

RadioUNet: Fast Radio Map Estimation With Convolutional Neural Networks

Ron Levie^{id}, Çağkan Yapar^{id}, *Graduate Student Member, IEEE*, Gitta Kutyniok, *Senior Member, IEEE*,
and Giuseppe Caire^{id}, *Fellow, IEEE*

Abstract—In this paper we propose a highly efficient and very accurate deep learning method for estimating the propagation pathloss from a point x (transmitter location) to any point y on a planar domain. For applications such as user-cell site association and device-to-device link scheduling, an accurate knowledge of the pathloss function for all pairs of transmitter-receiver locations is very important. Commonly used statistical models approximate the pathloss as a decaying function of the distance between transmitter and receiver. However, in realistic propagation environments characterized by the presence of buildings, street canyons, and objects at different heights, such radial-symmetric functions yield very misleading results. In this paper we show that properly designed and trained deep neural networks are able to learn how to estimate the pathloss function, given an urban environment, in a very accurate and computationally efficient manner. Our proposed method, termed RadioUNet, learns from a physical simulation dataset, and generates pathloss estimations that are very close to the simulations, but are much faster to compute for real-time applications. Moreover, we propose methods for transferring what was learned from simulations to real-life. Numerical results show that our method significantly outperforms previously proposed methods.

Index Terms—Convolutional neural networks, signal strength prediction, radio maps.

I. INTRODUCTION

IN wireless communications, the pathloss is a quantity that measures the loss of signal strength (reduction in power, or attenuation) between a transmitter (Tx) and receiver (Rx) due to large scale effects. The signal power attenuation may be caused by different factors, such as free-space propagation loss, reflections and diffraction from buildings, waveguide effect in street canyons, and obstacles blocking line of sight

between Tx and Rx. The pathloss function (sometimes referred to as *path gain* function or *radio map*), is a function that assigns to each Tx-Rx pair of locations x, y the corresponding large-scale signal attenuation $G(x, y)$. Notice that in addition to the large scale effects, wireless propagation is also subject to small-scale fading, due to the superposition of scattered wavefronts with different phases at the Rx location. Such small-scale effects are typically modeled as a Gaussian random variable H that, without loss of generality, can be normalized with unit second moment. Therefore, denoting by $Y = \sqrt{G(x, y)}HX + Z$ a signal sample at the Rx baseband output, where X is the transmitted signal sample with power P_{Tx} , H is the normalized small-scale fading, and Z is the additive noise with power spectral density N_0 , the received energy per sample is given by $\mathbb{E}[|Y|^2] = G(x, y)P_{Tx}/W + N_0$, where W denotes the signal bandwidth, and the Signal to Noise Ratio (SNR) at the input of the Rx baseband processor is given by $\text{SNR} = \frac{G(x, y)P_{Tx}}{N_0W}$. In this paper we develop a deep learning method for estimating radio maps, which we call RadioUNet.

A. Applications of Radio Maps

Many applications in wireless communication explicitly rely on the knowledge of the pathloss function, and thus, estimating pathloss is a crucial task. For example, in device-to-device (D2D) link scheduling, there exists a set of wireless devices that transmit signals to each other in pairs. A pair of devices that communicate defines a Tx-Rx link. The signal sent by a Tx is generally received by multiple Rx's beyond its intended destination, creating mutual interference between the links. While the general information theoretic setting for this problem is the Gaussian interference channel, whose capacity region and optimal coding techniques are still an open problem in general, a huge amount of work has been devoted to the problem of scheduling subsets of links to be active on the same time slot and frequency subband, such that their mutual interference is sufficiently weak and the multiuser interference can be treated as Gaussian noise. It turns out that in a particular regime of weak interference, *Treating Interference as Noise* (TIN) is information-theoretic approximately optimal [1]. Furthermore, efficient link scheduling and power control combined with TIN yields very good performance in comparison with classical interference avoidance schemes such as CSMA [2]. A practical such link scheduling algorithm developed by Qualcomm is FlashLinQ [3]. Recent works on information-theoretic inspired D2D link scheduling include [4], [5], that significantly improve upon FlashLinQ. A recent

Manuscript received May 3, 2020; revised September 25, 2020 and January 3, 2021; accepted January 20, 2021. Date of publication February 12, 2021; date of current version June 10, 2021. The work presented in this paper was partially funded by the DFG Grant DFG SPP 1798 “Compressed Sensing in Information Processing” through Project Massive MIMO-II, and by the German Ministry for Education and Research as BIFOLD - Berlin Institute for the Foundations of Learning and Data (ref. 01IS18037A). The associate editor coordinating the review of this article and approving it for publication was A. El Gamal. (Ron Levie and Çağkan Yapar contributed equally to this work.) (Corresponding author: Ron Levie.)

Ron Levie is with the Department of Mathematics, LMU Munich, 80331 München, Germany, and also with the Institute of Mathematics, TU Berlin, 10623 Berlin, Germany (e-mail: levie@math.lmu.de).

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

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).

Digital Object Identifier 10.1109/TWC.2021.3054977

1536-1276 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.

See <https://www.ieee.org/publications/rights/index.html> for more information.

more direct approach based on fractional programming optimization is provided in [6]. All these schemes somehow assume that the pathloss function between every Tx-Rx locations is known or can be accurately estimated via some probing scheme. A deep learning approach to D2D link scheduling is proposed in [7], which is implicitly based on the fact that interference is a decreasing function of distance and therefore that the pathloss function has a radial symmetry. Therefore, such scheme does not directly apply to more complicated urban propagation scenarios as considered in the present paper. From the above works it is clear that an accurate knowledge of the radio map for a specific environment is very important for efficient D2D links scheduling.

Another classical use-case example of radio maps is base station assignment, or user-cell site association, where the goal is to assign a set of wireless devices to a set of cellular base stations. In order to decide which device to assign to which station, it is important to know the radio map (e.g., see [8] and references therein).

Some additional applications that rely on the knowledge of the pathloss function are fingerprint based localization [9], physical-layer security [10], power control in multi-cell massive MIMO systems [11], user pairing in MIMO-NOMA systems [12], precoding in multi-cell large scale antenna systems [13], path planning [14], and activity detection [15].

B. Radio Map Prediction

A multitude of approaches for estimating the pathloss function have been proposed in the literature. For the sake of clarity, we can group these approaches in three categories.

Data driven interpolation methods assume that some measurements of the pathloss function are given at certain locations. These methods estimate the pathloss function at non-measured locations via some signal processing approach (e.g., *Kriging* [16]) and do not rely—or rely only lightly—on a model of the physical phenomenon. Beyond *Kriging*, other examples of such approaches are radial basis function interpolation [17, Sect 5.1], tensor completion [18], support vector regression [19], and matrix completion [20].

Model-based data fitting methods combine measurements of the pathloss function with a priori assumptions on the physical system to estimate the pathloss function at non-measured locations. For example, in tomography methods, the attenuation due to shadowing can be derived under some modeling assumptions from the so called spatial loss field (SLF), which in turn can be estimated from the measurements. Here, various assumptions on the underlying SLF can be imposed, e.g., low-rank structure [21], sparsity [22], and piecewise homogeneity [23], [24].

Last, *model-based prediction* estimates the pathloss function based only on available prior knowledge, e.g., physical considerations, without taking any measurements from the area of interest. Some examples are ray-tracing [25], dominant path model [26], and empirical models, e.g., [27].

C. Radio Map Prediction Using Deep Learning

Two recent papers proposed deep learning approaches for estimating radio maps [28], [29]. There, the neural network

is a function that returns an estimate of the pathloss for each input Tx-Rx locations. The network is trained on a fixed map and simulated pathloss values at a set of Tx-Rx locations. This procedure is a data-fitting method for the 4-dimensional (4D) function $G(x, y)$.¹ Different city maps require re-training the network and each trained network describes a specific map. In contrast, our RadioUNet learns to approximate the (outcome of the) underlying physical phenomenon, which is independent of a specific city map. Namely, the trained RadioUNet produces a radio map from any given Tx source and city map. We thus think of RadioUNet as a type of implicit simulation, given by the operations of its underlying convolution network. Even when the map is fixed, we show that RadioUNet significantly outperforms previous deep learning proposed methods.

There are several more papers on pathloss prediction that use fully connected neural networks, which do not take the city map information into consideration, and use additional information such as the height of the transmitter/receiver or the distance between them. For example, see the survey [30], and the papers [31]–[33]. These methods are clearly unsuited to predict the radio map as a function of the city map geometry, given as an input to the neural network, which is instead the focus of this paper.

Another recent work based on data-fitting to radio maps via deep learning, in the above fashion, is [34]. The authors of [34] also proposes a transfer learning approach to learn a radio map estimator corresponding to some antenna tilt T_B from a radio map estimator of another tilt T_A . There, it is assumed that there is a large amount of data to train the tilt T_A , and a small amount of data for the tilt T_B . We also consider a transfer learning approach, in which we train a radio map estimator on a large dataset of simulations, and transfer it to real-life with the aid of a small dataset of real-life measurements.²

Slightly after our work, a convolutional autoencoder network was proposed for spectrum map interpolation [35], where multiple transmitters with unknown locations operate simultaneously, and the city map is considered as an input along with measurements with known locations.

D. Our Contribution

In this paper we propose several versions of a radio map estimation method based on deep learning, which we term *RadioUNet*. In our setting, we consider mobile devices/base stations in an urban environment. Our deep learning based methods are efficient, estimating the whole radio map within an area of 256^2m^2 in an order of 10^{-3}sec to 10^{-2}sec , with root mean square accuracy of order 1dB, where the range of pathloss values from the noise floor to the maximal gain is 100dB. This is a mean accuracy of 1% (Root mean square error (RMSE) divided by the range). Some preliminary results were reported in [36], where the proof of concept of estimating radio maps with UNets was illustrated on a preliminary dataset of simulations based only on the city maps (buildings) but not including details such

¹Notice that when x and y are points on the plane \mathbb{R}^2 , the function $G(\cdot)$ has domain in \mathbb{R}^4 .

²Please see point 2) of Section I-D for the concept of “real-life” measurements used in this paper.

as cars along the streets. The source code of RadioUNet can be found at <https://github.com/RonLevie/RadioUNet>, and our dataset, RadioMapSeer, at <https://RadioMapSeer.github.io>. For reproducibility, see the compute capsule at <https://codeocean.com/capsule/ea977fe8-d945-4a49-8326-0c687f96f8ff/tree>.

1) *RadioUNet Methods*: Our radio map estimation methods are based on UNets [37] and their compositions. One version of RadioUNet (called RadioUNet_C) only uses as input the city map (i.e., the geometry of the urban environment), the Tx location, and no pathloss measurements. Thus, this method can be categorized as model-based simulation. However, as opposed to classical model-based simulation, our model is learned from training data. As such, on the one hand it does not have an explicit physically interpretable formulation, but on the other hand, its execution run-time (for the trained network) is much faster than existing model-based tools. Another model that we propose (called RadioUNets, with S for samples) takes as an additional input variable some measurements of the pathloss at a few locations. Thus, this method can be categorized as a model-based data fitting method. Another optional input variable is the locations of cars along the streets, which help predicting the shadowing effect due to the penetration of the signal through cars.

