. We repeat the latter step 50 times for each map, to represent different scenarios of BS deployment. The BS locations are chosen to be separated by at least 20 m. Moreover, the BSs and UEs are restricted to lie in the middle 150×150 and 164×164 boxes about the center of the 256×256 grid of the RadioMapSeer images, respectively. We ensured that each UE can receive the signals of at least one BS with a power above the noise floor (NF) (cf. Table IV for the number of BSs in coverage in the test set.)

For each of the above maps and BS location, we provide: 1) the radio map estimations via the RadioUNet model trained on DPM (see Subsection IV-A); 2) the simulated (by WinProp) radio maps obtained by DPM and IRT (cf. Sec. II-B); 3) a second version of the ray-racing
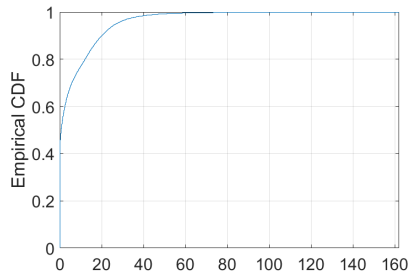
Fig. 1: Empirical CDF of the difference between the direct and dominant path in meters

simulations including the effects of cars in the streets, where each map consists of 100 cars (if the roads of the corresponding maps are of sufficient length to fit them, otherwise the maximum number that can fit) of size $2 \times 5 \times 1.5$ (width $\times$ length $\times$ height), randomly generated near and along or perpendicular to roads; 4) the B/W images of the city maps, along with the coordinates of the considered BSs and the UEs.

2) RadioToASeer Dataset: We generated another dataset based on the dominant path model (cf. Subsection II-B1) via the software *WinProp* [33], of the same settings as in RadioLocSeer Dataset, which provides ToA information of the dominant paths, to allow for comparisons between RSS and ToA ranging-based methods in realistic urban scenarios.

We call the straight line connecting Tx and Rx, which may go through obstacles (in this paper, buildings and cars), the *direct path*. The difference (error) between ranging with ToA measurement of the ray (range is calculated as: ToA $\times$ speed of light) and the direct path is called the *NLOS bias*, which is by definition non-negative (Zero when the link is in LOS, positive otherwise). The empirical cumulative distribution function (CDF) of the NLOS bias of the dataset is shown in Fig. 1. We see that 47% of the links in the RadioToASeer are in LOS.

We argue that even though the dominant path may be a NLOS path, and hence introduces an NLOS bias, it is quasi-optimal to use the ToA of such paths in ToA ranging-based algorithms. First, if the dominant path is LOS, then it is by definition the shortest path and the bias is zero. Second, in the NLOS situation, the NLOS bias of a dominant path is lower than that of a potential free space path between the same Tx and Rx, that undergoes reflections to reach its destination (cf. Section II-B1). Thus, relying on ToAs of the dominant paths would yield better performances than the ToAs of the other free space paths.

Notice that in NLOS conditions using the length (range) of the direct path (assuming building penetration) would also incur several problems. First, we note that empirical evidence shows that loss due to penetration through a building is around $15 - 20$dB [57]. In an urban scenario, the

associated direct path of an NLOS link is usually subject to blockage by numerous buildings. For devices with realistic noise figures and SNR requirements, the received power of such a path may go easily below the detection threshold. Furthermore, the bandwidth considered in the current paper does not allow for a well resolvability of the multi-path components. Indeed, in the considered setting, we have $\Delta\tau = 1/\text{W} = 100$ns, so, when multiplied by the speed of light $c \approx 3 \times 10^8$ m/s, the difference between the paths should be at least $\Delta d = c\Delta\tau = 30$m for resolvability. In our dataset, we observe that around 92% of NLOS direct paths don't satisfy this condition (cf. Fig 1), i.e., the difference of the length of the direct path and that of the dominant path is less than 30 m. Finally, whenever the direct path penetrates an obstacle, it incurs an *excess delay*

$$t_\Delta = (\sqrt{\epsilon_r} - 1)\frac{d_w}{c}, \tag{3}$$

where ϵ_r denotes relative electric permittivity and d_w stands for the thickness of the wall. Assuming a total length of 3m of penetrated wall of concrete with permittivity 4, we get an excess delay of 10 ns, which further makes the ToA of the direct and dominant path closer, deteriorating resolvability further: about 94% of NLOS direct paths don't satisfy this.

Hence, it is reasonable to ignore the direct paths and use the dominant paths instead.

Moreover, the dataset assumes exact ToA information, i.e., perfectly synchronized clocks at the BSs and UEs, and also no other additive noise on ToA measurements. Thus, all in all, evaluating the ToA ranging based-methods on this dataset yields essentially a best case of what is possible with ToA ranging in an urban environment.

Remark 2: *To the best of our knowledge, there exists no publicly-available dataset of measured radio maps reporting consistent signal strength and ToA with a fine resolution for the same environments. This is essential in order to make fair comparison of different methods, and we believe that our dataset will be useful for the research community in this area.*

C. Training

We perform supervised learning on the training set. We use *Adam* optimizer [58] with an initial learning rate of 10^{-5} and decrease the learning rate by a factor of 10 after 30 epochs. We set the total number of epochs for training as 50 and batch size as 15. To avoid overfitting, we pick the network parameters with the lowest validation error in the 50 training epochs. We used PyTorch [59] for the implementation[3].

---

[3]The code is available at https://GitHub.com/CagkanYapar/LocUNet. For reproducibility, see the compute capsule at https://codeocean.com/capsule/7149386/tree.

IV. NUMERICAL RESULTS

In this Section, we demonstrate the performance of LocUNet by numerical simulations. We assess the accuracy of LocUNet and the compared methods on the test set, namely, by MAE (1) with $\mathcal{S} = \mathcal{T}$, where $\mathcal{T}$ is the test set (cf. Subsection III-A).

First, we discuss the sources of mismatch between a true and estimated radio maps and describe two scenarios in order to investigate the robustness of our method to such realistic inaccuracies. Afterward, we illustrate these scenarios and LocUNet's estimation results under them. Then, we compare the performance of the presented method with state-of-the-art fingerprint-based methods and study the effect of number of BSs in coverage. Subsequently, we study numerically the effect of out-of-distribution measurements, and the location errors in the position labels. Then, evaluations of state-of-the-art ToA ranging-based methods are presented, which proves the superiority of LocUNet in numerous settings. Next, we present the baseline/relevant previous methods, discuss their relation with LocUNet, and present their numerical evaluations. Finally, we present the ablation study of input features, activation function of the layer before the CoM layer, and the loss function for training, which justify our choices.

All of the compared algorithms with CPU implementation were run on an Intel Core i7-8750H, and LocUNet was run on a GPU of NVIDIA Quadro RTX 6000. We report the approximate run-times of the LocUNet and the compared algorithms (all implemented using MATLAB, run on CPU), as well. For the sake of completeness, we note that none of our implementations were meant to optimize the run-time, and thus the reported run-times reflect only an indicative assessment.

A. Modeling of Mismatch in Radio Maps

As we argued before (cf. Section II-A), the RSS values are obtained by integrating the signal strength measurements over time and the channel bandwidth, which effectively averages out the small scale frequency selective fluctuations.

Given this, we identify two sources of mismatch between the actual RSS measured at the UE and the estimated radio map input to LocUNet. These are:

1) The mismatch between the real propagation phenomena and the simplified model which is used to estimate the radio maps. For example, a statistical model based on radial symmetry plus long-normal fluctuations exemplifies a very simplified model with large mismatch,

whereas models that use the geometry of the environment (e.g. ray tracing methods or their accurate approximations like RadioUNet) will yield much better radio map estimations.

