IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 11, 2019, accepted December 24, 2019, date of publication January 6, 2020, date of current version January 15, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2964103

Model-Aided Deep Learning Method for Path Loss Prediction in Mobile Communication Systems at 2.6 GHz

JAKOB THRANE, (Student Member, IEEE), DARKO ZIBAR, (Member, IEEE), AND HENRIK LEHRMANN CHRISTIANSEN, (Member, IEEE)
Department of Photonics Engineering, Technical University of Denmark, 2800 Kongens Lyngby, Denmark

Corresponding author: Jakob Thrane (jathr@fotonik.dtu.dk)

This work was supported by the Technical University of Denmark.

ABSTRACT Accurate channel models are essential to evaluate mobile communication system performance and optimize coverage for existing deployments. The introduction of various transmission frequencies for 5G imposes new challenges for accurate radio performance prediction. This paper compares traditional channel models to a channel model obtained using Deep Learning (DL)-techniques utilizing satellite images aided by a simple path loss model. Experimental measurements are gathered and compose the training and test set. This paper considers path loss modelling techniques offered by state-of-the-art stochastic models and a ray-tracing model for comparison and evaluation. The results show that 1) the satellite images offer an increase in predictive performance by $\approx$ 0.8 dB, 2) The model-aided technique offers an improvement of $\approx$ 1 dB, and 3) that the proposed DL model is capable of improving path loss prediction at unseen locations for 811 MHz with $\approx$ 1 dB and $\approx$ 4.7 dB for 2630 MHz.

INDEX TERMS 5G mobile communication, channel models, wireless communication, computer vision, machine learning, supervised learning.

I. INTRODUCTION

The fifth generation of mobile networks, 5G, seeks to expand the current mobile architecture with densification of base stations, also known as Heterogeneous UltraDense Network (H-UDN), to offer improved capacity and coverage for users. The densification results in low inter-site distances between terminals and base stations. Such a decrease in distance allows for improved radio conditions when operating at higher frequencies. For instance, Millimeter Waves (mmWaves) is expected to be an essential part of New Radio (NR) due to the large quantity of available spectrum but suffers over longer distances due to the increased path loss. The densification requires a change to the classic cellular architecture of having macrocells for coverage and capacity. A *Heterogeneous* mindset is set to replace the classical mindset where smaller base stations such as micro, pico, and even femtocells [1], [2] manage user data and the macrocells manage control signals and wide-area coverage. However, this approach poses a significant challenge in terms of network management and deployment strategies [3]. An essential element of planning and deploying mobile communication systems is the modelling of signal propagation and losses thereof. The task of such propagation models is two-fold, 1. to predict propagation for proposed deployment scenarios allowing evaluation before practically deploying such solutions and 2. be capable of improving the coverage and capacity of existing systems by dealing with so-called coverage holes.

Through recent years, as more advanced transmission technologies have made their way into mobile communication systems, more detailed channel modelling techniques have been required. While stochastic models are computationally simple and fast while keeping satisfactory accuracy, margins limit the predictive capabilities [4]. The rigid format of stochastic models causes coarse and inaccurate propagation prediction for scenarios different from the originated measurements. Such predictive inaccuracies result in calibration studies that seek to refit the empirically obtained models to different propagation scenarios. Ray-tracing is another method for accurately predicting the channel state given a propagation scenario. Ray-tracing, however, has one big

The associate editor coordinating the review of this manuscript and approving it for publication was Alberto Cano.

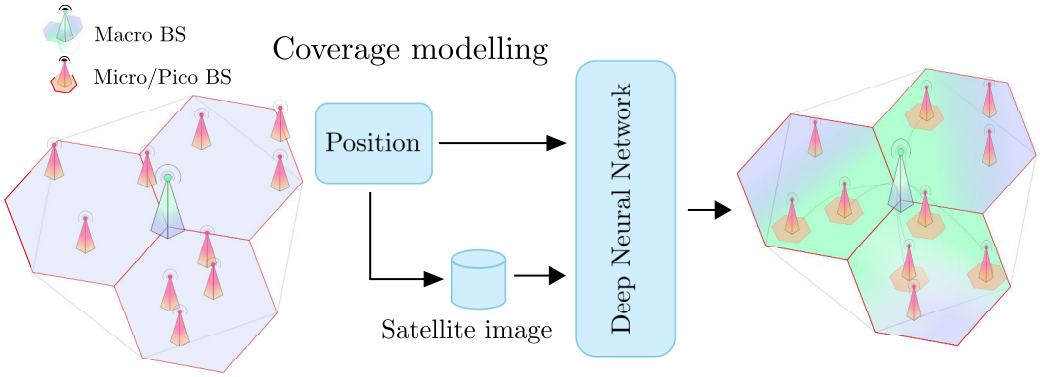


FIGURE 1. A complex coverage situation requires complex channel models for coverage prediction. A Deep Neural Network is proposed utilizing satellite images and position indicators for improved path loss prediction. The learned model can be queried using a satellite image and a position for the expected received power.

draw-back it is computationally expensive and geographical data exhaustive [5].

In the context of cellular planning and optimization, the purpose and application of channel modelling are vast. The complexity of channel models is highly relevant to the specific planning phase. For instance, in greenfield deployments, simple empirical path loss models are used. In urban settings where coverage already exists, the deployment and interference management increases in complexity, thus more advanced channel models are required. *Accurate channel models are essential to evaluate future generation mobile communication systems.* Channel models have stringent requirements and must consider a diverse selection of propagation scenarios to limit the need for time-consuming calibration and measurement studies.

Furthermore, cognitive networking is considered a necessary and essential element of future solutions [6]. Cognitive networking is expected to set forward tight requirements on channel models, not only in terms of accuracy but also in terms of computational performance. Thus, new channel models that can offer improvements in both aspects are of great interest. The primary focus of this work is the mean received power under slow fading impairments such as shadowing. Thus, the focus is on improving channel models for coverage modelling, where accurate link-budgets are essential.

A. CONTRIBUTIONS

In [7], Deep Learning (DL) is shown to be capable of inferring radio quality parameters using satellite images and can offer a metric of uncertainty using Bayesian approximation. The aim and novelty of this paper are to investigate the proposed DL method for radio propagation prediction and compare them to existing methodologies. More specifically:

- We propose an improved model for path loss prediction for use in mobile communication systems based on a DL framework utilizing satellite imagery and position indicators.
- We show how the proposed DL framework is capable of inferring features from satellite images and thereby produce improved prediction even in new/unseen environments.
- We compare the prediction of path loss at unseen locations for the proposed DL method and state-of-the-art channel models.

Stochastic models use an empirical-based path loss model. Such models are obtained and based on interpolation of experimental measurements. Thus they are regression methods that seek to predict a continuous value of path loss for a given distance of transmission.

The DL proposed is also of the type regression. However, we show that by using automated feature extraction of satellite images aided by a simple path loss model, we can approximate the large-scale attenuation by learned latent continuous variables. Furthermore, this achieves low data complexity as compared to traditional ray-tracing methods. It is thus interesting to compare the DL based models to traditional channel models such as empirical-based and ray-tracing-based.

B. STATE-OF-THE-ART

Mean path loss prediction has been subject to much research, both in terms of accuracy but also complexity. It is relevant to mention the latest empirical-based path loss models, as documented in 3GPP TR 38.901 and ITU-R M. 2412 [8], [9]. Single-slope distance path loss models characterize both models, and the Large-Scale Parameters (LSPs) are determined by a Gaussian distribution, thus making them simple and computationally fast.