2) *The Training Data*: We present a new dataset, called *RadioMapSeer*,³ of 56,000 simulated radio maps in different city locations and different Tx locations. Each simulation has a number of versions, generated using different types of *coarse* simulations (see details in Section III-A). In one type of simulation, cars are generated along the streets, and affect the outcome of the simulation. The cars serve as unpredictable obstacles perturbing the received signal strength. Alongside each simulation, the map of the city, the Tx locations and cars are also provided.

In addition, we present a smaller dataset of 1400 high accuracy simulations, with and without cars, called IRT4 (see Section III-A). In our setting, IRT4 serves as a surrogate for real-life measured radio maps, i.e., the effective ground truth with respect to which we calculate the prediction error. To imitate a realistic scenario, where the 1400 IRT4 simulations represent real-life measurements collected during a measurement campaign or even in real-time from user devices, each of the 1400 radio maps is only measured sparsely, e.g., we only have 300 receiver locations per map. At this point it is important to point out that the scope of our work is not to assess the accuracy of IRT4 with respect to real-life measurements. The main idea here is that the coarse simulations and IRT4 share the same basic underlying propagation phenomenon, but IRT4 has additional finer details which are not present in the coarse simulations. One goal is then to develop methods to predict such fine details even though in training we have access to a (large) dataset of coarse simulations, but only to a (small) set of sparse measurements of IRT4. Of course, when RadioUNet is employed in practice, the refined phenomenon should be taken as the actual real-life measurements.

³The dataset can be found at <https://RadioMapSeer.github.io>.

3) *Transferability to “Real-Life”*: One important aspect that we address in this paper is how to improve what RadioUNet learned from the coarse simulations to refined representations of the pathloss function. The ultimate goal for the sake of practical relevance is to transfer what RadioUNet learned from simulated data (the labeled training set) to real-life deployments. As a proof of concept, in the RadioMapSeer dataset we use the small set of high accuracy IRT4 simulations as a surrogate for actual real-life measurements. Through this proof of concept, we show that the proposed methods learn the “big-picture” coarse phenomenon from the large coarse simulation datasets, and use the additional smaller set of IRT4 sparse samples to refine and adapt the RadioUNet to the refined phenomenon, using a small subset of the trainable parameters. Given the fact that ray tracing commercial tools are routinely used for wireless networks layout planning (e.g., see [38] or <https://www.remcom.com/wireless-insite-em-propagation-software>), it is clear from engineering experience that high accuracy ray tracing can predict real-life measurements sufficiently well. This corroborates the significance of this proof of concept, where the role of real-life measurements (which are costly and difficult to obtain on such a large scale) is played here by the sparse but accurate IRT4 samples.

A second approach for transferability consists of training a RadioUNet to estimate radio maps from three *input feature channels*, the city map data, the Tx location, and some pathloss measurements. In this case the measurements are also taken from the same coarse simulation. However, once trained, the RadioUNet is then applied in “real life”, where real-life measurements of the pathloss are used as input. Again, in our experiments we test the transferability capability of this approach on the fine IRT4 simulation set.

4) *Applications*: Our RadioUNet can be directly applied to any of the problems mentioned before, where an accurate knowledge of the pathloss function between any Tx-Rx pair of locations is useful. In a dynamic environment, the set of refined measurements can be provided in real-time from the mobile devices, along with their position. For the sake of space limitation, in this work we demonstrate the potential of our radio map estimation method with two toy applications.

a) *Coverage Classification*: We show how to predict the service area of a Tx, and conversely, show how to estimate the domain where the Tx creates small interference with other devices.

b) *Pathloss Fingerprint Based Localization*: Using the estimated radio maps of a set of devices/base-stations with known location, the location of some other device d can be accurately computed if d reports the received signal gains from the base-stations.

II. BACKGROUND AND PRELIMINARIES

A. Wireless Communication

Consider a general Gaussian interference network with K Tx and N Rx devices located over a certain region of the 2D plane. Following the *Generalized Degrees of Freedom* (GDoF) oriented model in [1], it is useful to normalize the received signal such that the variance of the noise samples N_0 and the signal energy per symbol P_{Tx}/W are both equal to 1, and

define a parameter P such that the normalized received signal at each j -th Rx is given by

$$Y_j = \sum_{i=1}^K \sqrt{P\alpha_{i,j}} X_i + Z_j, \quad (1)$$

where $\alpha_{i,j} = \frac{\log \text{SNR}_{i,j}}{\log P}$ and $\text{SNR}_{i,j}$ is the SNR between Tx i and Rx j as defined in Section I. It turns out that the *GDoF region* of the underlying Gaussian interference network (i.e., a high-SNR representation of the capacity region) is defined by the exponents $\alpha_{i,j}$. Furthermore, under certain conditions of weak interference referred to as the TIN regime⁴ (see [1], [5] for the information-theoretic details) the GDoF region yields the actual *capacity region* within a bounded gap, independent of the SNR scale parameter P . These facts provide a strong evidence that the relevant notion of pathloss function is contained in the α 's exponents, i.e., the pathloss function should be estimated in logarithmic scale (in dB). Furthermore, from the theory in [1] it follows that negative values of the α 's exponents are irrelevant, that is, for the GDoF region it is sufficient to take the positive part of the $\alpha_{i,j}$'s. In practice, this means that we do not have to spend much effort in estimating very large negative values (in dB) of the pathloss function. As a matter of fact, it makes sense to truncate such function so that the received signal power is not too much smaller than the noise floor.

Driven by the above considerations, we define the pathloss in dB scale as $P_L = (P_{R_x})_{\text{dB}} - (P_{T_x})_{\text{dB}}$, where P_{T_x} and P_{R_x} denote the transmitted power and received power at the Tx and Rx locations, respectively. The truncation and rescaling of the pathloss function in dB scale in order to make it suitable for the proposed deep learning estimation method is given in Sections III-B and III-C.

B. Deep Learning

In this section we review some concepts from deep learning necessary for the understanding of this paper.

1) *Convolutional Neural Networks*: A *Convolutional neural network* (CNN) is a popular deep learning architecture, typically used in machine learning applications in imaging science [39], [40]. In our context, a *feature map* is a function from a 2D grid to some \mathbb{R}^N , where N is called the number of *feature channels*. If $N = 1$, we call the feature map a *gray level image*.

A CNN is defined by aggregating the following five basic computational steps as the layers of the network. i) In a *convolution layer* an input feature map is convolved with a filter kernel and added to some scalars called the *bias*. The number of input feature channels and output feature channels need not coincide. More accurately, let N be the number of input feature channels and M the number of output feature channels. Let f_1, \dots, f_N be the components of the input feature map. Note that each f_n is a gray level image, not a scalar. The components of the output feature map, g_m , are

defined for every $m = 1 \dots, M$ by

$$g_m = \sum_{n=1}^N f_n * y_{n,m} + b_m \quad (2)$$

where $*$ denotes convolution, and for each $m = 1, \dots, M$ and $n = 1, \dots, N$, $y_{n,m}$ is a gray level filter kernel, and b_m is the m -th component of the bias. ii) An *activation function* is any function applied on the entries of a feature map, and a typical choice is ReLU, defined by $r(z) = \max\{0, z\}$. iii) A *pooling layer* takes a feature map and down-samples it, e.g., by assigning the maximal entry of each 2×2 patch to the corresponding entry of the down-sampled feature map. iv) An *up-sampling layer* up-samples lower resolution feature maps to higher resolution ones. v) A *fully connected layer* is a general linear operator/matrix applied on the feature map, and added to some pre-defined bias. A CNN architecture is defined by choosing how to combine the above layers, choosing the number of feature channels, and choosing the shapes of the filter kernels. The trainable parameters are the filters, the fully connected matrices, and the biases.

2) *UNets*: UNet is a special CNN architecture, introduced in [37], and used in a multitude of applications, including image segmentation [41]–[44], video predicting [45], super resolution/image inpainting [46], inverse problems in imaging [47], image-to-image translation [48], and medical image analysis [49] to name a few.

UNets consist of convolution, pooling, up-sampling, and activation function layers, without fully connected layers. The UNet architecture is divided into two paths. The first portion of the layers gradually contracts the image as the layers deepen, and gradually increases the number of feature channels. This path—also called the *encoder*—is interpreted as a procedure for extracting “concepts” which become more complex/high-level and less spatially localized along the layers. The second portion of the layers—also called the *decoder*—expands the image as the layers deepen and reduces the number of feature channels gradually. This path is interpreted as a procedure of combining/synthesizing the concepts, layer by layer, to lower-level concepts, and eventually to an output image. The decoder layers are derived by up-sampling lower resolution images, and thus lack high resolution information on their own. To provide high resolution information to the decoder layers, the feature maps in the encoder layers are copied and concatenated to the corresponding feature maps of the decoder layers having the same resolution. This copying between non-neighboring layers is called *skip connection*.

Let \mathbf{p} denote the concatenation of all learnable parameters of the UNet, and let $U_{\mathbf{p}}$ denote the UNet, mapping input images \mathbf{f} to output images $U_{\mathbf{p}}(\mathbf{f})$. In supervised learning, a *training set* of many example inputs images \mathbf{f}_k and corresponding desired output images \mathbf{g}_k are given, where $k = 1, \dots, K$ and K is the size of the dataset. The goal is to fine tune the parameters \mathbf{p} of $U_{\mathbf{p}}$ so that $U_{\mathbf{p}}(\mathbf{f}_k) \approx \mathbf{g}_k$ for every $k = 1, \dots, K$. This is typically done by some variant of gradient descent, and the loss function to be optimized is

⁴In the information theoretic literature TIN stands for “Treating Interference as Noise” and the TIN regime is when TIN achieves the GDoF region, i.e., it is GDoF-optimal.

typically of the form

$$\mathcal{L}(\mathbf{p}) = \frac{1}{K} \sum_k \|\mathbf{g}_k - U_{\mathbf{p}}(\mathbf{f}_k)\| \quad (3)$$

for some norm, e.g., the root mean square norm. In *stochastic gradient descent* (SGD), k runs over one batch in each iteration.

3) *Curriculum Learning*: The SGD optimization procedure (and its variants) explores configurations of the parameters only along the 1D path of descent, which might miss good configurations. Namely, SGD searches the parameter space in a highly non-exhaustive manner. Thus, the expressive capacity of a network does not guarantee high quality trained networks. It is thus often important to lead gradient descent in a more deliberate way, and in some sense to “micro manage” the exploration of parameter configurations in the optimization process. One approach for achieving this is called *curriculum learning* [50]. In curriculum learning, training is divided into a *curriculum*, namely, a list of optimization problems, where the optimal solution of the previous problem is used as the initial guess for the next optimization problem. The idea is to first teach the network how to solve an easy to learn simplified version of the problem, and gradually to increase the complexity of the problem until reaching the original formulation of the loss function.