2) The mismatch due to the inaccurate knowledge of the propagation environment. This inaccuracy can be decomposed into two classes: **1)** Imperfect knowledge on position, shape and material information of the objects, **2)** Changes in the environment due to the objects with mobility. While the first class of inaccuracies could be alleviated by integrating public access city maps (e.g., *OpenStreetMap* [60]) with refined information from cameras, LiDAR, and satellites, the second class of inaccuracy is not easily solvable since it would require a dynamic re-computation at a fast rate of the radio maps for the time-varying environment.

Hence, in this work we focus on the inaccuracy due to mobile objects in the propagation environment, which we modeled as the presence of cars in the streets. In particular, these are unknown in the training phase of RadioUNet, and unknown at the run time of LocUNet. However, LocUNet can learn the discrepancy between the RadioUnet estimated radio map and the actual one during training, and (seamlessly) incorporate the statistics of this error in its location estimation.

In the following we designate two scenarios with increasing compliance with reality, encompassing the above mentioned sources of inaccuracies.

1) Nominal Scenario: In this *optimistic* scenario, we assume that the real radio maps are *exactly* governed by DPM and the radio map estimates are obtained by RadioUNet, which was trained in a supervised fashion to estimate radio maps under DPM assumption, as well. Moreover, we assume that there are *no addional unknown obstacles*, i.e., the *exact* information about the propagation environment is available to RadioUNet. Hence, LocUNet enjoys having access to very high accuracy radio maps, where the inaccuracy of the available radio maps with respect to the true radio maps is solely due to the prediction error of RadioUNet, addressing the first mentioned source of inaccuracy.

2) Robustness Scenario: This scenario encompasses all the inaccuracies enumerated previously, which are likely to occur in the deployment of the method. That is to say, we take into account the potential mismatch between the real radio maps (above mentioned first source of error), which we simulate (with WinProp [33]) by IRT as the real-surrogate (from which the UE measures its pathloss), and the estimated radio maps, which are obtained by RadioUNet that was trained on DPM dataset. We also add the effects of unknown obstructions to the real-surrogate IRT simulations, addressing the second above mentioned source of error. We model
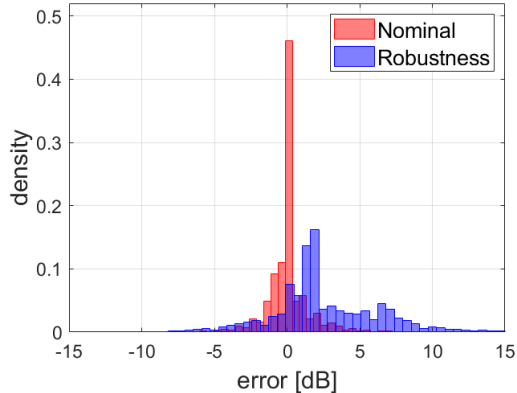
Fig. 2: Histograms of the differences between the estimated (by RadioUNet trained on DPM simulations) and the assumed ground-truth pathloss measurements in RadioLocSeer Dataset.

the obstructions as cars (100 per map in general, cf. Sec. III-B). The root mean square error between the estimated and the simulated real-surrogate pathloss measurements are found to be 1.66 dB and 4.80 dB, for the Nominal and the Robustness Scenarios, respectively, whereas the mean absolute errors are 0.87 dB and 3.55 dB. The histograms of the differences between the estimated and real-surrogate pathlosses are shown in Fig. 2 for each scenario.

B. Examples

In Fig. 3, we show the results of LocUNet for a typical test map in the previously explained Nominal and Robustness Scenarios. In the first two rows, we show the radio maps for the considered 5 BS locations, which serve as the surrogate ground-truth (From top to bottom): DPM ground-truth (for the pathloss measurements at UE in the Nominal Scenario), IRT with cars ground-truth (for the Robustness Scenario). In the third row, the DPM estimations via RadioUNet are shown. In the last row, we show the heatmaps and the locations estimations (i.e., the CoM of the quasi-heatmap, marked with yellow squares) of the LocUNet for the considered two scenarios, where we partitioned the quasi-heatmap in two partial heatmaps, representing the positive and the negative values (normalized by its max. and min. pixel value, respectively), which we call the *positive heatmap* and the *negative heatmap*. The true location of the UE is marked with green cross.

We note that the CoM of the quasi-heatmap is invariant to the multiplication of the quasi-heatmap by -1. In each scenario, we observe that one of the heatmaps (the positive ones for the first scenario, and the negative one for the second scenario) serves as a *belief map* about the location of the UE, while the other one as a *disbelief map*. We see that the non-zero pixels (there are more of them in the Nominal Scenario) of the disbelief maps are concentrated about the pixels of the buildings and are distanced from the neighborhood of the estimated location.
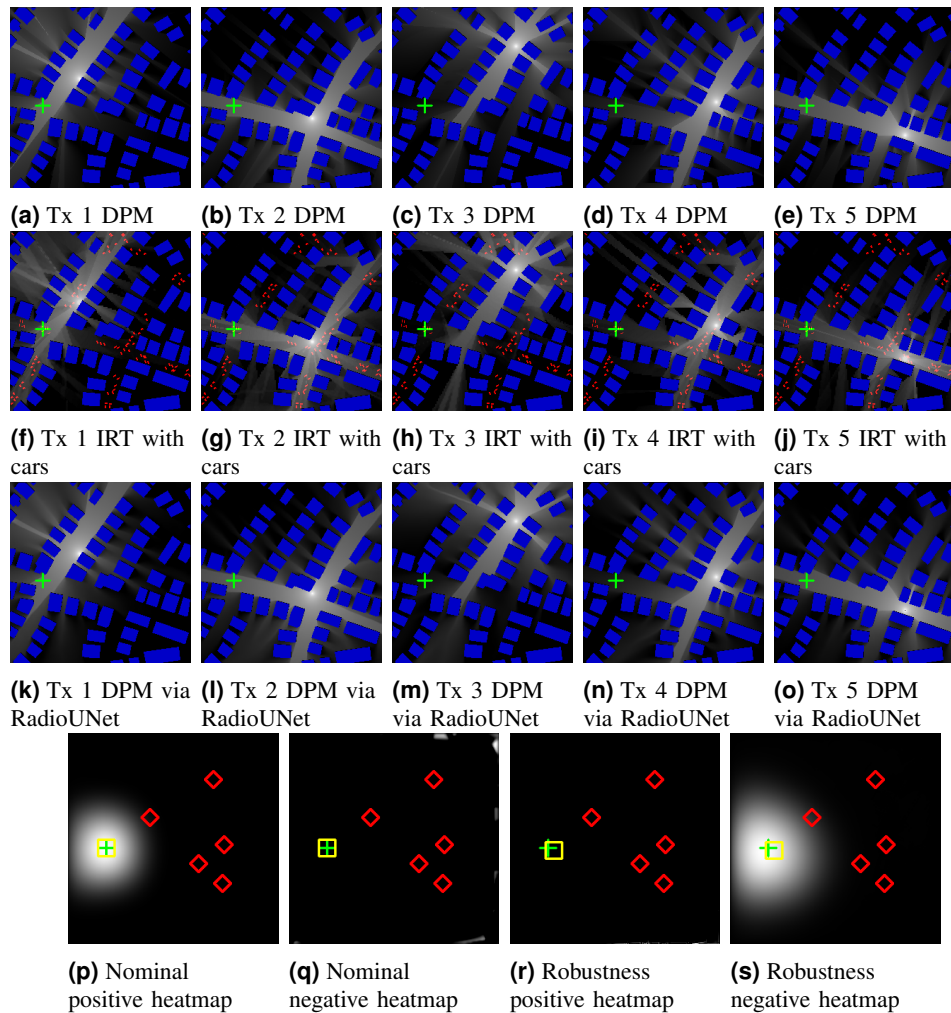
(a) Tx 1 DPM   **(b)** Tx 2 DPM   **(c)** Tx 3 DPM   **(d)** Tx 4 DPM   **(e)** Tx 5 DPM