To combat the rigid margins of the stochastic models, Neural Network (NN) have been successfully demonstrated to offer high performance and low complexity alternative to predicting path loss for wireless communication systems. The authors in [10] use inputs such as antenna separation distance, antenna height at both the transmitter and receiver, the clearance of the terrain, and the angle hereof and vegetation information. Most of such inputs are the result of features engineering by, e.g., inspecting satellite images. The authors in [11] document the use of NN with inputs such as clutter height, at 1800 MHz, and demonstrate predicting in routes different and not included in the training set. However,

it remains challenging to determine how inherently different the routes used for testing are from the routes included in the training set. Thus it is difficult to conclude that sufficient generalization is achieved. It is reasonable to assume that extrapolation is partly achieved for unknown propagation situations, which is not an ideal outcome. The primary reason for this is the problematic aspect of quantifying parameters (features) that are generalizable for radio propagation in the majority of propagation scenarios. In other words, which features should be engineered or selected to model radio propagation impairments effectively. Several features have a significant impact on the overall mean of path loss - for instance, the height of the transmitter and the receiver and the antenna separation. However, also impactful are large-scale fading impairments caused by vegetation and buildings. Such impairments are tricky to represent, i.e., engineer features for since they require statistical knowledge of the propagation scenario. For example, a feature to engineer could be the distance to the closest building. A comprehensive comparison of features and engineered features for NNs can be found in [12]. For frequencies below 1 GHz, features related primarily to distance achieve the best performing NN. For frequencies above 1 GHz, building and clutter related features have a significant impact on the predictive capabilities of the NN, related to the use of shorter wavelengths. It is well documented that the performance of adaptive models such as NN are limited by the features used [12], [13]. It is thus relevant to investigate approaches for feature engineering to improve predictive performance. Achieving generalization using such approaches is essential and paramount; otherwise, the trained models have little purpose. In this paper, we look at Deep Learning-based methodologies for achieving such properties.

DL has accelerated the field of Machine Learning (ML) and has offered significant improvements to model accuracy. The availability of raw data enables automated feature engineering through deep and layered network structures. Several authors have documented utilizing DL in Wireless communication and should be highlighted [14]–[16], however, to the best of the author's knowledge no one has utilized DL for propagation modelling using satellite imagery as we propose in this work.

C. PAPER OUTLINE

In Section II we introduce the basic principles of path loss modelling. Additionally, we detail how a Deep Neural Network (DNN) is used to learn a mapping function between input features, such as distance, position, and satellite images, to a received signal strength parameter. Finally, the proposed model architecture is described. In Section III, the experimental setup is detailed along with the structure of the finalized data set. Training of the DNN model and the resulting experiments are detailed in Section IV along with the best-found hyper-parameters. Additionally, in Section IV, traditional channel modelling methods used for validation and evaluation are explained. Results are presented in Section V, and the work is discussed in Section VI. Finally, a conclusion is presented in VII

NOTATION

G_{tx}	Transmission power and gain
PL(d)	Path loss
$L(d)$	Link budget
X_0	Gaussian noise with zero mean
σ	Local variability
$\mathbf{w}$	Vector of adaptive weights
θ	Hyperparameters
ϵ	Observation noise
B_{tx}	Transmitter indicator
$d_{lat,lon}$	Distance in latitude and longitude
d	3D Distance
$\mathbf{A}$	Satellite Image as $W \times H \times C$ matrix
x_n	Single input vector
$\mathbf{x}$	Vector of input vectors
t_n	Single target/observation
$\mathbf{t}$	Vector of targets/observations

II. DEEP LEARNING FOR PATH LOSS PREDICTION

The curve-fit provided by traditional mean path loss models are based on a theory of cause and effect, meaning the intent is to find the best parameters that can explain the observations. For instance, research has shown a single-slope distance model (Eq. 1) is capable of describing mean path loss satisfactory [17].

$$PL(d) = A + B \log 10(d) + X_0 \tag{1}$$

Some values of A and B as a function of features (such as frequency, transmitter height) have been found using interpolation of obtained measurements. Additionally, LSP can be modeled by a Gaussian distribution X_0 (log-normal) with mean zero and some σ denoting the *local variability*. Such a path loss model does not consider fast variations and assumes an average over samples obtained over a route with a length equal to tens of wavelengths [18]. This approach is a contrast to ML where the model is unknown. Machine Learning (ML) provides an extensive toolbox of adaptive models that are capable of learning representations and mapping functions. Thus the goal of ML is not only to discover the best parameters, but also the best model $f(\cdot)$. In this case, the best model that represents the path loss.

Supervised ML has the objective of mapping and learning representations between features and observations. For instance, how input relates to the outputs of a system. Discovering essential features can be difficult, and most times, it is difficult to formalize features. Such difficulties especially arise when considering the geo-statistics of the propagation area. A feature that adds useful information about buildings and their height for predicting path loss, how would it best be formalized? Would it be average building height in a straight line between the transmitter and receiver? Such a feature, in an urban scenario, would possibly be useless since multipath fading and scattering are

dominant [19]. So maybe the distance to the nearest building is more informative. In short, it can be challenging to engineer useful and representative features, which is why generalization of path loss prediction is an issue. *Performing regression-based prediction of path loss can turn into a series of extrapolation models since the geostatistics are difficult and time-consuming to formalize.*

DL provides with automated feature engineering through complex and *deep*-layered structures. In this work, features are learned from satellite images that can aid in predicting the received power of mobile communication systems. Geographical coordinates and a simple path loss model assist the model in the learning process. The selection of inputs are based on the following intuition and hypothesis:

- *Local coordinates* - contains scenario-specific information. Assists the model in deducing local propagation characteristics, like bearing of the transmitter and interference sources.
- *Satellite images* - contains information of local variability, e.g. large-scale fading impairments. Must be high enough resolution to be able to observe buildings, vegetation, and other structures that influence the magnitude of local variability.

A so-called Convoluted Neural Network (CNN) is used to obtain useful features from satellite images. Such networks use *convolutions* instead of regular multiplications. The aim is to obtain a regression model that can predict signal strength continuously; thus, the model is formalized as follows:

$$t_n = y(x_n, \mathbf{w}, \theta) + \epsilon \tag{2}$$

where y is the function to learn, x is the input, $\mathbf{w}$ are the adaptive weights, θ are the model hyper-parameters and ϵ is Gaussian distributed noise. We measure the LTE-A reference parameter, Reference Signal Received Power (RSRP), the observation is $t_n = \text{RSRP}$.

Different models and approaches can be used to learn such a function y. The model hyper-parameters (related to the complexity of the model) are tuned to the problem and the data. In this work, we use methodologies related to NN. Thus a hyper-parameter could, for instance, be the number of layers used.

We furthermore define the input to the model as follows:

$$x_n = [\text{lat}, \text{lon}, \text{B}_{tx}, \text{d}_{lat}, \text{d}_{lon}, d, A] \tag{3}$$

where lat, lon are the geographical coordinates of the receiver. B_{tx} indicates the transmitter (the dataset contains several transmitters). d_{lat}, d_{lon} denote the distance in the latitude and longitude direction respectively between the transmitter and the receiver. d denote the distance straight as the crow flies. All distance metrics are features and are computed based on the coordinates. A denote the corresponding satellite image. Thus the model is tasked with learning RSRP as a function of positions, distance, and satellite images.