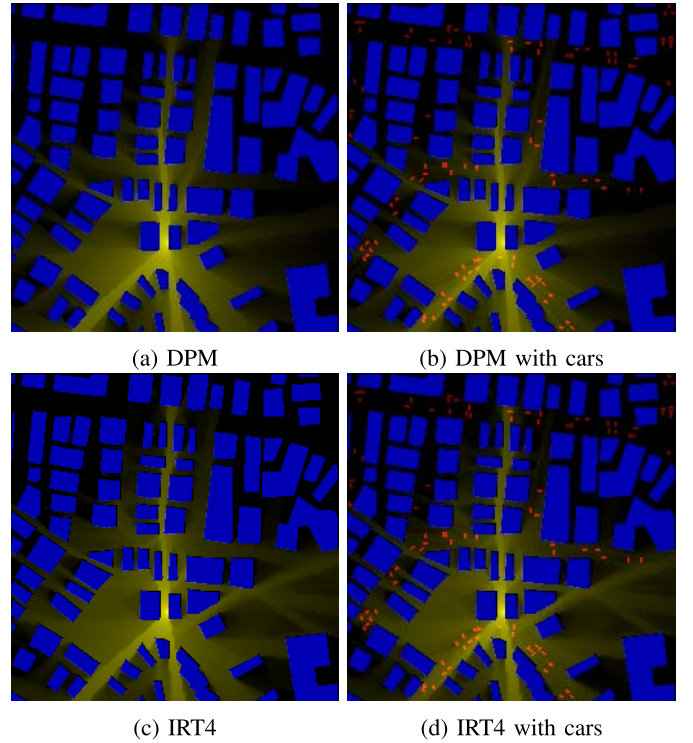
4) *Transfer Learning*: In some learning scenarios the training data does not represent exactly the data in the target application (e.g., when a large enough training set is difficult or costly to obtain). It is thus important to know whether the network, trained on one data distribution, performs well for another data distribution. The idea of training in one domain and testing in another domain is called *transfer learning* [51], [52]. The capacity of a network to perform well in new domains is called its *transferability*.

III. THE RADIOMAPSEER DATASET

In this section we introduce RadioMapSeer, a dataset of city maps with corresponding simulated radio maps that we have created and made available for this work.

A. General Setting

The RadioMapSeer dataset consists of 700 maps, 80 transmitter locations per map, and corresponding coarsely simulated radio maps. The coarse simulations are generated using the Dominant Path Model (DPM) method [26] and Intelligent Ray Tracing (IRT) [53] based on 2 interactions of the rays with the geometry, referred to here as IRT2. In addition, we have also generated fine simulations using IRT with 4 interactions (IRT4), for the first two transmitters of each map. The city maps are taken from *OpenStreetMap* [54] in the cities Ankara, Berlin, Glasgow, Ljubljana, London, and Tel Aviv. We set the heights of the transmitters, receivers and buildings as 1.5m, 1.5m, and 25m, respectively, which is relevant to device-to-device scenarios (see Section III-B for more details). All simulations were computed using the software *WinProp* [38]. For different situations (e.g., campus networks, cellular networks) one should generate new data sets accordingly. Some example radio maps from the dataset are



Parameter	Value
Number of transmitters	80
Frequency	5.9GHz
Bandwidth	10MHz
Pixel length	1 meter
Noise power spectral density	-174dBm/Hz
Transmit power	23dBm
Noise figure	0dB

(e) RadioMapSeer Dataset parameters

Fig. 1. RadioMapSeer examples and parameters. Buildings are blue, cars red, and pathloss yellow.

shown in Figure 1. All simulations are saved as dense sampling of the radio map in a 2D grid of $256 \times 256 \text{ m}^2$.

1) *Maps and Transmitters*: Each map is $256 \times 256 \text{ m}^2$ where buildings and roads are saved in the dataset as polygons. Each map is also converted to morphological 2D image (binary 0/1 pixel values with no intermediate gray levels) of 256×256 pixels, where each pixel represents one square meter. The interior of the buildings has pixel value = 1, and the exterior of the buildings has pixel value = 0. The transmitter locations are stored as numerical 2D values, and also given as morphological images, where the pixel in which the transmitter is located has value 1 and the rest is 0. Along with the city maps, roads are given both as polygonal lines and as morphological images with 1 on the road and zero outside. Cars are generated along and aside roads, and given as separate morphological images.

2) *Coarsely Simulated Radio Maps*: The coarse radio maps were generated using DPM and IRT2 with the radio network planning software *WinProp* [38]. Each simulated radio map stores at each pixel the pathloss between the pixel location and the transmitter location in dB.

To represent uncertainty in the dataset we consider two cases. First, a set of simulations on all city maps including the cars is produced using DPM and IRT2. These simulations are perturbations of the simulations based on the city map alone, without cars. We moreover provide separate datasets of perturbed city maps, where in each map of the original dataset m buildings are missing. We provide four such datasets with $m = 1, \dots, 4$.

3) *Higher Accuracy Simulations*: An additional smaller dataset of higher accuracy simulations is provided by using IRT4 with the same *WinProp* radio network planning software. Here, for each of the 700 maps we consider two transmitter locations. The goal of the higher accuracy simulations is to provide means of testing whether the network, trained on simulations, performs well in refined representations of pathloss functions. As already said before, we reiterate here that the high accuracy simulation serves as a surrogate for the real-life physical phenomenon, and it is useful for demonstrating the transferability property of our scheme.

4) *Pathloss Scale*: The pathloss values P_L are converted to *gray level* pixel values between 0 and 1 (see Subsection III-C). Therefore, each radio map is a gray level image of size 256×256 .

B. System Parameters

This study was originally motivated by device-to-device communications for safety in the context of intelligent transportation systems (ITS), currently based on the IEEE 802.11p standard. Accordingly, we consider a signal bandwidth W of 10MHz in the 5.9GHz band. We choose the transmitter power and thermal noise power spectral density as $(P_{Tx})_{dB} = 23dBm$ and $(N_0)_{dB} = -174dBm/Hz$ in compliance with IEEE 802.11p and assume an idealistic noise figure of 0dB at receivers (cf. Fig. 1(e) for a summary of the system parameters).

We express by $(\mathcal{N})_{dB} = 10 \log_{10} W N_0 + NF$ the noise floor in dB, with NF being the noise figure. We consider the points where the received signal power $(P_{Rx})_{dB} = P_L + (P_{Tx})_{dB}$ yields a signal-to-noise ratio above a desired SNR level, i.e. the points where $(SNR)_{dB} = (P_{Rx})_{dB} - (\mathcal{N})_{dB} \geq SNR_{thr}$ holds. Solving this for P_L we get the threshold $P_{L,thr}$ for the pathloss

$$P_L \geq P_{L,thr} = -(P_{Tx})_{dB} + SNR_{thr} + (\mathcal{N})_{dB}. \quad (4)$$

We call $P_{L,thr}$ the *pathloss threshold*. Consider for example the SNR requirement that the received signal power should be above the noise floor, i.e., when $SNR_{thr} = 0$. With the choice of parameters in Fig. 1(e), we find $P_{L,thr} = -127dB$.

One task of RadioUNet is to extract the area in the city map above the noise floor, given an input city map and transmitter location. To do this, the network must learn the physical phenomenon both above and below the noise floor. We thus truncate the pathloss values below another threshold $P_{L,tmc} < P_{L,thr}$. We choose $P_{L,tmc}$ such that the difference between the maximum pathloss M_1 in the dataset and $P_{L,thr}$ is approximately four times greater than the difference between $P_{L,thr}$ and $P_{L,tmc}$, i.e., $M_1 - P_{L,thr} = 4(P_{L,thr} - P_{L,tmc})$. The maximum and the minimum pathloss in the dataset are $-47.84dB$ and

$-186.41dB$, respectively. Note that the maximum is $-47.84dB$ and not $0dB$ since the pathloss is integrated over $1m^2$ pixels. To meet the previously mentioned condition, we set $P_{L,tmc} = -147dB$. Since any signal below $P_{L,thr}$ cannot be detected in practice, and is only used in simulation for theoretical reasons, we call $P_{L,tmc}$ the *analytic noise floor*.

C. Gray Level Conversion

We convert the pathloss values P_L to pixel values between 0 and 1 as follows. Denote by M_1 the maximal pathloss in all radio maps in the dataset, and define $f = \max\{\frac{P_L - P_{L,tmc}}{M_1 - P_{L,tmc}}, 0\}$. Here, $f = 0$ represents anything below the analytic noise floor, and $f = 1$ represents the maximal gain at the transmitter. Any intermediate value is referred to as a *gray level*.

Let us explain the importance of our gray level conversion when evaluating the performance of any pathloss estimation. We evaluate performance of any approximation $\tilde{f} : \mathcal{D} \rightarrow \mathbb{R}$ of a signal/image $f : \mathcal{D} \rightarrow \mathbb{R}$, where $\mathcal{D} = \{x_n\}_n$ is some finite grid in \mathbb{R}^2 , via the normalized mean square error (NMSE)

$$E = \frac{\sum_n |\tilde{f}(x_n) - f(x_n)|^2}{\sum_n |f(x_n)|^2}. \quad (5)$$

The numerator in (5) represents the absolute error, and the denominator represents the global magnitude of f . The coefficients $|\tilde{f}(x_n) - f(x_n)|^2$ and $|f(x_n)|^2$ having larger values affect the outcome of E the most, and small values are negligible. It is thus crucial to express the signal f in a representation in which the important parts of the signal obtain large values.

In our case, the representation of the radio map should be constructed in such a way that small powers contribute small values to E . Indeed, locations of small power represent a weak signal. If we represent the radio map as standard pathloss, in dB, the smaller the power in a certain location, the higher the magnitude of the pathloss, with negative sign. When the power goes to zero, the pathloss diverges to $-\infty$. In this representation, locations of a weak signal dominate the global magnitude of the radio map, and in general define a misleading concept of the “size” of the radio map. A similar situation occurs for the absolute error (the numerator of (5)).

As discussed in Section II-A, motivated by the GDoF region of a Gaussian interference network, we know that very large negative values of the pathloss are effectively irrelevant and should not dominate the overall error. Our gray level conversion resolves this issue. Indeed, anything below the noise floor, or more generally, below $P_{L,tmc}$, is deemed to be “too small to be interesting”, and set to zero. In contrast, the values of higher power, which are most important, are transformed to levels close to 1. We note that papers like [18], [20], [21] suffer from the aforementioned shortcoming, and it is thus difficult to interpret their reported performance.

When root mean square error (RMSE) is used, the gray level error is simply a scaling of the RMSE of the pathloss in dB (up to the truncation below the analytic noise floor). More precisely, we have

$$\sqrt{\sum_n |\tilde{P}_L(x_n) - P_L(x_n)|^2} = C \sqrt{\sum_n |\tilde{f}(x_n) - f(x_n)|^2},$$

where P_L is the pathloss in dB. For $\text{SNR}_{\text{thr}} = 0$ we have $C = 80$.