(f) Tx 1 IRT with cars   **(g)** Tx 2 IRT with cars   **(h)** Tx 3 IRT with cars   **(i)** Tx 4 IRT with cars   **(j)** Tx 5 IRT with cars

(k) Tx 1 DPM via RadioUNet   **(l)** Tx 2 DPM via RadioUNet   **(m)** Tx 3 DPM via RadioUNet   **(n)** Tx 4 DPM via RadioUNet   **(o)** Tx 5 DPM via RadioUNet

(p) Nominal positive heatmap   **(q)** Nominal negative heatmap   **(r)** Robustness positive heatmap   **(s)** Robustness negative heatmap

Fig. 3: A generic localization problem of our dataset, where the pathloss at the UE from each BS is moderate. **First row:** DPM true radio maps. **Second row:** IRT with cars true radio maps. Cars are shown in red. **Third row:** DPM estimated maps by RadioUNet. **Fourth row:** LocUNet results for the two scenarios. The normalized positive (by max. pixel value) and the negative (by min. pixel value) heatmaps of LocUNet before CoM layer are shown in gray-level. Buildings are blue, Tx locations are marked with red diamonds. Estimated and true locations are marked with yellow square and green cross, respectively.

C. Comparison with Pathloss Fingerprint-Based Methods

In Table III, we present the MAE accuracy of LocUNet in the scenarios described in Subsection IV-A, and compare its performance with the competing fingerprint-based methods. The data is provided by the RadioLocSeer Dataset III-B1. LocUNet is trained using the whole dataset with 5 anchor BSs, where it is ensured that the UEs can decode the beacons of at least one BS (See Table IV for the number of BSs in coverage in the test set). Here, we use the radio map estimations of all 5 BS, even when the UE is not in the coverage of a BS (i.e., UE is located at a zero pixel of the true radio map of a BS). In order to study the effect of number of BSs in reception during the test phase, we also test our method on the test set where the number of

TABLE III: Comparison with fingerprint-based methods. Average Euclidean distance (MAE) for 5 and 3 number of BSs, and approximate run-times of the evaluated methods.

Algorithm	5Tx	3Tx	ms
Nominal Scenario			
kNN [61] (k=16, 40)	6.95	17.27	~ 20
kNN [61] Cond.	6.93	16.56	~ 20
Adaptive kNN [62] (avg. k=2.50, 2.40)	7.49	16.18	~ 20
LocUNet	**4.80**	**10.70**	~ 5
Robustness Scenario			
kNN [61] (k=200,1500)	27.38	38.54	~ 20
kNN [61] Cond.	24.44	32.83	~ 20
Adaptive kNN [62] (avg. k=8.51, 5.08)	29.51	45.31	~ 20
LocUNet	**13.14**	**19.06**	~ 5

BSs in coverage is not more than 3.

We compare with two fingerprint-based methods, namely, *k-nearest neighbors (kNN)* method [61] and an adaptive kNN variant [62]. We determined the k parameters of the kNN algorithm by coarse grid-search for both of the 3 and 5 BS settings, as the one which yields the best performance in the training set. We also considered determining individual k parameters for each number of BSs in coverage (See the following Subsection IV-C1 for more) in the same way, and report the overall performance of this approach under the name *kNN Cond.* In the kNN methods, we have not excluded the pixels of buildings from consideration when finding the k neighbors (when a BS is not in coverage), and doing otherwise did not yield improvements.

We observe that LocUNet provides the best accuracy among the fingerprint based methods in all the scenarios, and LocUNet is especially good at dealing with the inaccuracies of the radio map estimations in the realistic setting, as witnessed in the results of the Robustness Scenario (Sec. IV-A). Thus, the results in this scenario are a more meaningful indicator for the performances of the compared algorithms. LocUNet outperforms the best performing competitor algorithm (kNN) by about 11m and 14m in MAE in the cases of 5 or 3 BSs per map.

Remark 3: *We attribute the success of LocUNet in accuracy and its robustness to the inaccuracies in radio map estimations to its convolutional nature, which takes fully into account the neighborhood/spatial relations in the radio map estimations, whereas the kNN method makes use of the spatial information only when the CoM of the k-nearest-neighbors are found in the final step. The kNN method can hence be seen as a convolutional method with convolution kernels of 1x1 pixels. Note that both kNN and our method perform CoM as the last step and the found k-nearest-neighbors can be interpreted as a heatmap. However, the k-nearest-neighbors are solely determined by assigning each pixel a distance in the signal space of the RSS values*

TABLE IV: Conditional performances of kNN and LocUNet.

No BS above NF:	1	2	3	4	5	Overall
kNN (k=200)	69.81	48.89	34.62	26.30	19.62	27.38
kNN Cond.	**46.25**	**35.56**	**29.88**	**24.28**	19.62	**24.44**
kNN Cond. w/ below NF	48.9	42.1	34.22	26.45	19.62	26.31
LocUNet	**30.51**	**21.89**	**16.31**	**13.28**	9.62	**13.14**
LocUNet Cond.	56	28.37	20.1	15.02	**9.06**	15.04
LocUNet Cond. w/ below NF	32.14	39.03	18.45	13.93	**9.06**	14.86
LocUNet Cond. 5	33.57	23.55	17.28	13.5	**9.06**	13.3
no instances in test set	3284	12478	26515	35167	72556	150000
%	2.19	8.32	17.68	23.44	48.37	100

under a metric (we took the usual "Euclidean", the sum of the squared residuals, and our experiments with the "Manhattan" metric did not yield improvements), which disregards any spatial relations of the pixels. LocUNet also compares the RSS measurements (represented via the constant image channel p_j) to the RSS values at each of the pixels (via the input radio map). However, LocUNet also utilizes the spatial information by propagating the RSS information from the spatial neighbors of each point via the convolutions.

1) Conditional Performance of the Compared Methods: Previously, we have evaluated LocUNet and the kNN algorithms for 5 BS deployments or for subsets of 3 BS deployments of them. However, not all BSs are necessarily in coverage (We only ensured each UE is at least in the coverage of one BS, cf. Sec. III-B), as happens often in the real life.

In Table IV we demonstrate the performance of LocUNet and kNN in the Robustness Scenario, conditioned on the number of BSs detectable at the UE. We compare the performance of training (for kNN, setting the k parameter) on the whole training set with that of training separately for different BS numbers. For the conditional training, we consider two variants, one which only uses the radio map estimation of a BS when the UE is in the coverage of the BS, and one which uses all the radio map estimations, even when a link between the UE and a BS is not strong enough for detection. We observe that such non-detectable signal information can be beneficial for LocUNet, while kNN demonstrates a better performance without taking such radio maps into consideration. We also observe that LocUNet performs the best when only one network with the whole training dataset is trained, whereas kNN achieves a better performance under conditional training. This can be attributed to the limited size of the dataset for low number of BSs (cf. bottom of the Table IV), causing LocUNet to overfit. In fact, the results with a network trained with only data from 5 BS in reception shows a deterioration with respect to training with the whole data. A detailed study of the effect of number of BSs in reception, dataset sizes, or neural network sizes could be interesting but is out of the scope of the current work.