A. DEEP NEURAL NETWORKS

NN, and in particular DNN, has proven to be useful in many nonlinear mapping problems [20]. NNs is a linear combination of basis functions which are transformed using nonlinear activation functions. A two-layered NN has the following form:

$$y_k(x, \mathbf{w}) = \sum_{j=1}^{M} w_{kj}^{(2)} h\left(\sum_{i=1}^{D} w_{ji}^{(1)} x_i + w_{j0}^{(1)} \right) + w_{k0}^{(1)} \tag{4}$$

where y_k is considered the k'th output. $\mathbf{w}$ are considered the adaptive weights for both layers with size M and D respectively, denoted (2) and (1). $h(\cdot)$ indicates the nonlinear activation function. In this work, the *Rectified Linear Unit (ReLU)* activation function is used. [21], [22].

A training set of inputs and observations are used to find the adaptive weights. The observations are denoted t_n while the inputs x_n. Given the training set, we thus seek to minimize the error provided by the sum-of-squares error function between the model output, given a set of weights, and the observations.

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^{N} ||y_n(x_n, \mathbf{w}) - t_n||^2 \tag{5}$$

Minimizing the above cost function corresponds to maximizing the likelihood function given the targets have noise that is Gaussian distributed. (E.g. ϵ in Eq. 2 is Gaussian distributed.) This is a fair assumption given the distribution of slow/large-scale variations (such as shadowing) are Gaussian distributed [18]. Minimizing the cost function is done iteratively with the use of the error backpropagation algorithm. In this work, the well known *Adam* optimizer is used [20]. Mini-batch training is furthermore used to compute the gradient of the cost function with respect to the weights for several data points. To combat over- and underfitting, the model, is regularized using standard principles such as batch normalization, dropout layers, and weight decay [21].

B. IMAGE PROCESSING

In order to effectively deal with satellite images, convolutional layers in the NN are used. The operation of convolutions in combination with NN structures has revolutionized image processing and gains hereof [23]. The operation of convolution can be defined as

$$s(t) = \int x(a) w(t - a) da \tag{6}$$

In the context of CNN x is the input and w is the kernel. The kernel is an adaptive filter that the network learns. Thus the layer can be seen as a filter bank with adaptive taps. If the training set contains representative examples, e.g., the image is related to signal attenuation, the cascaded architecture of these layers will result in useful latent features at the final layer. In this work, $i \in 0, 1, ..3$ layers are used. Each layer d_i consists of several basic operations used in CNN. As illustrated in Fig. 2 a structure of 2D convolutions, nonlinear transformations (ReLU), batch normalization and maximum

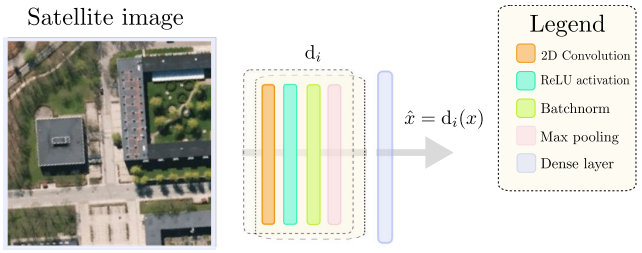


FIGURE 2. A cascaded structure of convolutional layers are used d_i. Each layer consists of several basic operations, such as ReLU activation, Batch normalization and max pooling. A linear set of weights is connected to the output of d_3.

pooling is sequentially used. A linear set of weights is added as an output layer. The architecture and the depth of the layers, and the kernel sizes used are detailed in Section IV.

C. TRAINING ARCHITECTURE

NN and DNN are universal approximators, and can be shown to be capable of approximating any continuous function in $\mathbb{R}^n$ [21]. The challenge remains in tuning such models and their hyper-parameters to achieve state-of-the-art performance. The search for such hyper-parameters is computationally expensive and is a major bottleneck in the training of deep models. The idea of introducing a *residual path loss* into the proposed model is similar to the work carried out in [24] where increased performance is observed training the model not only on the data but also using a simple physics model to assist in learning. The intuition here is that the model is tasked with learning the correction of the observations to the simple physics model. Such a learning task is easier and simpler than learning all possible explanations for all observations. It has been shown in [25] that expert knowledge (model-aided learning) can be embedded into wireless systems for optimization. Additionally, such hybrid-based models have shown to be effective for path loss prediction as documented in [26].

In this work, the proposed DL model uses a simple path loss model for assisting in the learning process. More specifically, we define the output of the simple path loss model as an estimated link budget, thus $L(d) = PL(d) + G_{tx}$, where G_{tx} is the estimated transmission power and related gain. The UMa_B model is used as a path loss model, and the link budget estimate is given as an input to the DNN. Furthermore, it is added to the output of the DNN, as illustrated in Fig. 3.

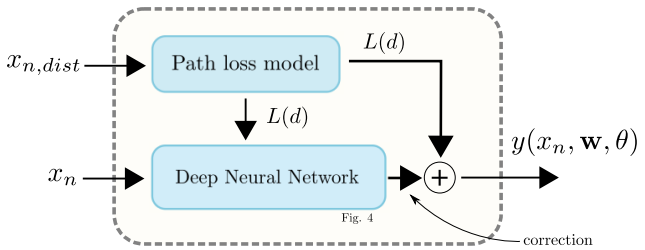


FIGURE 3. A simple path loss model is used in combination with a deep neural network. The DNN is thus tasked with learning a correction of a received power estimation. The structure of the DNN can be seen in Fig. 4.

In other words, the model is to learn a correction of the estimated path loss produced by the simple path loss model. Thus we can define

$$y(x_n, \mathbf{w}, \theta) = z([x_n, L(d)], \mathbf{w}, \theta) + L(d) \tag{7}$$

where $z(\cdot)$ is the DNN as detailed in the section below.

D. DEEP NEURAL NETWORK ARCHITECTURE

The proposed model $z(\cdot)$ consists of two NN and a CNN. The CNN is tasked with processing the satellite images, and a NN for managing engineered features and positional locators. The model concatenation can be observed in Fig. 4. Additional dense layers are added in sequence to the output of the CNN and the NN, termed NN2. One layer is used for adding the outputs of CNN and the NN. The sequentially used layers are added for enabling latent features to be a function of weighted positional locators and image features. This layer is directly connected to an output layer. The size of the layers and the overall architecture can be observed in Table 1 and 2.

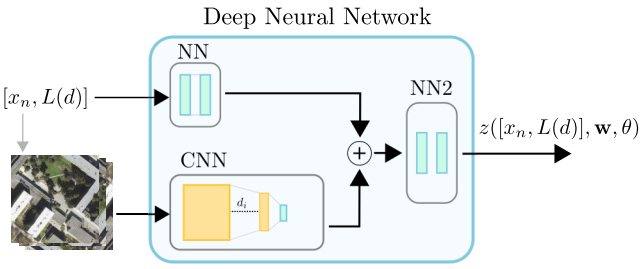


FIGURE 4. The DNN consists of a convolutional part, dealing with satellite images, and a regular dense NN dealing with the continuous features and inputs. The output of each module is added and condensed into several sequentially connected layers.

TABLE 1. Architecture of the CNN used for processing satellite images as detailed in Fig. 4.

	CNN
Input ch.	1
No. of convolutions	[200, 100, 50, 25, 12, 1]
Activation	*ReLU*
Kernel size	[(5,5), (3,3), (3,3), (3,3), (2,2), (2,2)]
Max pooling	2
Padding	2
Stride	1

TABLE 2. Architecture of the sub-models considered in the final model architecture.