Remark 1: It is worthwhile noticing that the RMSE of the pathloss in dB scale (up to rescaling as explained above) comes here not by accident, and in fact it is a very sensible choice for the approximation error from a communication theory significance viewpoint. The pathloss P_L in natural scale operates as a multiplier of quantity $\frac{P_{\text{Tx}}}{N_0 W}$ in order to yield the SNR at the Rx location. If P_L is perfectly known, then the pathloss prediction result can be immediately translated into a receiver rate result, using either the Shannon capacity formula for the appropriate channel model, or a table of rates versus Rx SNR according to the family of coding and modulation schemes specified by a particular communication standard. On the other hand, if P_L is known up to the approximation RMSE σ in dB scale, we can roughly model such approximation error as normal with standard deviation σ (see Figure 7).⁵ This means that the pathloss P_L in linear scale is known up to a log-normal shadowing fluctuation with parameter σ . Then, one can easily provide performance guarantees in terms of rate versus outage probability, where the latter can be evaluated from the tail of the log-normal distribution. Since these results are highly dependent on the system assumptions, we provide here just a simple example, leaving to the reader the generalization to any system of choice. Using the capacity formula for the AWGN channel and assuming a Tx transmission rate R bit/s/Hz, a receiver with pathloss (in linear scale) $P_L \times \Delta$, where P_L is the exact value and Δ is the log-normal error, can decode successfully if R is strictly less than the capacity $\log_2 \left(1 + P_L \times \Delta \times \frac{P_{\text{Tx}}}{N_0 W} \right)$. Then, the probability of (block) decoding error for such ideal coded system is given by the *information outage probability*, i.e., the probability that the Shannon capacity falls below the transmission rate R [56]. Using the fact that $10 \log_{10} \Delta$ is normally distributed $\mathcal{N}(0, \sigma^2)$, we obtain immediately:

$$P_{\text{out}}(R) = 1 - Q \left(\frac{(2^R - 1)_{\text{dB}} + (N_0 W)_{\text{dB}} - (P_{\text{Tx}})_{\text{dB}} - (P_L)_{\text{dB}}}{\sigma} \right),$$

where $Q(x) = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du$ is the Gaussian tail function. Hence, knowing P_L and the RMSE σ it is possible to obtain results in terms of receiver rate R for a given guaranteed block error probability $P_{\text{out}}(R)$ at any location x_n of the map. \diamond

IV. ESTIMATING RADIO MAPS VIA RADIOUNETS

In this section we introduce a number of methods, collectively called RadioUNet, that learn to estimate radio maps in different scenarios. We evaluate the accuracy of the proposed methods and compare them to state-of-the-art.

A. Motivation for RadioUNet

UNets have been extensively applied to imaging problems in the past few years with resounding success, and are considered

to be a baseline method for image-to-image tasks [57]. Our problem can be seen as mapping an image representing the city and Tx to an image representing the radio map, and hence using UNets is a natural choice. One advantage of using UNets in our case is that they respect the translation invariance symmetry of the physical phenomenon. Namely, this symmetry is built in to RadioUNet, and requires no training. Another strong point of UNets is the encoder-decoder interpretation, as we discuss next.

In Fig. 8 we show an example of a ground truth radio map generated by simulation, and the estimated radio map computed by the RadioUNet_C and RadioUNet_S. Aside from the low quantitative error, RadioUNet seems to synthesize radio maps from the urban geometry which qualitatively captures the correct shadow patterns. Note that the results in Fig. 8 are representative of the general quality of RadioUNet. One might naively interpret the success of the RadioUNet by postulating that it learns to mimic a physical model, like ray-tracing or some differential equation like Maxwell's equations. However, we believe that this is a misleading viewpoint. A more reasonable interpretation follows from the encoder-decoder description of general UNets. In the encoder path, the RadioUNet extracts complicated concepts about the geometry of the urban environment and the mutual relationship between the different geometric features, their location, and the location of the transmitter. Then, in the decoder path, the RadioUNet uses these concepts to synthesize the radio map. Thus, RadioUNet is based on extracting and analyzing *global* information about the urban environment, as opposed to classical physical models that are based on *local* information, like collisions with the geometry in ray-tracing and derivatives in differential equations. In this viewpoint, it is more fitting to compare RadioUNet to a highly skilled artist that draws radio maps from his/her perception of the urban environment as a whole, rather than comparing to a classical local physical model.

B. Different Setting in Radio Map Estimation

We consider the following scenarios for the input of the UNet, the map of the city, the learning setting, and the properties of the simulated dataset. The problem setting can be any combination of the choices presented in Subsections IV-B.1 and IV-B.2.

1) Network Input Scenarios:

a) **City Map and Transmitter Location:** In the first case, the UNet receives as input the map of the city and the Tx location as morphological images. From these two input feature channels the network estimates the radio map.

In this *accurate map scenario*, if the simulated dataset without cars is used, then the map without cars is given as input, while if the simulated dataset includes cars, then the map without cars is given as one feature channel, and the cars in an additional input feature channel.

When the map is accurate and the simulated data used for training is assumed to represent reality accurately, the radio map is uniquely determined by the map and the Tx location. Thus, the input feature channels are sufficient for high quality radio map reconstruction.

⁵Notice that approximating a random variable with zero mean and given standard deviation σ as normal $\mathcal{N}(0, \sigma^2)$ yields the maximum-entropy approximation, which has been widely used in statistics [55].

b) City Map, Transmitter Location, and Measurements:

In the second case, the UNet receives as input the two/three feature channels of map and Tx location as before, and an additional feature channel of measurements of the “true” radio map. The measurements are taken at some locations on the true map, i.e., their values are sampled from the target “ground truth”. This third feature channel is given as a gray level image, where in the pixels corresponding to the locations of the measurements the gray level value is the measurement. Non-measured pixels are set to zero. The network estimates the radio map from these three/four input feature channels.

This scenario is useful when the “nominal” map given as input feature channel does not represent reality completely accurately. Hence, the network learns a hybrid of a radio map estimation method based on the given map, which is not completely reliable, and an interpolation method of the accurate pathloss measurements. In this *non-accurate maps scenario*, a perturbed version of the ground truth maps is given as input to the UNet. We consider two types of perturbations: 1) the map is given with a one to four missing buildings; 2) the map is given without cars, but the ground truth simulation is computed with the cars.

Another source of inaccuracy, for which relying on measurements is useful, is the fact that training is done against coarse simulations, which are only approximations of reality (or, in our setting, approximations of IRT4).

2) Learning Scenarios:

a) Large and Dense Simulation Dataset: Here, the network is trained in supervised learning to predict a large dataset of 2D gray-level images representing dense measurements of radio maps on a fine grid. The images are the DPM simulations, the IRT2 simulations, both with or without cars, or random combinations of DPM and IRT2. In particular, the goal in the randomized simulation is to push the network to learn that it can only rely on the simulations for the big-picture behavior of radio maps, shared both by DPM and IRT2, but not on the fine details. This pushes the network to use additional information for refining the estimations, like the input measurements if given, or the smaller dataset of sparse IRT4 if given.

Transferring the trained network to the ground truth (IRT4 or real-life maps) is a *zero-shot generalization*. Namely, the network only learned to estimate the coarse simulations, not ever seeing the ground truth phenomenon, and we rely on the accuracy of the simulations, and optionally on the measurements, to predict the ground truth radio maps.

In case measurements are given as an input feature channel to the RadioUNet, real-life measurements would be given to the RadioUNet in the real-time operations, even though measurements from the crude simulation are used in training. Real-life measurements can be provided in real-time directly from the deployed devices, e.g., from the beacon signals of the transmitters, in the same way current systems report “Channel Quality Indicators” as measurements of the received signal strength. Hence, no costly measurement campaign is needed for training. The network can generalize well to real-life radio maps since it learned to interpolate the measurements, which are now accurate, while what was learned from the crude

First UNet										
Layer	In	1	2	3	4	5	6	7	8	9
Resolution	256	256	128	64	64	32	32	16	8	4
Channels	2/3/4	6	40	50	60	100	100	150	300	500
Filter	3	5	5	5	5	3	5	5	5	4
Layer	10	11	12	13	14	15	16	17	18	out
Resolution	8	16	32	32	64	64	128	256	256	256
Channels	300+300	150+150	100+100	100+100	60+60	50+50	40+40	20+6+2/3/4	20+2/3/4	1
Filter	4	4	3	3	6	5	6	5	5	-

Second UNet										
Layer	In	1	2	3	4	5	6	7	8	9
Resolution	256	256	128	64	64	32	32	16	8	4
Channels	3/4/5	20	30	40	50	60	70	90	110	150
Filter	3	5	5	5	5	3	5	5	5	4
Layer	10	11	12	13	14	15	16	17	18	out
Resolution	8	16	32	32	64	64	128	256	256	256
Channels	110+110	90+90	70+70	60+60	50+50	40+40	30+30	20+20+3/4/5	20+3/4/5	1
Filter	4	4	3	3	6	5	6	5	5	-

Fig. 2. RadioUNet architecture. *Resolution* is the number of pixels of the image in each feature channel along the x, y axis. *Filter* is the number of pixels of each filter kernel along the x, y axis. The input layer is concatenated in the last two layers.

simulations roughly guides the interpolation procedure to be physically feasible. We demonstrate this experimentally by training on coarse simulations and using IRT4 samples and targets (as a proxy for real-life measurements) in testing.

b) Large and Dense Simulation Dataset + Small Sparse Measured Dataset: Here, in addition to the large dataset of dense measurements, we also assume that we have a small dataset of sparse measurements taken from refined radio maps (the IRT4 simulations, which can be potentially replaced by real-life measurements). For each of the 700 maps of the RadioMapSeer dataset we consider two transmitter locations, and measurements in K receiver locations, where K is fixed, e.g., $K = 300$. In this scenario we first train a large network that estimates the crude simulations, using the large simulation dataset. Then, we improve the network output, using a smaller network, to match the small dataset of real-life measurements (see Subsection IV-C.2).

C. RadioUNet Architectures

The simplest RadioUNet comprises of one UNet. The input of the UNet has two, three or four feature channels, depending if measurements and cars are used, and the output is the one feature channel estimated radio map. In most architectures of RadioUNet we compose a second UNet on the first one. We call such an architecture a WNet (U + U makes a W). The input of the second UNet are the same as the inputs of the first UNet, plus an additional feature channel, the output of the first UNet. The architectures of our proposed UNets are reported in Fig. 2. The number of layers and feature channels were crudely searched to reduce overfitting and increase performance on the validation set, while not being too large to allow fast inference. The second UNet can be used for three different purposes, summarized in the following three subsections.

1) Retrospective Improvement: The idea here is to give RadioUNet a chance to improve its estimation in retrospective. The first UNet learns implicitly an algorithm for estimating the radio map from the input, by extracting high level concepts from the map and synthesizing a radio map from them. The philosophy here is that it would be beneficial to inspect the resulting estimation, and correct visible inconsistencies with

RadioUNet Accuracy						
Setting	RadioUNet c Test Accuracy					
	Coarse Simulations		Zero Shot IRT4		Adaptation to IRT4	
	NMSE	RMSE	NMSE	RMSE	NMSE	RMSE
Accurate Map						
deterministic DPM simulation	0.0075	0.02	0.0284	0.0384	0.0166	0.0292
deterministic IRT2 simulation	0.0219	0.032	-	-	0.0143	0.0271
non-deterministic simulation	0.0152	0.0272	0.0324	0.0405	0.0135	0.0262
Missing Four Buildings						
deterministic simulation	0.102	0.0742	0.1205	0.0759	0.1015	0.0735
non-deterministic simulation	0.1156	0.0769	0.1153	0.0783	0.1013	0.0726
Cars						
deterministic simulation with unknown cars	0.0132	0.0256	0.0357	0.0412	0.0249	0.0343
deterministic simulation with input cars	0.0092	0.0207	0.0315	0.0385	0.0201	0.0308
Setting	RadioUNet s Test Accuracy					
	Coarse Simulations		Zero Shot IRT4		Adaptation to IRT4	
	NMSE	RMSE	NMSE	RMSE	NMSE	RMSE
Accurate Map						
deterministic DPM simulation	0.0052	0.0164	0.0183	0.0307	0.0135	0.0262
non-deterministic simulation	0.0068	0.0183	0.0122	0.0245	0.0086	0.0209
Missing Four Buildings						
deterministic simulation	0.0321	0.0415	0.0409	0.0474	0.04	0.043
non-deterministic simulation	0.0443	0.039	0.0417	0.0437	0.0372	0.041
Cars						
deterministic simulation with unknown cars	0.0072	0.0187	0.0197	0.0304	0.0156	0.0269
deterministic simulation with input cars	0.0062	0.0173	0.0195	0.0305	0.0156	0.027

Fig. 3. Comparison of RadioUNet accuracy in different scenarios.

the map and with the physical phenomenon. To inspect the output of the first UNet, a second UNet extracts high level concepts from the estimated radio map, the city map, and all other inputs, and synthesizes from these concepts an improved estimation of the radio map. We observe that the retrospective improvement yields better performance especially when the first UNet is small (see Fig. 5). This WNet is thus a technique for reducing the size of the RadioUNet without degrading performance.