TABLE V: Neural networks of the Nominal and Robustness Scenarios evaluated under out-of-distribution pathloss measurements. Comparison with kNN where k parameter is fixed for the scenario of the interest. The accuracies under in-distribution measurements are shown in red.

	DPM	DPM w/ cars	IRT	IRT w/ cars
LocUNet Nominal	4.80	**8.80**	20.07	25.64
kNN Nominal (k=16)	6.95	11.29	24.38	28.56
LocUNet Robustness	12.94	12.52	**11.93**	13.14
kNN Robustness (k=200)	9.22	12.46	23.41	27.38

D. Robustness to Out-of-Distribution Pathloss Measurements

In this Section we show the performance of LocUNet and the kNN method for the above defined two scenarios under out-of-distribution (OOD) pathloss measurements, i.e., we test with measurements stemming from different distributions than the one during the training.

To this end, we take the trained LocUNet networks under these two scenarios, and kNN method with k values found via grid search in the training set for each scenarios. We then test both methods with pathloss measurements obtained from different simulations scenarios. In addition to the above mentioned two ground-truth scenarios (DPM and IRT with cars), we consider two more scenarios. **1**) DPM with cars, where the radio map is simulated with the DPM under the presence of additional (unknown to LocUNet) obstructions (cars), **2**) IRT simulations as in the Robustness Scenario, but without the presence of unknown cars. The data for these scenarios are also available in the RadioLocSeer Dataset. In Table V we observe that the network trained in the Nominal scenario suffers from substantial decrease of performance when the pathloss measurements stem from other scenarios, whereas the accuracy of the network trained in Robustness Scenario is not much affected, while being outperformed by the network trained in the Nominal Scenario when the measurements stem from DPM simulations, which demonstrates the importance of having high accuracy radio map estimations, when possible. We also notice that the network trained in Robustness Scenario performs even better when there are no unknown cars in the map. On the other hand, we observe that the performance of the kNN method is mainly dependent on the closeness of the real (surrogate) and the estimated maps, and different k (nearest neighbors) values do not lead to remarkable improvements in performance, especially when the mismatch between the estimated and true radio map is high.

E. Robustness to Location Errors

1) Location Errors in the Training Dataset: A crucial requirement of a localization method should be robustness to inaccuracies of the reported locations of the samples in the training dataset, as the very localization problem in dense urban setting, which the current paper is dealing with, is also present while collecting the labeled (here, the true locations) data. To
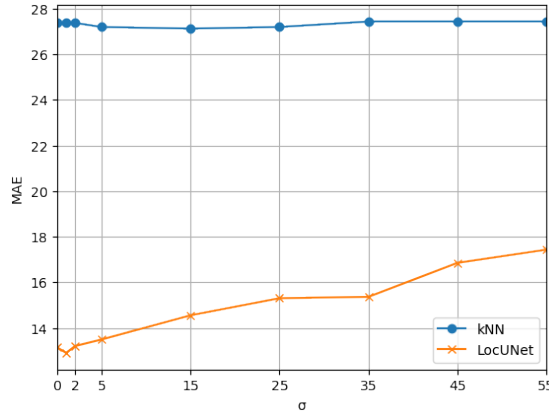
Fig. 4: Robustness to the inaccuracies in the reported positions of the training dataset samples

TABLE VI: Test accuracies under training with contaminated position labels. The numbers of training epochs of LocUNet that give rise to the best validation losses are reported in the last row.

Algorithm	$\sigma = 0$	$\sigma = 1$	$\sigma = 2$	$\sigma = 5$	$\sigma = 15$	$\sigma = 25$	$\sigma = 35$	$\sigma = 45$	$\sigma = 55$
kNN [61]	27.38	27.38	27.38	27.20	27.13	27.20	27.44	27.44	27.44
LocUNet	**13.14**	**12.90**	**13.20**	**13.49**	**14.55**	**15.30**	**15.36**	**16.85**	**17.43**
best # of epochs	20	17	30	14	8	8	8	9	6

assess the robustness of LocUNet to such inaccuracies, we consider additive Gaussian noise on each location coordinate of the training data. We report the test results for varying standard deviation of the noise in Fig. 4 and Table VI. We observe that even to a significantly high level of additive noise ($\sigma = 55$) to the positions in the training data, both LocUNet and kNN [61] show extremely high robustness. Another observation we made is a trend of decrease in the number of training epochs which yields the best validation loss with increasing σ, i.e., overfitting occurs earlier as σ increases and the generalization is more severely impaired.

2) Uncertainty of Anchor BS Locations: Throughout the paper, we focused on a setting with known anchor BS locations, which are naturally known to the service providers, and might be also available to the users (required for self-localization at the UEs), through open access sources such as https://www.cellmapper.net. Here, we present a brief numerical study of LocUNet, when the locations of the anchor BSs are *not* exactly known. As before, we model the uncertainty as an additive Gaussian random variable. Since generating a dataset for each realization of such uncertainty is unfeasible, we instead distort the position of the UE for each considered BS, and assume that the pathloss measurements stem from the distorted location (rounded so that it corresponds to a pixel in the radio map), while only the non-contaminated location is known. If the distorted location corresponds to a location in a building, we repeat the process of adding noise to the original location of the UE, until it does not. The results of LocUNet and kNN are reported in Table VII. In the first two rows, we show the accuracy of the algorithms

TABLE VII: Robustness to the inaccuracies in the anchor BS locations. For each transmitter, the reported pathloss stems from a position contaminated by additive Gaussian noise with given standard deviation σ, to emulate the uncertainty in the BS position.

Algorithm	$\sigma = 5$	$\sigma = 10$	$\sigma = 15$	$\sigma = 20$	$\sigma = 25$
kNN [61] Noisy Loc. ZS (k=200)	30.91	36.34	43.67	50.98	57.95
LocUNet Noisy Loc. ZS	**16.42**	**22.23**	**29.01**	**35.29**	**40.96**
kNN [61] Noisy Loc. (k=500,800,6000,10000,12000)	30.11	34.78	38.81	42.54	45.19
LocUNet Noisy Loc.	**16.26**	**20.52**	**25.43**	**29.55**	**34.09**

trained with data with exactly known transmitter locations and evaluate their performances under contaminated transmitter locations (Zero-Shot Scenario, ZS). In the last two rows, the results under training with distorted BS locations are presented, where we observe the beneficial effects of it with respect to the Zero-Shot Scenario on the accuracy of the both methods.

F. Comparison with ToA Ranging-Based Methods

Using the RadioToASeer Dataset (Sec. III-B2), we evaluate the accuracies of the state-of-the-art ToA ranging based algorithms, in our presented realistic urban setting, and compare with our method. The compared methods are based on the measured ToAs of the dominant paths of the beacon signals of BSs as explained in Section III-B2, found in our RadioToASeer Dataset. We assumed the availability of the ToA measurements even if the signal power of the dominant ray of interest falls below the noise floor. This further favors ToA-based schemes. In a separate set of experiments not shown here for space limitation we have verified that if only the rays with RSS above the noise floor are considered (average of 4.72 BS above the noise floor), similar results are obtained with 2-3 meters of degradation in the performances of SDP-based methods, and around 1 meter of improvement in bisection-based methods.