NN		NN2	
Layer size	[200, 200]	Layer size	[200, 16, 1]
Activation	*ReLU*	Activation	*ReLU*

III. EXPERIMENTAL SETUP

The campus area of the Technical University of Denmark was selected for conducting measurements as it consists of

suburban and urban characteristics such as abundant vegetation and condensed collections of 3 tall story buildings with very different building materials. Fig. 5 shows a map of the measurement area.

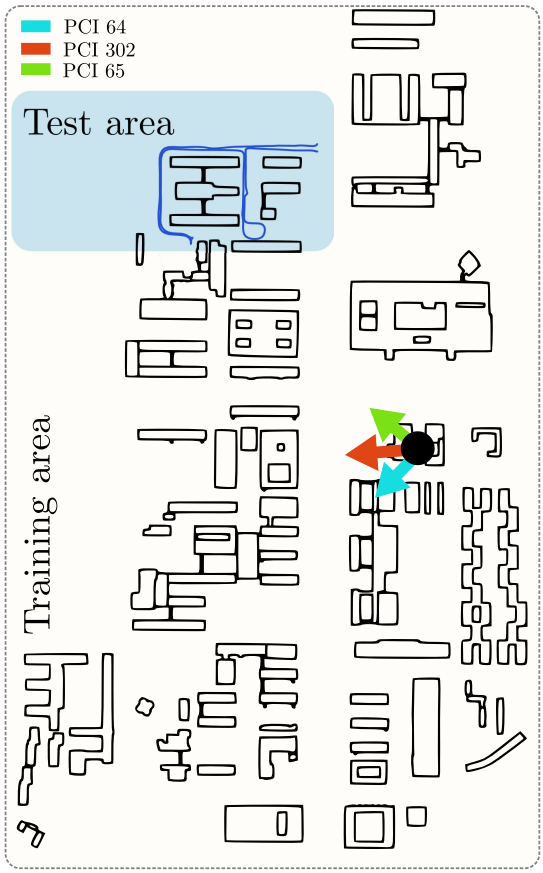


FIGURE 5. Measurement area where drive tests have been performed. The area is split into a training and test area. Measurements from the training area are used for the training set, while the measurements from the test area are used for the test set. The route used for training is not visualized.

A. RADIO MEASUREMENTS

Radio measurements were obtained using a Rohde & Schwarz (R&S) TSMW [27]. The hardware equipment is used along with the ROMES software from R&S [28], which is commonly used for drive testing. The area where measurements have been conducted can be seen in Fig. 5. A GPS module is integrated allowing for synchronization between radio measurements and GPS coordinates. The radio measurements were focused on downlink LTE-A frequencies in bands 20 and 7, more specifically EARFCN 6350 and 2850, which corresponds to frequencies 811 Mhz and 2630 MHz, respectively. Three base stations in the campus area were selected for the study. Transmitting from the same position, but with different configurations. PCI 64 and 65 are both operating at band 20 but considered two sectors of a cell site, while PCI 302 is operating at 2630 MHz and considered a single sector. 20 MHz of bandwidth is considered.

ROMES offers a parallelized capture of radio measurements, allowing for up to 32 independent measurements, each synchronized with GPS positioning. Roughly 14 km of road was driven, and a measurement was taken approximately every 2-4 ms. The speed of the vehicle was kept constant on longer stretches of road, thus approximated equal distance between measurement points. Parking lots and turning areas resulted in a reduced distance between measurements.

The resulting dataset (for both training and testing) was thus of size $\sim$ 60000 data points. The route used for testing is highlighted in Fig. 5. The bearing of 811 MHz cell resulted in fewer drive-able roads than the bearing of the two cells of 2630 MHz, which is the reason for fewer data points above 1100m.

B. SATELLITE IMAGES

Satellite images have, in recent years, been accessible with close to no cost. Even high-resolution satellite images can be obtained for free through services such as Mapbox [29]. For each GPS position (and thus radio measurement), a satellite image is obtained. Given the number of radio measurements, the same amount of satellite images is required. For this work, the static image rest API from Mapbox was used. The only pre-processing of the satellite images done was a rotation. The images were rotated according to the transmitter. More specifically, the direction of the transmitter is always down/south in the images. The bearing (the rotation angle) was computed using the GPS coordinates of the known measurements. Thus the centre of the corresponding satellite image is the GPS position. An example of this is seen in Fig. 6. The rotation of the images is primarily based on the intuition behind propagation, e.g., if the link state is Line-of-Sight (LOS), the majority of the signal propagation is given in a direct and straight line as the crow flies. A Mercator zoom level of 17 was chosen for each image. This constitutes 0.75 meter/pixel, thus the area covered by the satellite images of size 256 $\times$ 256 pixels is roughly $\sim$ 185 $\times$ 185 m [30]. Such a zoom level enables a clear indication of buildings, vegetation, and roads. This also means that nearby measurements have significant similar images causing an overlap of the area covered. The intuition here is that large-scale fading has a decorrelation distance. e.g., nearby spatial positions

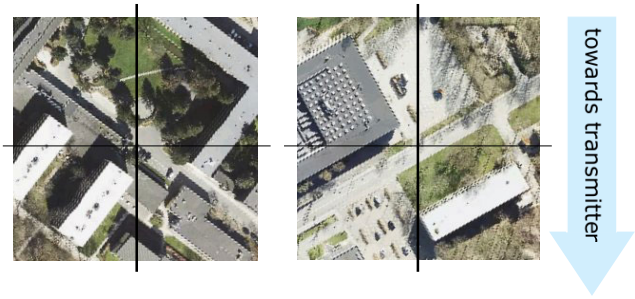


FIGURE 6. The images are rotated according to the position of the transmitter. Down/South in the image is towards the transmitter.

experience a similar magnitude of, for instance, shadowing. Thus the overlap of the satellite images is to increase the magnitude of latent features able to explain such large-scale fading impairments.

Each measurement is mapped to an image of resolution $256 \times 256 \times 3$ (three colour channels, RGB) producing the final dataset.

C. DATASET

The route illustrated in Fig. 5 is used for testing and remains unseen during training. The size of the test route is ~ 7000 samples, meaning the remainder $60000 - 7000 = 53000$ was using for training. A training/validation split of 75:25 was selected.

The complete and final dataset is a combination of both engineered features and obtained satellite images, as detailed in Eq. 3. A few features have been engineered, such as distance to the transmitter in both latitude and longitude direction. The transmitters have been denoted using a binary one-hot encoding to separate information from different transmitters at identical positions. The dataset can be found at [31].

IV. MODEL TRAINING AND VALIDATION

The model is implemented with the framework Pytorch [32], which enables easy use of NN methodologies and error minimization using back-propagation. The source code for the model and the dataset can be found here: `https://github.com/jakthra/PathLossPredictionSatelliteImages`.

Experiments were conducted for most hyper-parameters such as dropout probability, layer sizes, the minibatch size, the regularization magnitude (L1 and L2). The experiments were conducted using a random search (not exhaustive) of the hyper-parameters. Such a search entails sampling the hyper-parameters using a uniform distribution [33].