The WNet is trained in a curriculum. The first UNet is trained first to estimate the ground truth radio maps, with MSE loss. In the second phase, the weights of the first UNet are frozen, and the second UNet is trained to estimate the ground truth radio maps with MSE loss.

2) *Adaptation to Refined Measurements*: Here, we first train the first UNet to estimate coarse simulations from the large dataset with MSE loss. The simulations may be randomized or deterministic. After training, the weights of the first UNet are frozen, and the second UNet is trained to improve the estimation of the first UNet on the small dataset of IRT4.

The IRT4 training consists of sparse images, namely, for each map, there are K Rx locations $\{x_k\}_{k=1}^K$, and the pathloss $f(x_k)$ is only known for these locations. We typically take $K = 300$. The loss function for the second UNet is the weighted MSE, with weights $W_k = \frac{1}{K}$ for the points $\{x_k\}_{k=1}^K$, and weight 0 for the unmeasured points. We train the adaptation UNet in two steps. First, we train a retrospective improvement UNet on the coarse dataset, and then we further train this UNet on the sparse IRT4 dataset.

RadioUNet Accuracy on 200-test set and 100-test subset						
Setting	RadioUNet c Test Accuracy					
	Coarse Simulations		Zero Shot IRT4		Adaptation to IRT4	
	200-set RMSE	100-set RMSE	200-set RMSE	100-set RMSE	200-set RMSE	100-set RMSE
Accurate Map						
deterministic DPM simulation	0.0203	0.0206	0.038	0.0384	0.0288	0.0295
Missing Four Buildings						
deterministic simulation	0.0733	0.073	0.0736	0.0723	0.0685	0.0693
Setting	RadioUNet s Test Accuracy					
	Coarse Simulations		Zero Shot IRT4		Adaptation to IRT4	
	200-set RMSE	100-set RMSE	200-set RMSE	100-set RMSE	200-set RMSE	100-set RMSE
Accurate Map						
non-deterministic simulation	0.0225	0.0226	0.0249	0.025	0.0217	0.0217
Missing Four Buildings						
non-deterministic simulation	0.0459	0.0462	0.0456	0.0445	0.0442	0.0468

Fig. 4. Performance of selected RadioUNet methods on the 400/100/200 train/validation/test split. The last 100 test examples of the 200 test set are the 100 test examples of the original 500/100/100 split. The performance on the 200 and 100 test sets is comparable, indicating that the 100 test set is not “too small”.

3) *Thresholder*: A thresholder second UNet is used in the service area classification method. The goal of the second UNet here is to take the estimated radio map of the first UNet and to produce a service map from it. More details are given in Subsection VI-A.

D. Training

The 700 maps of the RadioMapSeer dataset are randomly split into 500 training maps, 100 validation maps, and 100 test maps. The random split is fixed, and available in the project web page⁶. We aim in this split to have a large enough training set, for avoiding overfitting, and a large enough test set, to avoid test set bias. To illustrate that this split is reasonable, we also consider a 400/100/200 train/validation/test split, where the last 100 test example of the 200 are the 100 test examples of the original 500/100/100 split. After training on the 400/100 train/validation split, the error on the 200 test set is very close to the error on the last 100 test examples (see Fig. 4). Hence, there does not seem to be a visible bias in the original 100 test set.

We perform supervised learning on the RadioMapSeer dataset. The loss function is the MSE between the inferred radio maps by RadioUNet and the simulation radio maps from the training set. Training of all methods was performed with Adam [58], with learning rate of 10^{-4} . We take 50 epochs for each UNet, no regularization, and batch size 15. To alleviate overfitting, out of the 50 epochs we pick the model with smallest error in the validation set. Lastly, the models are tested either on the coarse simulations on the test maps, or on the IRT4 simulations on the test maps. Performance is evaluated by RMSE on the gray levels and by NMSE (normalized MSE). Note that the RMSE in dB is 80 times the RMSE of gray level.

E. RadioUNet Performance

In Fig. 3 we report the results in all of the above settings. Recall that RadioUNet_C and RadioUNet_S denote the

⁶<https://github.com/RonLevie/RadioUNet>

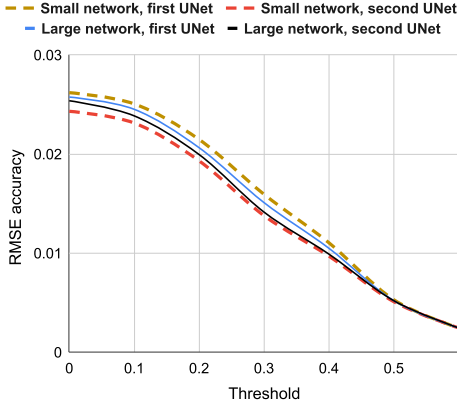


Fig. 5. Accuracy of RadioUNet_C of difference network sizes and different pathloss thresholds $P_{L,thr}$. The small network has 6,109,271 parameters and large one has 25,411,831 parameters. We plot the accuracy of the RadioUNets with and without the retrospective improvement. Small networks outperform large networks when both have retrospective improvement. The accuracy of RadioUNet with pathloss threshold at pixel value 0.6 is comparable to the quantization error of the png image file.

RadioUNet based on no input measurements and input measurements, respectively. From the table we can observe that both the adaptation method to sparse IRT4 samples, and the training with randomized coarsely simulated maps, promote transferability. All accuracies are given both in NMSE and RMSE. RMSE is the square root of the MSE on the whole test set. The pathloss threshold is taken as $P_{L,thr} = -127\text{dB}$. The best results on IRT4 for each category are marked in bold face. RadioUNet_S was trained and tested with a random number of input measurements between 1 and 300. *Zero-shot IRT4* means testing the methods, trained on coarse simulations, on IRT4. *Adaptation to IRT4* means training a second small UNet to match the sparse IRT4 measurements. All architectures are based on the WNetS of Fig. 2, where for zero-shot transfer the second UNet is a retrospective improvement, and for adaptation to sparse IRT4, the second UNet is the adaptor. The receiver points of the sparse IRT4 dataset are randomly generated for each map, and fixed forever. For RadioUNet_C, the sparse IRT4 dataset has 300 receivers per transmitter. For RadioUNet_S, the sparse IRT4 dataset has 600 receivers per transmitter, out of them 1 to 300 random points are taken as input points of the RadioUNet_S. The training loss is computed for all 600 points. To show that the higher transferability of the random simulations is not simply because IRT2 is closer to IRT4 than DPM, we also include the scenario where the deterministic simulation is IRT2. This produces inferior results compared to the random simulations.

In Fig. 5 we compare RadioUNet_C with and without retrospective improvement for different pathloss thresholds. The results demonstrate that the retrospective improvement UNet is effective when the first UNet is small, thus making it a useful strategy for reducing the network size for the same accuracy. In Fig. 6 we compare the performance of different RadioUNet_S methods on maps with various numbers of missing buildings. We observe that the strategy of combining random coarse simulations with an adaptor UNet to IRT4 promotes transferability.

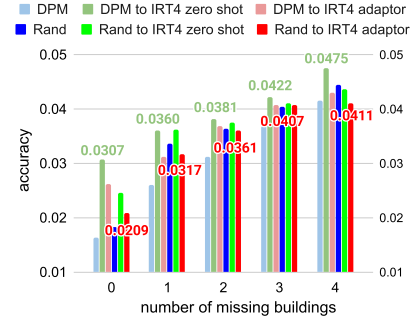


Fig. 6. Accuracy of RadioUNet_S with different numbers of missing buildings, different types coarse simulations, and different transfer methods to sparse IRT4.

V. COMPARISON OF RADIOUNET TO STATE-OF-THE-ART

In Fig. 9(b) we present the performance of different methods of radio map estimation. For methods that depend on samples, we use an input map with four missing buildings, and for methods that do not rely on samples we use the full map. Apart from the fact the RadioUNet outperforms the data driven interpolation methods, the tomography method and the previously proposed deep learning approach significantly, these other methods need a separate training/optimization to fit the model to *each* map. Particularly, variations in the environment, like moving cars, requires re-computing the methods, which is not efficient. RadioUNets, in comparison, are trained offline only once, and are then employed in any environment very efficiently. RadioUNet can deal with cars by using the measurements input, where the network is trained on a dataset of simulations with cars. All GPU methods ran on Nvidia Quadro GV100, and CPU methods on Intel Core i7-8750H.

A. Comparison to Model-Based Simulation

We compare the run-time of RadioUNet⁷ with DPM, IRT2 and IRT4. To penalize RadioUNet, we compare run-time on an Intel Core i7-8750H CPU, which is a highly non-optimal platform for convolution networks. RadioUNet estimates radio maps roughly one to three orders of magnitude faster than the simulation methods. In our experiments, WinProp completes a DPM simulation in roughly an order of 1sec on the CPU, and IRT2 and IRT4 take an order of 10sec and 10²sec, respectively. RadioUNet takes an order of 10⁻¹sec on the CPU, and 10⁻³sec to 10⁻²sec on NVIDIA Quadro GV100 GPU.

B. Comparison to Data Driven Interpolation

Next, we compare RadioUNet_C and RadioUNet_S with data driven interpolation methods: radial basis function (RBF) interpolation using multiquadric function [17, Sect 5.1] and tensor completion [18]. For the data driven methods we set to zero the gray level values inside the known buildings of the map post-processing, thus using the urban geometry data. Without this step, data driven interpolation methods obtain a very poor accuracy since they are not able to recover the sharp building edges. In Fig. 9(b), we plot the average NMSE over

⁷Notice that the run-time is the computation time of the *trained* network. This does not include the training, which is done offline and once for all.

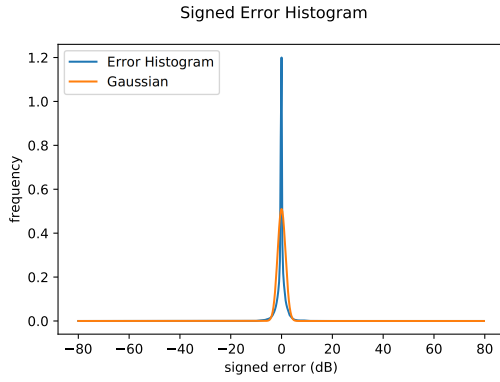


Fig. 7. Blue: the histogram of the difference $f(x) - \tilde{f}(x)$ between the radio map and its approximation in dB in the test set. The histogram is computed for RadioUNet_C on DPM with accurate map. Red: a Gaussian distribution with the same first and second moments.