In practice, there are several factors which deteriorate the accuracy of ToA measurements at a UE, e.g., clock offset and skew (synchronization errors), quantization of the measured values, additional thermal noise, or failures in resolving the multipath components. In the numerical experiments, we only took into account the inaccuracy due to the quantization (in LTE, ToAs are quantized to intervals of 9.77 meters), the thermal noise and the clock error and modeled them altogether with a single additional Gaussian noise variable with zero mean and standard deviation σ, as proposed in [63]. As in [63], we took $\sigma = 20$ meters, and also evaluated the methods with $\sigma = 10$m and under no measurement noise, $\sigma = 0$ (cf. Table VIII). For all the cases, all the 5 BSs or a subset of 3 BSs are taken into consideration.

For the robust ToA methods [23]–[25], we pick the NLOS bias b at CDF 0.9 (20 m), as suggested in [26], and also at CDF 0.5 (0.7 m) (cf. Fig. 1).

TABLE VIII: Comparison with ToA ranging-based methods. Absolute Euclidean distance (MAE) accuracies and approximate run-times of compared algorithms under different additional noise and number of BSs in the map.

no BS:	5	3	5	3	5	3	Run-time
Standard dev. of noise:	$\sigma = 0$		$\sigma = 10$		$\sigma = 20$		$\sim$ ms
POCS [20], [21]	37.75	46.15	39.36	47.26	41.27	48.72	15
SDP [22]	**6.81**	**13.95**	15.52	27.81	24.88	41.02	600
Robust SDP 1 [23], $b = 20$	9.98	17.10	17.02	27.59	27.76	40.37	600
Robust SDP 1 [23], $b = 0.7$	7.04	15.38	17.13	28.12	28.42	41.74	600
Robust SDP 2 [24], $b = 20$	12.53	19.63	18.66	28.59	27.81	39.93	600
Robust SDP 2 [24], $b = 0.7$	7.14	15.63	17.10	28.10	28.39	41.72	600
Bisection-based robust method [25], $b = 20$	9.16	15.87	**14.42**	**25.40**	**23.30**	**38.14**	16
Bisection-based robust method [25], $b = 0.7$	9.49	14.95	14.80	27.35	24.09	40.75	16
Max. correntropy criterion method [26]	12.44	18.40	20.54	31.27	31.91	45.68	30
LocUNet Nominal & Robustness Scenarios	4.80	10.70	13.14	19.06	13.14	19.06	5

We report the simulation results in Table VIII. We added the results of LocUNet in the Nominal and Robustness Scenarios, at the bottom of the Table, for comparison. The results of LocUNet under the Nominal and Robustness Scenarios should be compared with the ToA ranging-based methods under no additive measurement noise and under noisy measurements, respectively. We observe that LocUNet outperforms all the compared ToA-based methods in all the settings.

G. Comparison with Baselines in Image Processing

To demonstrate the validity of our method, we perform comparisons with baselines. We relate our method to the localization problems (e.g. human/hand pose estimation) in image processing, which can be generally characterized as a so-called *keypoint* regression problem [64]. We describe in the following four approaches, with increasing level of accuracy, where our method falls into the last category. We report our findings based on the numerical experiments we conducted, and discuss the results.

1) Direct classification: For the localization problem at hand, a classification approach seems to be the most appropriate at first sight. In this case, the task of the DNN is to approximate a one-hot map with the "hot" pixel at the desired location. The classical loss function in this case is the cross-entropy loss between the desired one-hot map and the heatmap output by the DNN (both maps can be interpreted as probability distributions on the set of all possible locations). While this approach may be appropriate for classification of categorical data (e.g., handwriting of letters and numbers), it is inappropriate for localization since the cross-entropy loss does not capture the proximity of the estimated position in the Euclidean distance sense. In fact, our regression approach based on MAE or MSE is much more meaningful in this sense. Our implementation of the classification approach did not converge, as expected.
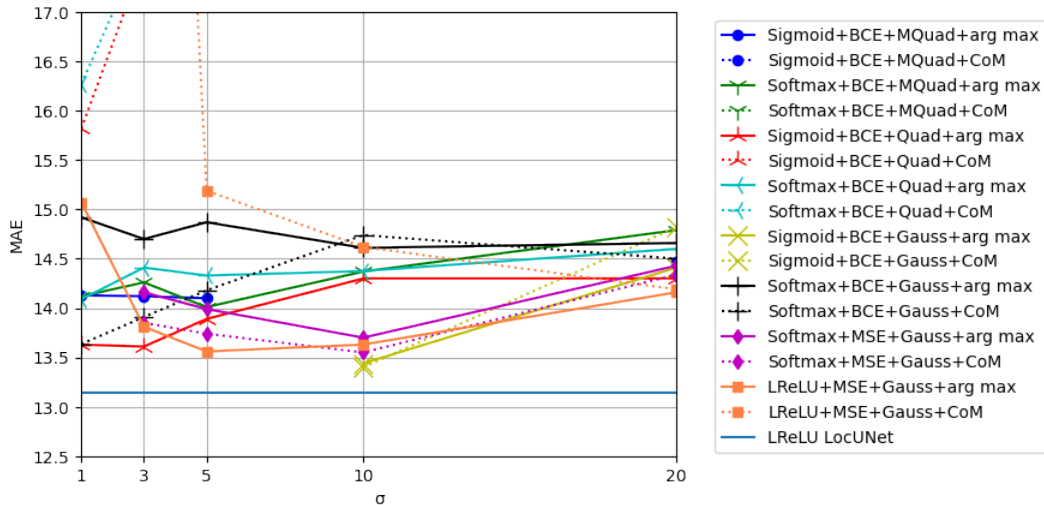
Fig. 5: Heatmap regression with different activations, heatmap loss functions, heatmap shapes vs LocUNet. Abbreviations: **(Gauss)**: Gaussian, **(MQuad)**: Inverse Multiquadric, **(Quad)**: Inverse Quadratic, **(BCE)**: Pixelwise binary cross-entropy.

TABLE IX: 5 best performing choices for heatmap regression vs LocUNet

Method	σ	MAE
Sigmoid+BCE+Gaussian+CoM	10	**13.40**
Sigmoid+BCE+Gaussian+argmax	10	13.44
Softmax+MSE+Gaussian+CoM	10	13.55
LeakyReLU+MSE+Gaussian+argmax	5	13.56
Sigmoid+BCE+Inverse quadratic+argmax	3	13.61
LocUNet	-	**13.14**

2) Direct regression: Another reasonable approach could be the direct regression of the pixel coordinates under a distance based loss, where instead of the CoM layer, a fully connected layer would be used as the last layer. This flattening at the last layer impairs the spatial generalizability of the network. In particular, the translation equivariance, which is crucial for the localization problem, is lost. Our experiments with this approach did not converge either. Similar observations were made in some previous works, e.g. in [65].

3) Heatmap regression: To preserve the spatial information, the flattening of the layers has to be avoided. Heatmap regression based methods use fully convolutional networks to maintain the spatial information. In such methods, the networks are trained to match a (ground truth) heatmap, which are usually generated as 2D Gaussian distributions (standard deviation σ being a hyperparameter to be tuned) centered on the ground truth pixels. The values of the heatmap for the *Gaussian* case is obtained by $e^{-\frac{r^2}{2\sigma^2}}$, where r is the distance to the ground truth pixel from the pixel of interest. Additionally, we used two other radial basis functions for heatmap generation, namely, *Inverse Multiquadric*: $1/\sqrt{1 + r^2/(2\sigma^2)}$ and *Inverse Quadratic*: $1/(1 + r^2/(2\sigma^2))$.