Data augmentation has been used for improving generalization, thus minimizing the gap between training error and test error. A random affine transformation is used, which keeps the centre of the image invariant but rotates and shears the image randomly. An example of such can be observed in Fig. 7. Random rotation of ± 20 degrees has been used,

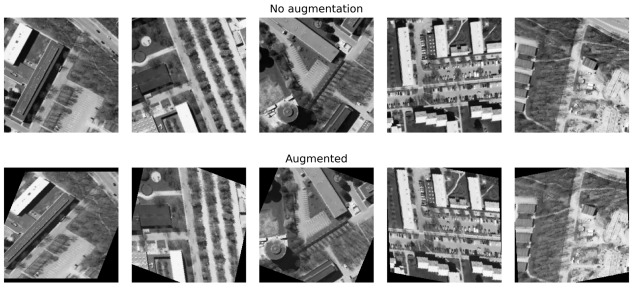


FIGURE 7. Examples of augmentation. Top row are satellite images rotated according to the transmitter but without transformation/augmentation applied. The bottom row is with transformation, thus the original images are augmented.

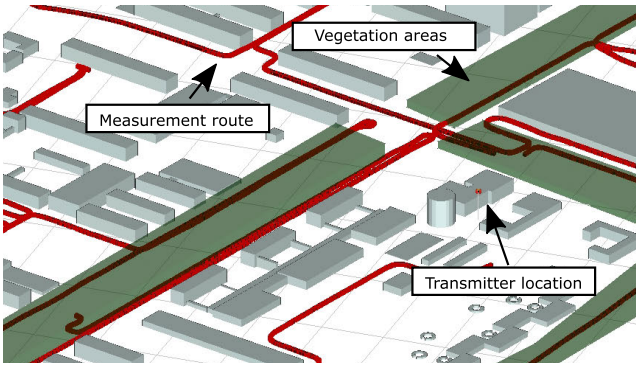


FIGURE 8. The route used for drive-testing is imported into a ray-tracing model, where buildings of the area have been imported using LIDAR and known vegetation areas are approximated.

with a ± 10 degree of shear. Data augmentation furthermore increases the size of the dataset as it produces many copies of the same measurement, but offers slightly different input images. Data augmentation is applied at every training iteration by applying a random transformation of the original image. Several epochs of training thus result in several training examples, with different transformations, of the original input image.

Additionally, the conversion to grey-scale, e.g., one channel instead of 3, offered improved generalization. A batch size of 30 was used, along with a weight decay of 0.0028. A learning rate scheduler was furthermore used, stepping the learning rate with a factor of 10 if no improvement on the test set is observed for several epochs. More details are available in the Github repository.

The final architecture of the model, as proposed in Fig. 4 are detailed in Table 1 and Table 2. In the remainder of this section, we describe and detail the ray-tracing and empirical model used for validating and evaluating the trained model.

A. RAY-TRACING

Ray-tracing requires GIS data such as radar data. This data is relatively inexpensive (if not able for free); however, obtaining updated newly scanned datasets can be expensive. If the datasets are too old, it does not constitute a fair reflection of the propagation environment. Urban environments are especially prone to outdated datasets since frequent construction is common. The authors in [34] highlight this by showing the poor resulting accuracy when not maintaining and updating GIS data. Additionally, the authors in [35] show that a significant improvement for path loss prediction (indoor) can be achieved by managing how the rays are launched depending on the receiver locations.

A ray-tracing model of the University Campus was constructed using a procedure as follows:

1) Obtained LIDAR scans of University Campus with a resolution of ~ 4.5 points per m^2 [36]
2) Obtained footprints of buildings in the study area from OpenStreetMap [37]

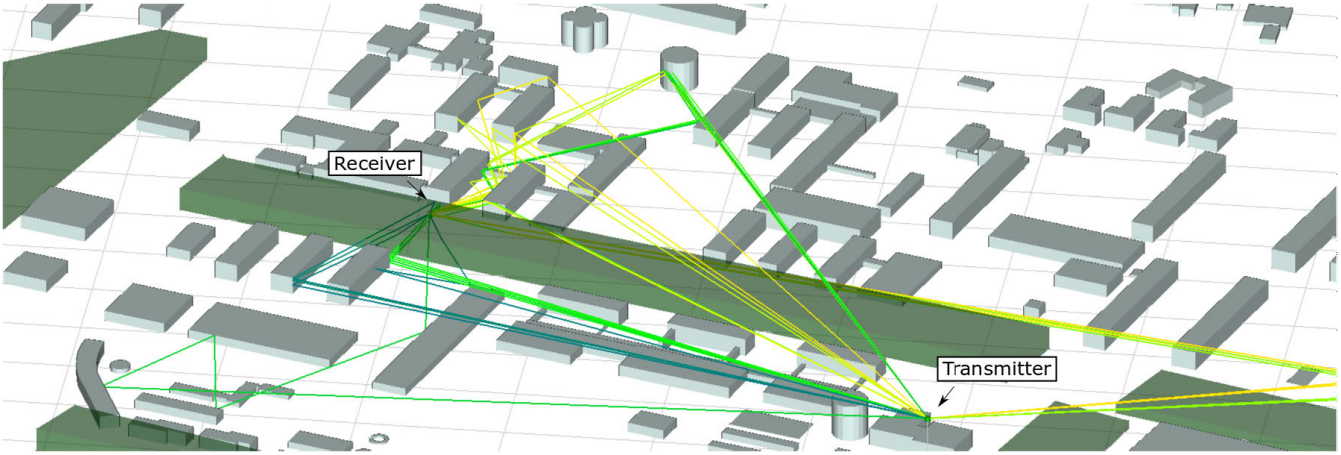


FIGURE 9. 25 Propagation paths for a single receiver point simulated by Remcom Wireless Insite. The polygon faces in the simulation define the propagation paths and are furthermore limited to a maximum of 25 paths. The 25 most likely propagation paths are computed for all ∼ 60000 data points.

3) Open-source software QGIS was used to extract vector shapes of buildings and their respective height.
4) Vector shapes and terrain data was added to the 3D model in the ray-tracing software. In this case, the Remcom ray-tracing solution was used [38]
5) Approximations of materials and their permittivity were defined along with transmitter and receiver configurations.

The above procedure produces a 3D model of the University campus. The position of measurements for 811 and 2630 MHz were imported to the model and simulated. The resulting dataset offers received power at measured locations according to the defined 3D model. Such an approach allows for a direct comparison of ray-tracing and experimental measurements. We have previously shown in [39] that state-of-the-art empirical path loss models achieve similar performance, and in some cases, outperform, that of an implemented ray-tracing model. In this work, the ray-tracing model is evaluated concerning LSP.

Remcom Wireless Insite [38] is used as the ray-tracing engine, and the properties of the simulation environment can be observed in Table 3. The number of paths is limited to 25, propagated from a standard half-wave dipole antenna array. The antennas in the ray-tracing engine were modelled to approximate the configuration of the measured cells and the respective sectors. The transmission power is set to 43 dBm with a sectorized SISO antenna definition (120 degrees of bearing with three sectors). The antenna is placed at the height of 30 m. The permittivity of the building materials (Concrete/Brick) is 4.4 to 5.3 F/m. A full 3D ray-tracing approach is used, accelerated by a GPU. Thus the number of faces define the overall complexity. The number of successful paths vary from `Rx` point to `Rx` point, and is determined by the number of faces present in the simulation environment. An example of the computed propagation paths can be observed in Fig. 9 for a single receiver point.

TABLE 3. Properties of the ray-tracing model implemented in Remcom.

Reflections	6
Diffractions	1
Area Size	14 km^2
Number of buildings	3917
Number of faces	16563
Building material	Concrete/Brick

B. EMPIRICAL PATH LOSS MODELS

The general parameter generation of 3GPP 38.901 and ITU-R M. 2412 consists of similar steps and input parameters. The path loss models used for the 3GPP 38.901 and ITU-R M.2412 are similar and are based on one same studies with small differences. In short, ITU-R M.2412 offers 2 channel models **A** and **B**, each for different propagation scenarios such as urban, suburban, or rural. The latter of the models (**B**) is identical to that of 3GPP 38.901. Thus, we refer to Urban Macro (UMa)_A as the definition in ITU-R M.2412 and UMa_B as the definition in TR 38.901 by 3GPP. Fig. 10 shows the predictive capability of the UMa_B along with the measurements conducted.