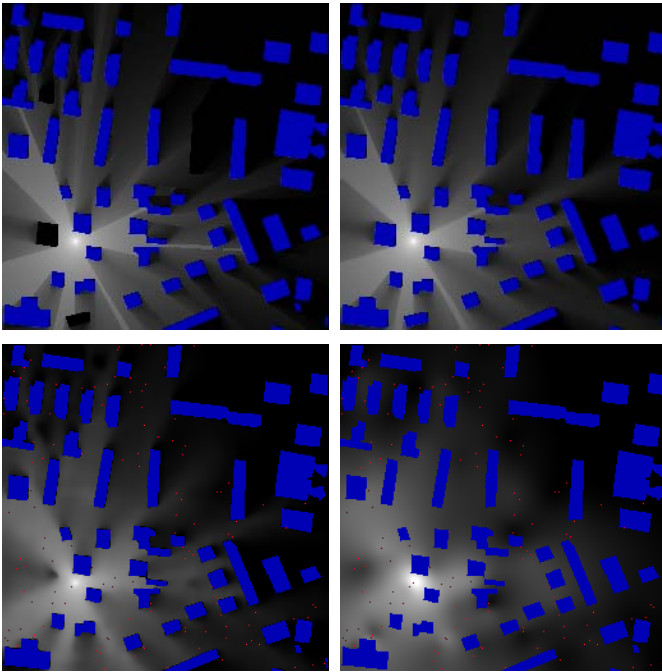


Fig. 8. Comparison of RadioUNet with RBF with four missing buildings in the input. From top-left to bottom-right. 1:Ground truth radio map. 2:RadioUNet_C with all buildings. 3:RadioUNet_S with missing buildings. 4:RBF. The measured 127 locations for both RadioUNet_S and RBF are marked in red. For RBF the transmitter is also a measurement, and the known buildings are set to zero post-processing. Known buildings are marked in blue.

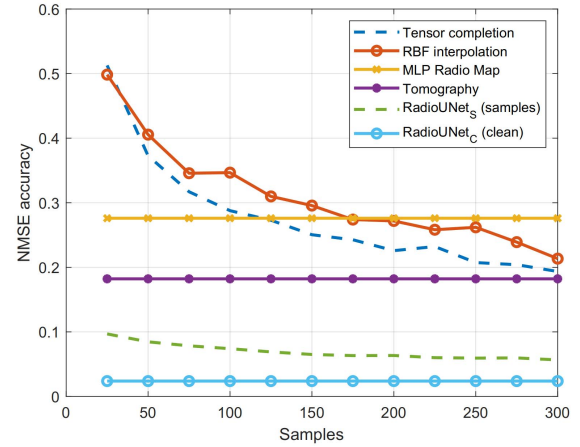
80 Txs of RadioUNet_S and of the two data driven interpolation methods as a function of the number of samples. Both versions of RadioUNet clearly outperform state-of-the-art. Aside from that, RadioUNet, implemented on a CPU, is roughly one order of magnitude faster than RBF interpolation, and three orders of magnitude faster than tensor completion interpolation.

C. Comparison to Model-Based Data Fitting

We compare RadioUNet with a tomography method. In general, tomography methods model the attenuation in the channel strength as the sum of a distance dependent pathloss and a shadowing term which models the attenuation due to obstructions. To model shadowing, a spatial loss field $L : \mathbb{R}^2 \rightarrow \mathbb{R}$ (SLF) is defined. For each spatial location y , the value

Fully Connected Radio Map Network								
Layer	In	1	2	3	4	5	6	7
Neurons	4	64	200	2000	4000	4000	2000	64
								1

(a) Architecture of the MLP of [29].



(b) Estimation error of the radio map reconstruction methods as a function of the number of measurements. We chose Map 12 from the test set, on which RadioUNet performs worse than the average test map (three times the average NMSE). RadioUNet_C, Tomography, and deep learning one-step (MLP), are based on no samples, and are given as horizontal baselines.

Fig. 9. Comparison of RadioUNet with state-of-the-art.

$L(y)$ in a sense models the transparency of y , where $L(y) = 0$ models free space, and $L(y) > 0$ represents a “translucent” obstacle. The shadowing term from the Tx location x to the Rx location y is computed as the integral of L in a narrow oval for which the transmitter and receiver sit on the edges of the largest diameter. More generally, the oval can be replaced by some other shape, which may be trainable.

Note that as opposed to ray-tracing methods, tomography method do not consider at all wave propagation phenomena like diffraction and reflections, and only model the attenuation due to the penetration of the signal through material. For high frequency signals, the attenuation due to penetration in urban environments is very large, which make tomography method less realistic than DPM and IRT.

In tomography methods (e.g., [21]–[24], [59]), the SLF is typically estimated from observed pathloss values between samples transmitter-receiver pairs, by solving an inverse problem. In our situation the problem is easier, since we are given the city map. Thus, the SLF outside the buildings, in free space, is known to be zero. Moreover, the building material is constant, and thus it is natural to consider an SLF with one value f inside buildings, and 0 outside. Hence, the computation of the SLF is reduced to finding the scalar f for which the tomography method gives a radio map as close as possible to the ground truth radio map. This method takes an order of 10^2 sec to run.

D. Comparison to Deep Learning Data Fitting

We compare RadioUNet to the deep learning one-step prediction approach of [29]. We note that the two-step prediction approach of [29] did not perform well in our setting. As explained in Subsection I-C, this method is a data-fitting of a

multilayer perceptron (MLP) to a 4D radio map of a specific city map. The network receives the transmitter and receiver 2D locations and returns the estimation of the pathloss for this pair. The network architecture is reported in Fig. 9(a). For a fixed map, the 80 transmitters are split to 60 training, 10 validation and 10 test transmitters. The network is trained and tested against all receiver locations in the 256×256 grid. This method takes an order of 10sec to estimate all 256×256 pixels, which must be computed separately.

E. Complexity Comparison

Consider a radio map of $n \times n$ pixels. We compare the asymptotic complexity of all considered methods.

- *RadioUNet*: In our UNets, the resolution of Layer l is $O(n^2 4^{-l})$, and the number of channels c_l increase in l slower than 2^l . Moreover, the convolution kernels are spatially localized. This means that the complexity is $O(n^2 \sum_l 4^{-l} c_l c_{l-1}) = O(n^2)$.
- *MLP Regression*: Each estimation takes $O(1)$ operations, and there are n^2 estimations to cover the map. Thus the complexity of $O(n^2)$. We note that in practice the constant here is large, since the MLP must be able to realize a complicated function on a 4D domain.
- *DPM and IRT*: In both methods there is a pre-processing step in which a graph that represents the line of sight between different wall and edge segments, and receiving points, is constructed. For a map with W wall and edge segments. The complexity of this step is $O(W(W + n^2))$. Let us roughly estimate this complexity in terms of n . Since walls are one dimensional, a reasonable estimation of the number of wall pixels in a dense urban environment is An for some $A > 1$. Another reasonable assumption is that, on average, each set of $B > 1$ wall/edge pixels are grouped to one segment. This puts the pre-processing complexity at $O(n^3)$. The pre-processing complexity poses a lower bound to the complexity of both DPM and IRT, regardless of the number of interactions. In IRT, the complexity grows exponentially with respect to the number of interactions (multiplied by n^2). In another version of DPM, preprocessing is not required [60]. In this algorithm, the so called *step 3* dominates complexity, with $O(n^3)$ operations. Indeed, the average path has $O(n)$ pixels, and there are n^2 receiving points.
- *Tomography*: On average, the shape in which the spatial loss field is integrated has proportional area to n^2 . Thus, since all receiver points are computed separately, the complexity is $O(n^4)$ with a small constant.
- *RBF*: For k measurements, RBF takes $O(k^3)$ operations. In typical situations we take k proportional to n , in which case the complexity is $O(n^3)$.
- *Tensor Completion*: In our case we have a matrix completion problem. The dominating term in each iteration is due to SVD within the shrinkage operation [61]. Thus the complexity for T iterations is $O(Tn^3)$.

VI. APPLICATIONS

In this section we demonstrate the usefulness of RadioUNet with two simple applications and also discuss some future applications as future work.

A. Coverage Classification

Service area classification shows up in two situations. In the first problem, given a Tx-Rx link, we would like to know if the received signal strength is large enough. In the second problem, given two Tx-Rx links, we would like to know if the interference caused by one link on the other is low enough. In both cases, the goal is to classify if the pathloss of a certain Tx is above or below some threshold at the location of some Rx. For a fixed Tx location x , let $f(y)$ denote the radio map at location y . We define the *coverage map* as the thresholding function

$$C(y) = \begin{cases} 0 & \text{if } f(y) \leq T, \\ 1 & \text{if } f(y) > T, \end{cases} \quad (6)$$

where T is a threshold in gray scale. For the first problem, depending on the system requirements, T is some value above the noise floor. For example, for high bit rates the signal has to arrive with high SNR, so a typical value for T might be pixel value 0.5 (see e.g., [62]). For the second problem, a typical choice for T is the noise floor, which is pixel value 0.2 for us.

Our goal is to predict the coverage map from the input city and transmitter location. Note that in principle UNets are expressive enough to predict coverage maps, since coverage maps are a sub-phenomenon of radio maps, and UNets are expressive enough to predict radio maps. However, this naive point of view disregards the fact the the gradient descent optimization procedure is highly non-exhaustive, and only searches parameter configurations along a 1D path. As it turns out, simple UNets fail to learn meaningful predictions of coverage maps. Intuitively, radio maps are more predictable than coverage maps since shadow patterns are always associated with simple concepts like building corners and spatial relations between building, receiver locations, and the location of the transmitter. In contrast, in the coverage map most shadow edges disappear and are “absorbed” by one or the domains above or below T .

For the architecture to successfully predict the coverage map, it must first understand the underlying phenomenon of radio maps. We thus consider a WNet architecture, where the first UNet is RadioUNet, and predicts the radio map from the city and transmitter inputs, and the second UNet receives the predicted radio map as input, along with the map and the transmitter location, and computes the coverage map from them. We call the second UNet the thresholding UNet, or TUNet. We call this architecture the Coverage WNet, or CWNNet in short.