After the neural network is trained, the position is inferred by taking either the pixel with the

TABLE X: Performance of LocUNet in MAE and RMSE with different activation functions at the last layer before CoM, with different loss functions for training, and with different additional input features. n.c.: no convergence.

Training Loss:			MAE				MSE			
Inputs	# parameters	Test Metric	Leaky ReLU	ReLU	Softmax	Sigmoid	Leaky ReLU	ReLU	Softmax	Sigmoid
wC wT	22,332,381	MAE	**13.14**	n.c.	13.25	13.15	**14.45**	15.41	15.04	15.04
		RMSE	**21.78**	n.c.	23.53	23.09	**21.93**	22.81	23.72	22.35
w/oC wT	22,331,676	MAE	13.32	n.c.	**12.88**	13.47	**14.32**	n.c.	14.54	14.66
		RMSE	23.66	n.c.	**22.60**	23.79	21.98	n.c.	22.53	**21.88**
wC w/oT	22,328,856	MAE	**12.86**	n.c.	13.00	13.27	14.75	n.c.	**14.45**	14.91
		RMSE	**22.26**	n.c.	23.27	23.37	**21.95**	n.c.	22.07	21.99
w/oC w/oT	22,328,151	MAE	**13.05**	n.c.	13.08	13.36	16.24	n.c.	**14.40**	14.69
		RMSE	**22.37**	n.c.	22.55	23.12	23.44	n.c.	**21.7**	22.21

largest value in the heatmap (the *mode*) or the center of mass (the *mean*), as the heatmap can be interpreted as an a posteriori probability distribution of the UE position, and the mode and mean correspond to *maximum a posteriori* and *posterior mean* estimation, respectively. Hence, these methods are trained against a loss which is different from the actual metric of interest - the Euclidean distance. Our experiments with this approach with the mentioned inference functions (i.e., $\arg\max$ and CoM), and with different activation functions, heatmap loss functions, and heatmap shapes have demonstrated worse yet close performances to LocUNet for specific values of σ, which are shown in Fig. 5 and summarized in Table IX.

4) Spatial regression: This approach, which LocUNet is special case of, is based on relying on the heatmaps only as a latent representation, and defining the loss directly on the location coordinates. The inference of the coordinates are achieved by a differentiable function (mostly by CoM as in our paper), which allows end-to-end training. Equipped with the spatial generalization capability thanks to the convolutional network, together with the direct regression of the coordinates, such methods outperform the heatmap regression approach [64], [66]–[69].

H. Ablation Study

In this section, we discuss the effectiveness of our neural network design by comparing with other reasonable approaches. In the following, we present the impact of the additional input features, the choice of the last activation function of the network, and the choice of the loss function during the training. Resulting accuracies in MAE and root-mean-square error (RMSE) are shown in Table X.

1) Impact of the Additional Input Features: Our observation is that the impact of the additional input features (cf. Sec. III-A) on the performance is not significant. Hence, for the sake of simplicity of the exposition, throughout the paper we have taken all the additional features as inputs to the LocUNet, when comparisons with other methods are carried out. In Table X we

TABLE XI: Improvements (in percentage) observed in Table X in MAE, RMSE metrics by training with MAE and MSE losses, with respect to training with MSE and MAE losses, respectively.

Inputs	Improvement	Leaky ReLU	Softmax	Sigmoid
wC wT	MAE	**9.07**%	**11.9**%	**12.58**%
	RMSE	-0.69%	-0.81%	3.21%
w/oC wT	MAE	6.98%	11.42%	8.10%
	RMSE	**7.10**%	0.31%	8.03%
wC w/oT	MAE	**12.80**%	10.03%	11.00%
	RMSE	1.39%	5.15%	5.91%
w/oC w/oT	MAE	**19.65**%	9.17%	9.03%
	RMSE	-4.78%	3.7%	3.94%

show the results of the numerical experiments with (**w**) and without (**w/o**) city (**C**) and Tx (**T**) maps as additional input features.

2) Impact of the Activation Function Before the CoM Layer: All of the above mentioned previous works in localization forced the heatmaps by the activation functions to become non-negative, or normalized them to represent a probability distribution (e.g. by Softmax as the activation function). Interestingly, our experiments with several activation functions (ReLU, Softmax, Sigmoid and Leaky ReLU), have shown, that allowing negative values in the heatmap (by using Leaky ReLU as the activation function), performed generally the best among all the activation functions (cf. Table X). Here, this quasi-heatmap can be interpreted as a quasi-probability and the CoM as then the mean under a quasi-probability distribution, which is the MMSE estimator (See also the Examples in Sec. IV-B and the *belief/disbelief map* interpretation therein). A study of this interesting observation is, however, beyond the scope of this paper.

3) Impact of the Loss Function in the Training: Based on the results in the Table X, we compare the influence of the training loss function on the evaluation metrics of MAE and RMSE in Table XI, where we show the improvement (in percentage) in MAE (RMSE) when trained with MAE (MSE) loss, with respect to when trained with MSE (MAE) loss. In mathematical terms, we calculate $\text{Imp}_{\text{RMSE}} = 100 \times \frac{\text{RMSE}_{\text{MAE}} - \text{RMSE}_{\text{MSE}}}{\text{RMSE}_{\text{MAE}}}$, $\text{Imp}_{\text{MAE}} = 100 \times \frac{\text{MAE}_{\text{MSE}} - \text{MAE}_{\text{MAE}}}{\text{MAE}_{\text{MSE}}}$, where the terms in fractions denote the performance in terms of $\text{TestMetric}_{\text{TrainingLoss}}$.

We observe that in most of the cases training with MAE loss yielded more gains in MAE in comparison to the performance gains achieved in RMSE when trained with MSE.

V. CONCLUSIONS

In this paper we presented a deep learning method to localize a UE based on the measured signal strengths from a set of BSs. Our method, based on the estimates of the true radio maps of each BS, is tailored to work in realistic propagation environments in urban settings. The proposed approach does not necessitate currently uncommon hardware neither at UEs nor at

BSs, and does not require an extensive measurement campaign, and thus can be easily integrated in the localization practices that are already in use. Our method outperforms the state-of-the-art localization algorithms in realistic urban scenarios and enjoys high robustness to inaccuracies in the radio map estimates, and also to inaccuracies in the positions labels of the training set. We also presented two new datasets, which can be used to evaluate the performances of pathloss fingerprint and ToA ranging-based algorithms in realistic urban environment scenarios.

VI. Acknowledgment

References

[1] Ç. Yapar, R. Levie, G. Kutyniok, and G. Caire, "LocUNet: Fast urban positioning using radio maps and deep learning," in *ICASSP 2022 - 2022 IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2022, pp. 4063–4067.

[2] J. Warrior, E. McHenry, and K. McGee, "They know where you are [location detection]," *IEEE Spectrum*, vol. 40, no. 7, pp. 20–25, 2003.

[3] G. Bresson, Z. Alsayed, L. Yu, and S. Glaser, "Simultaneous localization and mapping: A survey of current trends in autonomous driving," *IEEE Transactions on Intelligent Vehicles*, vol. 2, no. 3, pp. 194–220, 2017.

[4] S. Severi, H. Wymeersch, J. Härri, M. Ulmschneider, B. Denis, and M. Bartels, "Beyond GNSS: Highly accurate localization for cooperative-intelligent transport systems," in *IEEE Wireless Comm. and Networking Conf. (WCNC)*, 2018, pp. 1–6.

[5] E. Pournaras, "Proof of witness presence: Blockchain consensus for augmented democracy in smart cities," *Journal of Parallel and Distributed Computing*, vol. 145, pp. 160–175, 2020.