V. RESULTS

We compare the predictive performance of the proposed modelling technique with traditional modelling approaches on the forementioned test set. The error in terms of Root-Mean-Squared-Error (RMSE) (lower is better) can be observed in Fig. 11. The final obtained model utilizing satellite images is thus seen to offer a performance increase (of $\approx$ 1 dB for 811 MHz, and $\approx$ 4.7 dB for 2630 MHz) compared to the traditional approaches.

We compare the different learned models as follows; 1) a data-driven approach (no path loss model, thus no correction to be learned), where both images and features are utilized.

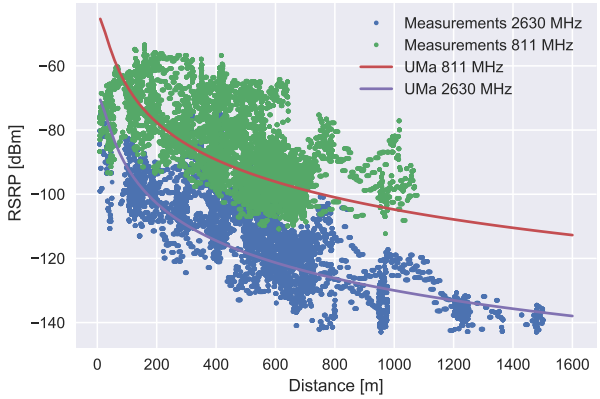


FIGURE 10. Experimental measurement of RSRP for 811 and 2630 MHz as a function of distance. The path loss model UMa_B is shown for both frequencies.

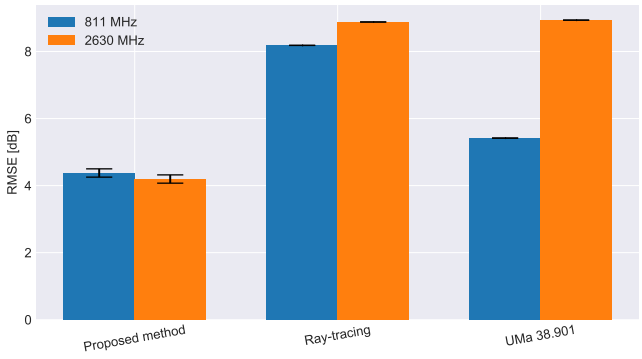


FIGURE 11. RMSE comparison of the proposed method to traditional modelling techniques. The standard deviation σ for the different experiments of training the DNN model is shown as errorbars.

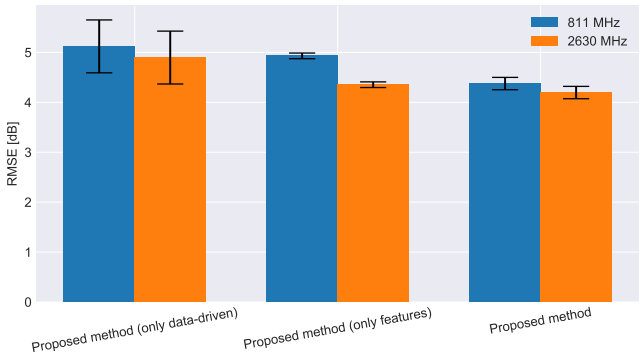


FIGURE 12. RMSE comparison of each modelling approach. A fully data-driven model (without the simple path loss model), and the proposed method without satellite images is included for reference.

2) a model using only features and 3) The full model where both images and features are used. These are compared in Fig. 12. A performance increase is observed when including images to the finalized model, as opposed to just utilizing the engineered features. By including images, an increase in predictive performance is observed to be $\approx$ 0.8($\pm$0.2) dB. This illustrates the improved generalization by including images to the model. The model-aided approach, where the model is tasked with learning a correction factor, offer an increase in the predictive performance of $\approx$ 1($\pm$0.15) dB for both 811 and 2630 MHz. Additionally, the data-driven approach is more prone to over/underfitting as highlighted by the increased σ. It was observed during training that data augmentation is necessary and effective for reducing the overfitting properties of the model during training. An example of the test and training error with and without data augmentation can be observed in Fig. 13. It can be seen that a severe overfit is present if no data augmentation is applied. This is observed by the test error diverging from the training error, also known as the *generalization gap*. This was remedied in [7] using regularization hyperparameters. However, data augmentation in combination with tuned regularization parameters has since shown to be the better performing approach.

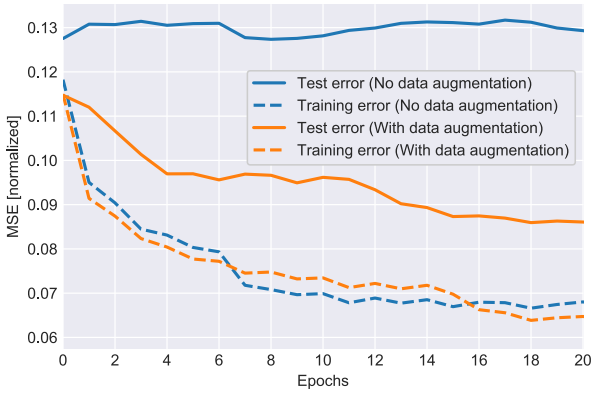


FIGURE 13. Example of the test error with and without data augmentation applied (shown in solid). The training error is shown as the dashed line for both approaches.

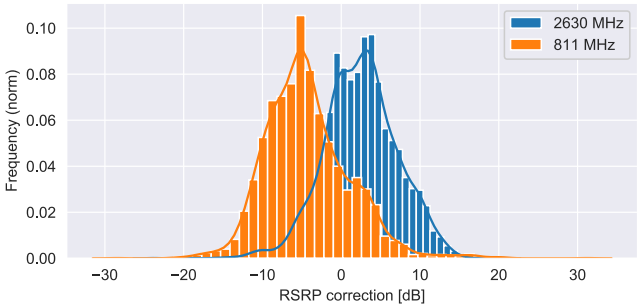
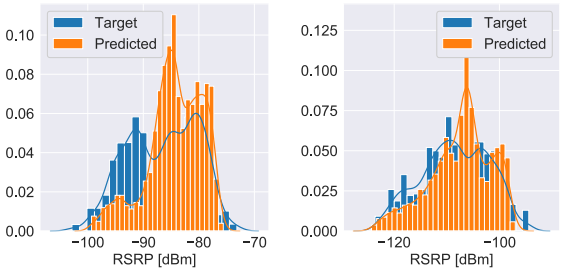


FIGURE 14. Histogram of path loss correction for RSRP given by the model for 811 and 2630 MHz on the test set.

The distribution of the learned correction of the trained model is observed in Fig. 14. The distributions are Gaussian-shaped, which corresponds to the distribution of large-scale fading. The model has thus learned that the correction is Gaussian distributed around the mean predicted power by the simple path loss model. The slight offset of the distributions from having a mean zero illustrates an offset in calibrating the simple path loss models. The correction distribution thus indicates the need for re-calibrating by $\approx$ 4 dB for 811 MHz, and $\approx$ $-$3 dB for 2630 MHz. It is important to note that no prediction improvement was found by re-calibrating the distributions to have mean zero.