To train CWNNet we use curriculum learning. We first train the RadioUNet as before. We then freeze the RadioUNet, and train the TUNet in a curriculum as explained next. As it turns out, the discontinuous nature of the coverage map is still too challenging for the TUNet to learn directly. Instead, we relax the coverage map to a soft coverage map $C_\alpha(y) = \text{sigmoid}(\alpha(f(y) - T))$ where α is a parameter that determines how soft the transition between 0 and one 1 is. We interpret $C_\alpha(y)$ as the probability of location y being in the coverage area. In the curriculum we first train the TUNet to predict

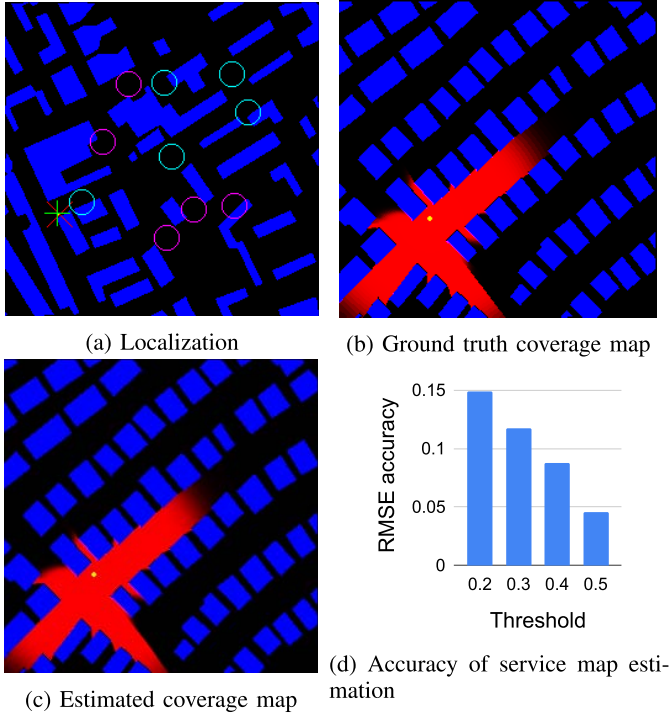


Fig. 10. **Top-left:** Localization result. Green +: true Rx position, red X: estimated Rx position, yellow: pixels of the localization intersection, magenta circles: Tx's of the best localization result out of the R , green circles: the rest of the Tx's. **Top-right and bottom-left:** Coverage map results with threshold 0.5. Red: coverage map. Blue: city map. Yellow: transmitter. **Bottom-right:** accuracy of service map estimation for different thresholds in RMSE.

$C_\alpha(y)$ with $\alpha = 1$, and gradually increase α . We end up with $\alpha = 128$, which we judge to be high enough to represent a sharp transition.

The accuracy of SWNet for different thresholds and an example service map are presented in Fig. 10.

B. Pathloss Based Fingerprint Localization

Suppose that a device is simultaneously in the coverage of several base stations located at Tx points x_1, \dots, x_K , and reports the strengths g_k (converted in gray scale) of their corresponding beacon signals. Let $f_k(y)$ denote the estimated radio map for Tx location $x = x_k$, for $k = 1, \dots, K$. For some $\epsilon > 0$, we define the ϵ -level set for level g_k as

$$L_k^\epsilon = \{z \in \Gamma : |f_k(z) - g_k| \leq \epsilon\}, \quad (7)$$

where Γ is the discrete grid (domain of the radio map, in our case the 256×256 grid). Then, in order to identify the location of the receiver y we can consider the intersection of the ϵ -level sets $S = \bigcap_{k=1}^K L_k^\epsilon$. If this set is localized about a single point, then we have located y with high probability.

Assuming that the reported values $\{g_k\}_k$ are equal to the true radio map values, if for some k the radio map prediction error satisfies $|f_k(y) - g_k| > \epsilon$ then $y \notin L_k^\epsilon$ and y will not be contained in the intersection S . We call such k an outlier. In contrast, if we choose K too small, then S will contain multiple points and the localization is ambiguous. Hence, the method works well when the estimated radio maps are accurate and the number of reported signal strengths K is large enough but not too large.

To alleviate the effect of outliers, instead of computing a single intersection we can select random subsets of $J < K$ Tx's and consider the intersection of the corresponding ϵ -level sets. We also take random ϵ values for each map, since different maps have different unknown accuracies. Repeating this random selection R times, we generate R candidate sets, some of which may be empty and some of which may contain multiple points. For the R' non-empty outcomes we compute a score for the quality of the result, and pick the outcome with the best score. For example, we use the variance of the localization outcome. Let S_t be the localization outcome of sample t , where $t = 1, \dots, R'$. Then, we define the expected position given S_t as $\hat{y}_t = \sum_{z \in S_t} \frac{z}{|S_t|}$, and the associated variance

$$V_t = \sum_{z \in S_t} \frac{|z - \hat{y}_t|^2}{|S_t|},$$

where $|z - \hat{y}|$ is the Euclidean distance between z and \hat{y}_t in \mathbb{R}^2 and $|S_t|$ is the area of S_t . Since smaller variance means better localization, we pick the non-empty localization outcome with smallest variance. In this paper we mention this approach just as an example of the use of accurate radio map estimation. In future work we will deal with improving the pathloss based localization with more sophisticated localization extraction, and using additional signal fingerprints.

In Fig. 10(a) we present an example localization result with $K = 10$, $J = 5$, $R = 5$, $\epsilon = 0.03$. The best outcome has a standard deviation of 0.5 meters. The distance between the estimated and true receiver location is 1.58 meters.

VII. CONCLUSION

In this paper we introduced RadioUNet, a deep learning method for simulating radio maps given a city geometry, Tx location, and optionally some pathloss measurements and car locations. For training RadioUNet, we introduced the new dataset RadioMapSeer, which we hope will be used for developing deep learning methods for pathloss prediction by other researchers as well. We developed approaches for transferring what was learned on the large dataset of coarsely simulated radio maps to real-life, and demonstrated the superior performance of our methods with respect to state-of-the-art, both in run-time and accuracy.

ACKNOWLEDGMENT

The authors would like to thank Ibrahim Rashdan for a fruitful discussion on the impact of cars on the pathloss function.

REFERENCES

- [1] C. Geng, N. Naderializadeh, A. S. Avestimehr, and S. A. Jafar, "On the optimality of treating interference as noise," *IEEE Trans. Inf. Theory*, vol. 61, no. 4, pp. 1753–1767, Apr. 2015.
- [2] G. Bianchi, "Performance analysis of the IEEE 802.11 distributed coordination function," *IEEE J. Sel. Areas Commun.*, vol. 18, no. 3, pp. 535–547, Mar. 2000.
- [3] X. Wu *et al.*, "FlashLinQ: A synchronous distributed scheduler for peer-to-peer ad hoc networks," *IEEE/ACM Trans. Netw.*, vol. 21, no. 4, pp. 1215–1228, Aug. 2013.
- [4] N. Naderializadeh and A. S. Avestimehr, "ITLinQ: A new approach for spectrum sharing in device-to-device communication systems," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1139–1151, Jun. 2014.

- [5] X. Yi and G. Caire, "Optimality of treating interference as noise: A combinatorial perspective," *IEEE Trans. Inf. Theory*, vol. 62, no. 8, pp. 4654–4673, Aug. 2016.
- [6] K. Shen and W. Yu, "FPLINQ: A cooperative spectrum sharing strategy for device-to-device communications," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Aachen, Germany, Jun. 2017, pp. 2323–2327.
- [7] W. Cui, K. Shen, and W. Yu, "Spatial deep learning for wireless scheduling," *IEEE J. Sel. Areas Commun.*, vol. 37, no. 6, pp. 1248–1261, Jun. 2019.
- [8] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "Optimal user-cell association for massive MIMO wireless networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1835–1850, Mar. 2016.
- [9] S. Nikitaki, G. Tsagkatakis, and P. Tsakalides, "Efficient training for fingerprint based positioning using matrix completion," in *Proc. Eur. Signal Process. Conf. (EUSIPCO)*, Bucharest, Romania, Aug. 2012, pp. 195–199.
- [10] Z. Utkovski, P. Agostini, M. Frey, I. Bjelakovic, and S. Stanczak, "Learning radio maps for physical-layer security in the radio access," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Cannes, France, Jul. 2019, pp. 1–5.
- [11] T. V. Chien, T. N. Canh, E. Björnson, and E. G. Larsson, "Power control in cellular massive MIMO with varying user activity: A deep learning solution," *IEEE Trans. Wireless Commun.*, vol. 19, no. 9, pp. 5732–5748, 2020.
- [12] S. Park, A. Q. Truong, and T. H. Nguyen, "Power control for sum spectral efficiency optimization in MIMO-NOMA systems with linear beamforming," *IEEE Access*, vol. 7, pp. 10593–10605, 2019.
- [13] A. Ashikhmin and T. Marzetta, "Pilot contamination precoding in multi-cell large scale antenna systems," in *Proc. IEEE Int. Symp. Inf. Theory*, Cambridge, MA, USA, Jul. 2012, pp. 1137–1141.
- [14] S. Zhang and R. Zhang, "Radio map based path planning for cellular-connected UAV," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Waikoloa, HI, USA, Dec. 2019, pp. 1–6.
- [15] Z. Chen, F. Söhrabi, and W. Yu, "Sparse activity detection for massive connectivity," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1890–1904, Apr. 2018.
- [16] M. L. Stein, *Interpolation of Spatial Data: Some Theory for Kriging*. New York, NY, USA: Springer, 1999.
- [17] C. M. Bishop, *Neural Networks for Pattern Recognition*. New York, NY, USA: Oxford Univ. Press, 1995.
- [18] D. Schaufele, R. L. G. Cavalcante, and S. Stanczak, "Tensor completion for radio map reconstruction using low rank and smoothness," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Cannes, France, Jul. 2019, pp. 1–5.
- [19] R. Timoteo, D. Cunha, and G. Cavalcanti, "A proposal for path loss prediction in urban environments using support vector regression," in *Proc. Adv. Int. Conf. Telecommun. (AICT)*, Paris, France, Jul. 2014, pp. 119–124.
- [20] S. Chouvardas, S. Valentin, M. Draief, and M. Leconte, "A method to reconstruct coverage loss maps based on matrix completion and adaptive sampling," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Shanghai, China, Mar. 2016, pp. 6390–6394.
- [21] D. Lee, S.-J. Kim, and G. B. Giannakis, "Channel gain cartography for cognitive radios leveraging low rank and sparsity," *IEEE Trans. Wireless Commun.*, vol. 16, no. 9, pp. 5953–5966, Sep. 2017.
- [22] D. Romero, D. Lee, and G. B. Giannakis, "Blind radio tomography," *IEEE Trans. Signal Process.*, vol. 66, no. 8, pp. 2055–2069, Apr. 2018.
- [23] D. Lee, D. Berberidis, and G. B. Giannakis, "Adaptive Bayesian radio tomography," *IEEE Trans. Signal Process.*, vol. 67, no. 8, pp. 1964–1977, Apr. 2019.
- [24] D. Lee and G. B. Giannakis, "A variational bayes approach to adaptive channel-gain cartography," in *Proc. ICASSP-IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Brighton, U.K., May 2019, pp. 8434–8438.
- [25] K. Rizk, J.-F. Wagen, and F. Gardiol, "Two-dimensional ray-tracing modeling for propagation prediction in microcellular environments," *IEEE Trans. Veh. Technol.*, vol. 46, no. 2, pp. 508–518, May 1997.
- [26] R. Wahl, G. Wölfl, P. Wildbolz, and F. Landstorfer, "Dominant path prediction model for urban scenarios," in *Proc. IST Mobile Wireless Commun.*, 2005, pp. 1–5.
- [27] T. Zugno, M. Drago, M. Giordani, M. Polese, and M. Zorzi, "Towards standardization of millimeter wave Vehicle-to-Vehicle networks: Open challenges and performance evaluation," 2019, *arXiv:1910.00300*. [Online]. Available: <http://arxiv.org/abs/1910.00300>
- [28] T. Imai, K. Kitao, and M. Inomata, "Radio propagation prediction model using convolutional neural networks by deep learning," in *Proc. Eur. Conf. Antennas Propag. (EuCAP)*, Krakow, Poland, Mar. 2019, pp. 1–5.
- [29] K. Saito, Y. Jin, C. Kang, J.-I. Takada, and J.-S. Leu, "Two-step path loss prediction by artificial neural network for wireless service area planning," *IEICE Commun. Exp.*, vol. 8, no. 12, pp. 611–616, 2019.
- [30] S. I. Popoola *et al.*, "Determination of neural network parameters for path loss prediction in very high frequency wireless channel," *IEEE Access*, vol. 7, pp. 150462–150483, 2019.
- [31] S. P. Sotirioudis and K. Siakavara, "Mobile radio propagation path loss prediction using artificial neural networks with optimal input information for urban environments," *AEU-Int. J. Electron. Commun.*, vol. 69, no. 10, pp. 1453–1463, Oct. 2015.
- [32] S. P. Sotirioudis, S. K. Goudos, K. A. Gotsis, K. Siakavara, and J. N. Sahalos, "Application of a composite differential evolution algorithm in optimal neural network design for propagation path-loss prediction in mobile communication systems," *IEEE Antennas Wireless Propag. Lett.*, vol. 12, pp. 364–367, 2013.
- [33] I. Povescu, I. Nafomita, P. Constantinou, A. Kanatas, and N. Moraitis, "Neural networks applications for the prediction of propagation path loss in urban environments," in *Proc. IEEE VTS 53rd Veh. Technol. Conf., Spring*, vol. 1, May 2001, pp. 387–391.
- [34] C. Parera, Q. Liao, I. Malanchini, C. Tatino, A. E. C. Redondi, and M. Cesana, "Transfer learning for tilt-dependent radio map prediction," *IEEE Trans. Cognit. Commun. Netw.*, vol. 6, no. 2, pp. 829–843, Jun. 2020.
- [35] Y. Teganya and D. Romero, "Data-driven spectrum cartography via deep completion autoencoders," in *Proc. ICC-IEEE Int. Conf. Commun. (ICC)*, Dublin, Ireland, Jun. 2020, pp. 1–7.
- [36] 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 *Proc. ICASSP-IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2020, pp. 8678–8682.
- [37] 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, Switzerland: Springer, 2015, pp. 234–241.
- [38] R. Hoppe, G. Wölfl, and U. Jakobus, "Wave propagation and radio network planning software WinProp added to the electromagnetic solver package FEKO," in *Proc. Int. Appl. Comput. Electromagn. Soc. Symp. Italy (ACES)*, Florence, Italy, Mar. 2017, pp. 1–2.
- [39] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 512, pp. 436–444, May 2015.
- [40] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Nov. 1998.
- [41] E. Shelhamer, J. Long, and T. Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 4, pp. 640–651, Apr. 2017.
- [42] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [43] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 4th Int. Conf. 3D Vis. (3DV)*, Stanford, CA, USA, Oct. 2016, pp. 565–571.
- [44] O. Çiçek, A. Abdulkadir, S. S. Lienkamp, T. Brox, and O. Ronneberger, "3D U-Net: Learning dense volumetric segmentation from sparse annotation," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI 2016*. Cham, Switzerland: Springer, 2016, pp. 424–432.
- [45] M. Mathieu, C. Couprie, and Y. LeCun, "Deep multi-scale video prediction beyond mean square error," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, San Juan, Puerto Rico, May 2016, pp. 1–14.
- [46] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Honolulu, HI, USA, Jul. 2017, pp. 1132–1140.
- [47] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep convolutional neural network for inverse problems in imaging," *IEEE Trans. Image Process.*, vol. 26, no. 9, pp. 4509–4522, Sep. 2017.
- [48] Z. Yi, H. Zhang, P. Tan, and M. Gong, "DualGAN: Unsupervised dual learning for Image-to-Image translation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2868–2876.