[6] R. Di Taranto, S. Muppirisetty, R. Raulefs, D. Slock, T. Svensson, and H. Wymeersch, "Location-aware communications for 5G networks: How location information can improve scalability, latency, and robustness of 5G," *IEEE Signal Processing Magazine*, vol. 31, no. 6, pp. 102–112, 2014.

[7] A. H. Sayed, A. Tarighat, and N. Khajehnouri, "Network-based wireless location: Challenges faced in developing techniques for accurate wireless location information," *IEEE Signal Processing Magazine*, vol. 22, no. 4, pp. 24–40, 2005.

[8] A. Amini, R. M. Vaghefi, J. M. de la Garza, and R. M. Buehrer, "Improving GPS-based vehicle positioning for intelligent transportation systems," in *2014 IEEE Intelligent Vehicles Symposium Proceedings*, 2014, pp. 1023–1029.

[9] I. Guvenc and C. Chong, "A survey on TOA based wireless localization and NLOS mitigation techniques," *IEEE Communications Surveys and Tutorials*, vol. 11, no. 3, pp. 107–124, 2009.

[10] R. M. Vaghefi and R. M. Buehrer, "Cooperative sensor localization with NLOS mitigation using semidefinite programming," in *2012 9th Workshop on Positioning, Navigation and Communication*, 2012, pp. 13–18.

[11] H. Liu, H. Darabi, P. Banerjee, and J. Liu, "Survey of wireless indoor positioning techniques and systems," *IEEE Trans. on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 37, no. 6, pp. 1067–1080, 2007.

[12] R. M. Vaghefi, M. R. Gholami, and E. G. Ström, "Bearing-only target localization with uncertainties in observer position," in *2010 IEEE 21st Int. Symp. on Personal, Indoor and Mobile Radio Comm. Workshops*, 2010, pp. 238–242.

[13] R. M. Vaghefi and R. M. Buehrer, "Received signal strength-based sensor localization in spatially correlated shadowing," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, 2013, pp. 4076–4080.

[14] A. Zanella, "Best practice in RSS measurements and ranging," *IEEE Communications Surveys and Tutorials*, vol. 18, no. 4, pp. 2662–2686, 2016.

[15] N. Patwari, A. Hero, M. Perkins, N. Correal, and R. O'Dea, "Relative location estimation in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2137–2148, 2003.

[16] T. S. Rappaport, *Wireless communications: Principles and practice*, 2nd ed.   Prentice Hall, 2002.

[17] H. Laitinen, J. Lahteenmaki, and T. Nordstrom, "Database correlation method for GSM location," in *IEEE VTS 53rd Vehicular Technology Conference, Spring 2001. Proceedings (Cat. No.01CH37202)*, vol. 4, 2001, pp. 2504–2508 vol.4.

[18] G. Giorgetti, S. Gupta, and G. Manes, "Localization using signal strength: To range or not to range?" in *Proceedings of the First ACM International Workshop on Mobile Entity Localization and Tracking in GPS-Less Environments*, ser. MELT '08.   New York, NY, USA: Association for Computing Machinery, 2008, p. 91–96.

[19] A. Karttunen, A. F. Molisch, S. Hur, J. Park, and C. J. Zhang, "Spatially consistent street-by-street path loss model for 28-GHz channels in micro cell urban environments," *IEEE Trans. Wireless Comm.*, vol. 16, no. 11, pp. 7538–7550, 2017.

[20] M. R. Gholami, H. Wymeersch, E. G. Ström, and M. Rydström, "Wireless network positioning as a convex feasibility problem," *EURASIP Journal on Wireless Communications and Networking*, vol. 2011, no. 1, p. 161, 2011.

[21] A. O. Hero and D. Blatt, "Sensor network source localization via projection onto convex sets (POCS)," in *Proceedings. (ICASSP '05). IEEE Int. Conf. on Acoustics, Speech, and Signal Proc., 2005.*, vol. 3, 2005, pp. iii/689–iii/692 Vol. 3.

[22] R. M. Vaghefi, J. Schloemann, and R. M. Buehrer, "NLOS mitigation in TOA-based localization using semidefinite programming," in *2013 10th Workshop on Positioning, Navigation and Communication (WPNC)*, 2013, pp. 1–6.

[23] G. Wang, H. Chen, Y. Li, and N. Ansari, "NLOS error mitigation for TOA-based localization via convex relaxation," *IEEE Transactions on Wireless Communications*, vol. 13, no. 8, pp. 4119–4131, 2014.

[24] H. Chen, G. Wang, and N. Ansari, "Improved robust TOA-based localization via NLOS balancing parameter estimation," *IEEE Transactions on Vehicular Technology*, vol. 68, no. 6, pp. 6177–6181, 2019.

[25] S. Tomic, M. Beko, R. Dinis, and P. Montezuma, "A robust bisection-based estimator for TOA-based target localization in NLOS environments," *IEEE Communications Letters*, vol. 21, no. 11, pp. 2488–2491, 2017.

[26] W. Xiong, C. Schindelhauer, H. Cheung So, and Z. Wang, "Maximum correntropy criterion for robust TOA-based localization in NLOS environments," *arXiv e-prints*, p. arXiv:2009.06032, Sep. 2020.

[27] Q. D. Vo and P. De, "A survey of fingerprint-based outdoor localization," *IEEE Comm. Surveys Tutorials*, vol. 18, no. 1, pp. 491–506, 2016.

[28] E. Laitinen, J. Talvitie, E.-S. Lohan, and M. Renfors, "Comparison of positioning accuracy of grid and path loss-based mobile positioning methods using received signal strengths," in *Proc. Signal Process. and Appl. Math. for Electron. and Commun. (SPAMEC)*, Cluj-Napoca, Romania, 2011.

[29] W. Y. Al-Rashdan and A. Tahat, "A comparative performance evaluation of machine learning algorithms for fingerprinting based localization in DM-MIMO wireless systems relying on big data techniques," *IEEE Access*, vol. 8, 2020.

[30] D. Burghal, A. T. Ravi, V. Rao, A. A. Alghafis, and A. F. Molisch, "A Comprehensive Survey of Machine Learning Based Localization with Wireless Signals," *arXiv e-prints*, p. arXiv:2012.11171, Dec. 2020.

[31] K. Rizk, J. F. Wagen, and F. Gardiol, "Two-dimensional ray-tracing modeling for propagation prediction in microcellular environments," *IEEE Trans. Vehic. Tech.*, vol. 46, no. 2, pp. 508–518, May 1997.

[32] Z. Yun and M. F. Iskander, "Ray tracing for radio propagation modeling: Principles and applications," *IEEE Access*, vol. 3, pp. 1089–1100, 2015.

[33] R. Hoppe, G. Wölfle, and U. Jakobus, "Wave propagation and radio network planning software WinProp added to the electromagnetic solver package FEKO," in *Proc. Int. Appl. Computational Electromagnetics Society Symp. - Italy (ACES)*, Florence, Italy, March 2017, pp. 1–2.

[34] R. Wahl, G. Wölfle, P. Wertz, P. Wildbolz, and F. Landstorfer, "Dominant path prediction model for urban scenarios," in *14th IST Mobile and Wireless Communications Summit*, 2005.

[35] I. Stepanov and K. Rothermel, "On the impact of a more realistic physical layer on MANET simulations results," *Ad Hoc Networks*, vol. 6, no. 1, pp. 61–78, 2008.