The distributions of both the model output and the measurements are shown in Fig 15. A significant better distribution fit (visually) is observed for 2630 MHz than 811 MHz.

For 811 MHz (Fig. 15a), a cluster of measurements seem to be difficult to predict (at around −95 dBm).



(a) 811 MHz (b) 2630 MHz

FIGURE 15. Distribution of RSRP for 811 (a) and 2630 (b) MHz on the test set and the model output.

VI. DISCUSSION

The satellite images offer much information, and the entirety is not necessarily relevant for radio performance prediction, especially at lower frequencies. The use of such images results ultimately in a model that is harder to train since the latent features obtained by the CNN might be sparse and, to some extent, memorization of the training data. The use of other information sources such as LIDAR scans seem appropriate and might offer an improved level of information to predict the propagation of radio waves. Furthermore, a timeline of images is required to consider the influence of seasonal changes on radio propagation, for instance, the leaves on trees.

Finally, given the size of the training set, it is believed that simplistic satellite images or even just vectorized maps might offer comparable or even improved results. This approach will also require less complicated model selection techniques which will thus lead to reduced training time. Finally, more experimental data is required to validate and assess the performance of the proposed method. The results provided in this work does *indicate* the use of satellite images enable interpolation between unseen areas and measurements, thus presenting a generalization of signal attenuation at higher frequencies.

The optimization of Hyper-parameters, for instance, the number of convolutional layers and their respective size, is challenging. Deep Learning models have many hyper-parameters and doing a systematic grid search for each is not feasible in terms of training time. Obtaining a better choice of hyper-parameters, than documented in this paper, is a possibility. The completed hyper-parameter optimization and thus model selection procedures of this work consist of 300 experiments and is considered the most time-consuming aspect of obtaining such a solution. For such reasons comparing complexity can be tricky. We experienced roughly 240 minutes of training time. For comparison, we experienced roughly 120 minutes for producing ray-tracing results for the test route, e.g. the approximately 7000 data points. Both accelerated with a GPU Nvidia 1080 Ti. Predicting the test route using the DL model was completed in $\sim$ 3 minutes.

We argue that a reduction of the overall data complexity is achieved while improving the prediction accuracy, as satellite images and the features used are simple and easy to obtain. In terms of model complexity, it is difficult to argue that any gains have been achieved. These gains suffer significantly due to expensive model selection experiments. The model selection and training of the model can, in principle, only be done once, due to the improved generalization properties of including images. If done so, the model complexity lies only in the prediction time and the memory required, both of which, is significantly lower than a ray-tracing approach.

Future work will be invested in obtaining a dataset that is decoupled from the campus area, as has been used for training. A difference in building placement and overall architecture could have a significant impact on the test performance of the system. However, this remains to be tested. The dataset and model have been made public and can be downloaded for free. Any extension to either the model architecture, hyper-parameters or the dataset is welcomed.

VII. CONCLUSION

Accurate path loss prediction with improved generalization using satellite images can be achieved with the use of convoluted neural networks. A gain of $\approx$ 1 dB has been achieved at 811 MHz, and $\approx$ 4.7 dB at 2630 MHz, compared to traditional modelling techniques such as ray-tracing and empirical models. Additionally, utilizing a simple path loss model to assist the neural network with learning offer improved predictive performance with a gain of $\approx$ 1 dB. Including satellite images increase the predictive performance additionally by $\approx$ 0.8 dB. Additionally, we conclude that the complexity of such deep neural networks is primarily associated with model selection principles and the run-time hereof. Finally, the proposed approach would benefit from an increase in data, as to quantify the generalization achieved by including metadata such as satellite images.

ACRONYMS

Notation	Description
CNN	Convoluted Neural Network.
DL	Deep Learning.
DNN	Deep Neural Network.
H-UDN	Heterogeneous UltraDense Network.
LOS	Line-of-Sight.
LSP	Large-Scale Parameter.
ML	Machine Learning.
mmWave	Millimeter Wave.
NN	Neural Network.
NR	New Radio.
R&S	Rohde & Schwarz.
ReLU	Rectified Linear Unit.
RMSE	Root-Mean-Squared-Error.
RSRP	Reference Signal Received Power.
UMa	Urban Macro.

REFERENCES

[1] S. Yunas, M. Valkama, and J. Niemelä, "Spectral and energy efficiency of ultra-dense networks under different deployment strategies," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 90–100, Jan. 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7010521/

[2] X. Ge, S. Tu, G. Mao, and C. X. Wang, "5g ultra-dense cellular networks," *IEEE Trans. Wireless Commun.*, vol. 23, no. 1, pp. 72–79, Feb. 2016. [Online]. Available: http://ieeexplore.ieee.org/document/7422408/

[3] A. Taufique, M. Jaber, A. Imran, Z. Dawy, and E. Yacoub, "Planning wireless cellular networks of future: Outlook, challenges and opportunities," *IEEE Access*, vol. 5, pp. 4821–4845, 2017. [Online]. Available: http://ieeexplore.ieee.org/document/7883847/

[4] C.-X. Wang, J. Bian, J. Sun, W. Zhang, and M. Zhang, "A survey of 5G channel measurements and models," *IEEE Commun. Surveys Tutr.*, vol. 20, no. 4, pp. 3142–3168, 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8424015/

[5] Z. Yun and M. F. Iskander, "Ray tracing for radio propagation modeling: Principles and applications," *IEEE Access*, vol. 3, pp. 1089–1100, 2015. [Online]. Available: http://ieeexplore.ieee.org/document/7152831/

[6] Z. Zhang, G. Yang, Z. Ma, M. Xiao, Z. Ding, and P. Fan, "Heterogeneous ultradense networks with NOMA: System architecture, coordination framework, and performance evaluation," *IEEE Veh. Technol. Mag.*, vol. 13, no. 2, pp. 110–120, Jun. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8351641/

[7] J. Thrane, M. Artuso, D. Zibar, and H. L. Christiansen, "Drive test minimization using deep learning with Bayesian approximation," in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Aug. 2018, pp. 1–5.

[8] *TR 138 901 - V14.3.0 - 5G; Study on Channel Model for Frequencies From 0.5 to 100 GHz*, 3GPP document TR 38.901 version 14.3.0 Release 14. [Online]. Available: http://www.etsi.org/standards-search

[9] "Guidelines for evaluation of radio interface technologies for IMT-2020," ITU-R, Tech. Rep. M.2412-0, 2017. [Online]. Available: https://www.itu.int/pub/R-REP-M.2412-2017

[10] E. Ostlin, H. Zepernick, and H. Suzuki, "Macrocell path–loss prediction using artificial neural networks," *IEEE Trans. Veh. Technol.*, vol. 59, no. 6, pp. 2735–2747, Jul. 2010. [Online]. Available: http://ieeexplore.ieee.org/document/5466252/

[11] S. I. Popoola, E. Adetiba, A. A. Atayero, N. Faruk, and C. T. Calafate, "Optimal model for path loss predictions using feed-forward neural networks," *Cogent Eng.*, vol. 5, no. 1, Feb. 2018, Art. no. 1444345. [Online]. Available: https://www.cogentoa.com/article/10.1080/23311916.2018.1444345

[12] S. I. Popoola, A. Jefia, A. A. Atayero, O. Kingsley, N. Faruk, O. F. Oseni, and R. O. Abolade, "Determination of neural network parameters for path loss prediction in very high frequency wireless channel," *IEEE Access*, vol. 7, pp. 150462–150483, 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8865025/

[13] S. P. Sotiroudis, S. K. Goudos, and K. Siakavara, "Neural networks and random forests: A comparison regarding prediction of propagation path loss for NB-IoT networks" in *Proc. 8th Int. Conf. Mod. Circuits Syst. Technol. (MOCAST)*, May 2019.