- [49] G. Litjens *et al.*, “A survey on deep learning in medical image analysis,” *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
- [50] Y. Bengio, J. Louradour, R. Collobert, and J. Weston, “Curriculum learning,” in *Proc. 26th Annu. Int. Conf. Mach. Learn.* New York, NY, USA: Association Computing Machinery, 2009, pp. 41–48.
- [51] S. J. Pan, I. W. Tsang, J. T. Kwok, and Q. Yang, “Domain adaptation via transfer component analysis,” *IEEE Trans. Neural Netw.*, vol. 22, no. 2, pp. 199–210, Feb. 2011.
- [52] K. Weiss, T. M. Khoshgoftaar, and D. Wang, “A survey of transfer learning,” *J. Big Data*, vol. 3, no. 1, pp. 1–40, Dec. 2016.
- [53] T. Rautiainen, G. Wölflé, and R. Hoppe, “Verifying path loss and delay spread predictions of a 3d ray tracing propagation model in urban environment,” in *Proc. IEEE 56th Veh. Technol. Conf.*, Vancouver, BC, Canada, vol. 4, Dec. 2002, pp. 2470–2474.
- [54] OpenStreetMap contributors. (2017). *Planet Dump*. [Online]. Available: <https://planet.osm.org> and <https://www.openstreetmap.org>
- [55] T. M. Cover, *Elements of Information Theory*. Hoboken, NJ, USA: Wiley, 1999.
- [56] E. Biglieri, J. Proakis, and S. Shamai, “Fading channels: Information-theoretic and communications aspects,” *IEEE Trans. Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [57] A. Hauptmann and J. Adler, “On the unreasonable effectiveness of CNNs,” 2020, *arXiv:2007.14745*. [Online]. Available: <http://arxiv.org/abs/2007.14745>
- [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, pp. 1–15.
- [59] M. A. Gutierrez-Estevez, R. L. G. Cavalcante, and S. Stanczak, “Non-parametric radio maps reconstruction via elastic net regularization with multi-kernels,” in *Proc. IEEE 19th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Kalamata, Greece, Jun. 2018, pp. 1–5.
- [60] G. Wölflé and F. M. Landstorfer, “Field strength prediction with dominant paths and neural networks,” in *Proc. MIOP*, Sindelfingen, Germany, Apr. 1997, pp. 216–220.
- [61] S. Gandy, B. Recht, and I. Yamada, “Tensor completion and low-n-rank tensor recovery via convex optimization,” *Inverse Problems*, vol. 27, no. 2, Jan. 2011, Art. no. 025010.
- [62] *Intelligent Transport Systems (ITS); Access Layer Specification for Intelligent Transport Systems Operating in the 5 GHz Frequency Band*, Standard EN 302 663 V1.2.1, ETSI, 7 2013.



Ron Levie received the Ph.D. degree in applied mathematics from Tel Aviv University, in 2018. From 2018 to 2020, he was a Post-Doctoral Researcher with the Research Group Applied Functional Analysis, Institute of Mathematics, Technische Universität Berlin, Germany. Since 2021, he has been a Researcher with the Bavarian AI Chair for Mathematical Foundations of Artificial Intelligence, Department of Mathematics, Ludwig Maximilian University of Munich, Germany. His current research interests include mathematical theory and applica-

tion of deep learning, applied harmonic and functional analysis, image processing, and applications to telecommunication. He won excellence awards for his M.Sc. and Ph.D. studies, and a Post-Doctoral Researcher Minerva Fellowship.



Çağkan Yapar (Graduate Student Member, IEEE) received the B.Sc. degree in electrical and electronics engineering from Boğaziçi University, in 2012, and the M.Sc. degree in electrical engineering and information technology from TU München, in 2015. He is currently a Doctoral Researcher with the Communications and Information Theory Chair, TU Berlin, Germany. His research interests include information theory, communications, machine learning, and signal processing.



Gitta Kutyniok (Senior Member, IEEE) received the Diploma degree in mathematics and computer science and the Ph.D. degree from Universität Paderborn, Germany, and the Habilitation degree in mathematics from the Justus-Liebig Universität Gießen, in 2006.

From 2001 to 2008, she held visiting positions at several U.S. institutions, including Princeton University, Stanford University, Yale University, Georgia Institute of Technology, and Washington University in St. Louis, and was a Nachdiplomlecturer with ETH Zurich in 2014. In 2008, she became a Full Professor of mathematics with the Universität Osnabrück, and moved to Berlin three years later, where she held the Einstein Chair of the Institute of Mathematics, Technische Universität Berlin and a courtesy appointment with the Department of Computer Science and Engineering until 2020. She is currently the Bavarian AI Chair of Mathematical Foundations of Artificial Intelligence, Ludwig-Maximilians Universität München, and an Adjunct Professorship in machine learning with the University of Tromsø. Her main research interests include in the areas of applied harmonic analysis, artificial intelligence, compressed sensing, high-dimensional data analysis, imaging science, inverse problems, machine learning, partial differential equations, and applications to life sciences and telecommunication.

Dr. Kutyniok became a member of the Berlin-Brandenburg Academy of Sciences and Humanities in 2017 and SIAM Fellow. She received various awards for her research, such as an Award from the Universität Paderborn in 2003, the Research Prize of Gießen and a Heisenberg-Fellowship in 2006, the von Kaven Prize by the DFG in 2007, and an Einstein Chair in 2008. She gave the Noether Lecture at the ÖMG-DMV Congress in 2013 and the Hans Schneider ILAS Lecture at IWOTA in 2016. She received the Francqui Chair of the Belgian Francqui Foundation in 2020, and has been holding the first Bavarian AI Chair at LMU since 2020. She was Chair of the SIAM Activity Group on Imaging Sciences from 2018 to 2019 and the Co-Chair of the first SIAM conference on Mathematics of DataScience in 2020. She also serves as the Founding Vice-Chair of the new SIAM Activity Group on Data Science. She was the Scientific Director of the Graduate School BIMoS, TU Berlin, from 2014 to 2020. She is currently the Chair of the GAMM Activity Groups on Mathematical Signal- and Image Processing and Computational and Mathematical Methods in Data Science. She is also the main Coordinator of the Priority Programm of the German Research Foundation on theoretical foundations of deep learning.



Giuseppe Caire (Fellow, IEEE) was born in Torino, in 1965. He received the B.Sc. degree in electrical engineering from the Politecnico di Torino, in 1990, the M.Sc. degree in electrical engineering from Princeton University, in 1992, and the Ph.D. degree from the Politecnico di Torino, in 1994.

He was a Post-Doctoral Research Fellow with the European Space Agency (ESTEC, Noordwijk, The Netherlands) from 1994 to 1995, an Assistant Professor in telecommunications with the Politecnico di Torino, an Associate Professor with the University of Parma, Italy, a Professor with the Department of Mobile Communications, Eurecom Institute, Sophia-Antipolis, France, a Professor of Electrical Engineering with the Viterbi School of Engineering, University of Southern California, Los Angeles, CA, USA. He is currently an Alexander von Humboldt Professor with the Faculty of Electrical Engineering and Computer Science, Technical University of Berlin, Germany. His main research interests include in the field of communications theory, information theory, channel, and source coding with particular focus on wireless communications. He received the Jack Neubauer Best System Paper Award from the IEEE Vehicular Technology Society in 2003, the IEEE Communications Society and Information Theory Society Joint Paper Award in 2004 and in 2011, the Okawa Research Award in 2006, the Alexander von Humboldt Professorship in 2014, the Vodafone Innovation Prize in 2015, an ERC Advanced Grant in 2018, the Leonard G. Abraham Prize for Best IEEE JSAC Paper in 2019, and the IEEE Communications Society Edwin Howard Armstrong Achievement Award in 2020. He was a recipient of the 2021 Leibniz Prize of the German National Science Foundation (DFG). He has served in the Board of Governors of the IEEE Information Theory Society from 2004 to 2007, and as officer from 2008 to 2013. He was the President of the IEEE Information Theory Society in 2011.