[36] A. Del Corte-Valiente, J. M. Gómez-Pulido, O. Gutiérrez-Blanco, and J. L. Castillo-Sequera, "Localization approach based on ray-tracing simulations and fingerprinting techniques for indoor–outdoor scenarios," *Energies*, vol. 12, no. 15, 2019. [Online]. Available: https://www.mdpi.com/1996-1073/12/15/2943

[37] M. Raspopoulos, C. Laoudias, L. Kanaris, A. Kokkinis, C. G. Panayiotou, and S. Stavrou, "3D ray tracing for device-independent fingerprint-based positioning in WLANs," in *2012 9th Workshop on Positioning, Navigation and Communication*, 2012, pp. 109–113.

[38] P. S. Maher and R. A. Malaney, "A novel fingerprint location method using ray-tracing," in *GLOBECOM 2009 - 2009 IEEE Global Telecommunications Conference*, 2009, pp. 1–5.

[39] K. El-Kafrawy, M. Youssef, A. El-Keyi, and A. Naguib, "Propagation modeling for accurate indoor WLAN RSS-based localization," in *2010 IEEE 72nd Vehicular Technology Conference - Fall*, 2010, pp. 1–5.

[40] F. Firdaus, N. A. Ahmad, and S. Sahibuddin, "Accurate indoor-positioning model based on people effect and ray-tracing propagation," *Sensors*, vol. 19, no. 24, 2019.

[41] M. M. Butt, A. Rao, and D. Yoon, "RF fingerprinting and deep learning assisted UE positioning in 5G," *arXiv preprint arXiv:2001.00977*, 2020.

[42] G. Wölfle, R. Hoppe, D. Zimmermann, and F. M. Landstorfer, "Enhanced localization technique within urban and indoor environments based on accurate and fast propagation models," in *European Wireless*, February 2002, pp. 3799–3803.

[43] M. N. de Sousa, R. L. Cardoso, H. S. Melo, J. W. C. Parente, and R. S. Thomä, "Machine learning and multipath fingerprints for emitter localization in urban scenario," in *Developments and Advances in Defense and Security*, Á. Rocha and R. P. Pereira, Eds.   Singapore: Springer Singapore, 2020, pp. 217–230.

[44] R. Levie, C. Yapar, G. Kutyniok, and G. Caire, "RadioUNet: Fast radio map estimation with convolutional neural networks," *IEEE Transactions on Wireless Communications*, vol. 20, no. 6, pp. 4001–4015, 2021.

[45] D. Li, Y. Lei, H. Zhang, and X. Li, "Outdoor positioning based on deep learning and wireless network fingerprint technology," *International Journal of RF and Microwave Computer-Aided Engineering*, vol. 30, no. 12, p. e22444, 2020.

[46] K. Elawaad, M. Ezzeldin, and M. Torki, "DeepCReg: Improving cellular-based outdoor localization using CNN-based regressors," in *2020 IEEE Wireless Communications and Networking Conference (WCNC)*, 2020, pp. 1–6.

[47] P. Garau Burguera, "Logical radio maps for user localization in a real outdoor radio environment," Master's thesis, Aalto University. School of Electrical Engineering, 2020. [Online]. Available: http://urn.fi/URN:NBN:fi:aalto-2020122056430

[48] F. Mahdavi, H. Zayyani, and R. Rajabi, "RSS localiation using an optimized fusion of two deep neural networks," *IEEE Sensors Letters*, vol. 5, no. 12, pp. 1–4, 2021.

[49] J. F. Li, Y.-X. Ye, A.-N. Lu, M.-Y. You, K. Huang, and B. Jiang, "Wireless localization based on deep learning: State of art and challenges," *Mathematical Problems in Engineering*, vol. 2020, p. 5214920, 2020.

[50] A. Zaidi, F. Athley, J. Medbo, U. Gustavsson, G. Durisi, and X. Chen, *5G Physical Layer: principles, models and technology components*.   Academic Press, 2018.

[51] R. Hoppe, G. Wölfle, and F. Landstorfer, "Fast 3-D ray tracing for the planning of microcells by intelligent preprocessing of the data base," in *3rd European Personal and Mobile Communications Conference (EPMCC)*, 1999.

[52] G. Durgin, N. Patwari, and T. S. Rappaport, "An advanced 3D ray launching method for wireless propagation prediction," in *1997 IEEE 47th Vehicular Technology Conference. Technology in Motion*, vol. 2, 1997, pp. 785–789 vol.2.

[53] O. Ronneberger, O. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015*, N. Navab, J. Hornegger, W. M. Wells, and A. F. Frangi, Eds. Cham: Springer International Publishing, 2015, pp. 234–241.

[54] R. Levie, C. Yapar, G. Kutyniok, and G. Caire, "Pathloss prediction using deep learning with applications to cellular optimization and efficient D2D link scheduling," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 8678–8682.

[55] C. Yapar, R. Levie, G. Kutyniok, and G. Caire, "Dataset of pathloss and ToA radio maps with localization application," 2022. [Online]. Available: https://dx.doi.org/10.21227/0gtx-6v30

[56] ——, "Dataset of pathloss and ToA radio maps with localization application," *arXiv preprint arXiv:2212.11777*, 2022.

[57] G. Durgin, T. S. Rappaport, and Hao Xu, "Measurements and models for radio path loss and penetration loss in and around homes and trees at 5.85 GHz," *IEEE Transactions on Communications*, vol. 46, no. 11, pp. 1484–1496, 1998.

[58] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Diego, CA, USA, May 2015.

[59] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai, and S. Chintala, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems 32*. Curran Associates, Inc., 2019, pp. 8024–8035.

[60] OpenStreetMap contributors, "Planet dump retrieved from https://planet.osm.org ," https://www.openstreetmap.org, 2017.

[61] P. Bahl and V. N. Padmanabhan, "RADAR: An in-building RF-based user location and tracking system," in *Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies (Cat. No.00CH37064)*, vol. 2, 2000, pp. 775–784 vol.2.

[62] J. Oh and J. Kim, "Adaptive k-nearest neighbour algorithm for WiFi fingerprint positioning," *ICT Express*, vol. 4, no. 2, pp. 91 – 94, 2018, sI on Artificial Intelligence and Machine Learning.

[63] F. Pérez-Cruz, P. M. Olmos, M. M. Zhang, and H. Huang, "Probabilistic time of arrival localization," *IEEE Signal Processing Letters*, vol. 26, no. 11, pp. 1683–1687, 2019.

[64] C. Tensmeyer and T. Martinez, "Robust keypoint detection," in *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, vol. 5, 2019, pp. 1–7.

[65] B. Brousseau, J. Rose, and M. Eizenman, "Hybrid eye-tracking on a smartphone with CNN feature extraction and an infrared 3D model," *Sensors*, vol. 20, no. 2, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/2/543

[66] A. Nibali, Z. He, S. Morgan, and L. Prendergast, "Numerical coordinate regression with convolutional neural networks," *arXiv preprint arXiv:1801.07372*, 2018.

[67] J. P. Robinson, Y. Li, N. Zhang, Y. Fu, and S. Tulyakov, "Laplace landmark localization," in *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2019, pp. 10 102–10 111.

[68] X. Sun, B. Xiao, F. Wei, S. Liang, and Y. Wei, "Integral human pose regression," in *Proc. of the European Conf. on Computer Vision (ECCV)*, Sep. 2018.

[69] D. C. Luvizon, H. Tabia, and D. Picard, "Human pose regression by combining indirect part detection and contextual information," *Computers & Graphics*, vol. 85, pp. 15–22, 2019.