[14] T. J. O'Shea and J. Hoydis, "An introduction to deep learning for the physical layer," Feb. 2017, *arXiv:1702.00832*. [Online]. Available: https://arxiv.org/abs/1702.00832

[15] S. Dorner, S. Cammerer, J. Hoydis, and S. T. Brink, "Deep learning based communication over the air," *IEEE J. Sel. Top. Signal Process.*, vol. 12, no. 1, pp. 132–143, Feb. 2018. [Online]. Available: https://ieeexplore.ieee.org/document/8214233/

[16] F. A. Aoudia and J. Hoydis, "End-to-end learning of communications systems without a channel model," Apr. 2018, *arXiv:1804.02276*. [Online]. Available: https://arxiv.org/abs/1804.02276

[17] C. Phillips, D. Sicker, and D. Grunwald, "Bounding the practical error of path loss models," *Int. J. Antennas Propag.*, vol. 2012, Mar. 2012, Art. no. 754158. [Online]. Available: http://www.hindawi.com/journals/ijap/2012/754158/

[18] F. Perez-Fontan, P. M. Espineira, *Modeling the Wireless Propagation Channel: A Simulation Approach With MATLAB*. Hoboken, NJ, USA: Wiley, 2008.

[19] D. N. C. Tse and P. Viswanath, *Fundamentals of Wireless Communications*. Cambridge, U.K.: Cambridge Univ. Press, 2005.

[20] C. M. Bishop, *Pattern Recognition and Machine Learning*. New York, NY, USA: Springer, 2006.

[21] M. A. Nilsen, *Neural Networks and Deep Learning*. Determination Press, 2015. Accessed: Nov. 2019. [Online]. Available: http://neuralnetworksanddeeplearning.com/

[22] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. Cambridge, MA, USA: MIT Press, 2016.

[23] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, May 2015. [Online]. Available: http://www.nature.com/articles/nature14539

[24] A. Zeng, S. Song, J. Lee, A. Rodriguez, and T. Funkhouser, "TossingBot: Learning to throw arbitrary objects with residual physics," Mar. 2019, *arXiv:1903.11239*. [Online]. Available: https://arxiv.org/abs/1903.11239

[25] A. Zappone, M. Di Renzo, M. Debbah, T. T. Lam, and X. Qian, "Model-aided wireless artificial intelligence: Embedding expert knowledge in deep neural networks for wireless system optimization," *IEEE Veh. Technol. Mag.*, vol. 14, no. 3, pp. 60–69, Sep. 2019. [Online]. Available: https://ieeexplore.ieee.org/document/8765703/

[26] B. J. Cavalcanti, G. A. Cavalcante, L. M. De Mendonça, G. M. Cantanhede, M. M. De Oliveira, and A. G. D'Assunção, "A hybrid path loss prediction model based on artificial neural networks using empirical models for LTE and LTE-A at 800 MHz and 2600 MHz," *J. Microw., Optoelectron. Electromagn. Appl.*, vol. 16, no. 3, pp. 708–722, Sep. 2017.

[27] "R&S TSMW universal radio network analyzers user manual," Rohde & Schwarz GmbH & Co. KG, Munich Germany, Tech. Rep. 1171.5960.32-8, 2017.

[28] "ROMES4 drive test software: Mobile coverage and QoS measurements in mobile networks," Rohde & Schwarz GmbH & Co. KG, Munich Germany, Tech. Rep. 5214.2062.12, 2018.

[29] *Mapbox*. Accessed: Nov. 2019. [Online]. Available: https://mapbox.com

[30] *OpenStreetMapWiki*. Accessed: Nov. 2019. [Online]. Available: https://wiki.openstreetmap.org/

[31] J. Thrane and H. L. Christiansen, "Mobile communication system measurements and satellite images," IEEE Dataport, Tech. Univ. Denmark, Lyngby, Denmark, 2019. [Online]. Available: http://dx.doi.org/10.21227/1xf4-eg98

[32] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer. (2017). *Automatic Differentiation in PyTorch*. https://www.semanticscholar.org/paper/Automatic-differentiation-in-PyTorch-Paszke-Gross/b36a5bb1707bb9c70025294b3a310138aae8327a

[33] J. Bergstra and Y. Bengio, "Random search for hyper-parameter optimization," *J. Mach. Learn. Res.*, vol. 13, pp. 281–305, 2012.

[34] E. M. Vitucci, V. Degli-Esposti, F. Fuschini, J. S. Lu, M. Barbiroli, J. N. Wu, M. Zoli, J. J. Zhu, and H. L. Bertoni, "Ray tracing RF field prediction: An unforgiving validation," *Int. J. Antennas Propag.*, vol. 2015, pp. 1–11, Aug. 2015. [Online]. Available: http://www.hindawi.com/journals/ijap/2015/184608/

[35] F. Hossain, T. Kim Geok, T. Abd Rahman, M. Nour Hindia, K. Dimyati, S. Ahmed, C. P. Tso, A. Abdaziz, W. Lim, A. Mahmud, T. C. Peng, C. P. Liew, and V. Thiruchelvam, "Indoor 3-D RT radio wave propagation prediction method: PL and RSSI modeling validation by measurement at 4.5 GHz," *Electronics*, vol. 8, no. 7, p. 750, Jul. 2019. [Online]. Available: https://www.mdpi.com/2079-9292/8/7/750

[36] *Kortforsyningen*. Accessed: Nov. 2019. [Online]. Available: https://kortforsyningen.dk/

[37] *Geofabrik Download Server*. Accessed: Nov. 2019. [Online]. Available: https://download.geofabrik.de/europe/denmark.html

[38] *Electromagnetic Simulation Software Remcom Wireless Insite*. Accessed: Nov. 2019. [Online]. Available: https://www.remcom.com/

[39] J. Thrane, D. Zibar, and H. L. Christiansen, "Comparison of empirical and ray-tracing models for mobile communication systems at 2.6 GHz," in *Proc. IEEE 90th Veh. Technol. Conf. (VTC-Fall)*, Sep. 2019, pp. 1–5.

JAKOB THRANE (Student Member, IEEE) received the M.Sc. degree in telecommunication engineering from the Technical University of Denmark (DTU), in 2016, where he is currently pursuing the Ph.D. degree in machine learning and deep learning applications to mobile networks. He previously worked with optical communication and optimization of physical layer parameters. His main areas of research are wireless transmission, mobile networks, and machine learning.

DARKO ZIBAR (Member, IEEE) received the M.Sc. degree in telecommunication and the Ph.D. degree in optical communications from the Technical University of Denmark, in 2004 and 2007, respectively. He is currently an Associate Professor with the DTU Fotonik, Technical University of Denmark, and the Head of the Machine Learning in Photonics (M-LiPS) Group. His research efforts are currently focused on the application of machine learning methods to optical communication, ultrasensitive amplitude and phase detection, and optical fiber sensing systems.

HENRIK LEHRMANN CHRISTIANSEN (Member, IEEE) received the M.Sc.E.E. and Ph.D. degrees specializing in telecommunications from the Technical University of Denmark. He has several years of experience in the telecom industry. He is currently an Associate Professor in mobile communication with the Technical University of Denmark. His main areas of research are mobile network architectures, and mobile fronthaul and backhaul networks.