Deep Learning-based Signal Strength Prediction Using Geographical Images and Expert Knowledge

Jakob Thrane[1], Benjamin Sliwa[2], Christian Wietfeld[2], and Henrik L. Christiansen[1]

[1]Department of Photonics Engineering, Technical University of Denmark, 2800 Kongens Lyngby, Denmark
[2]Communication Networks Institute, TU Dortmund University, 44227 Dortmund, Germany
e-mail: [1]{jathr, hlch}@fotonik.dtu.dk [2]{Benjamin.Sliwa, Christian.Wietfeld}@tu-dortmund.de

Abstract—Methods for accurate prediction of radio signal quality parameters are crucial for optimization of mobile networks, and a necessity for future autonomous driving solutions. The power-distance relation of current empirical models struggles with describing the specific local geo-statistics that influence signal quality parameters. The use of empirical models commonly results in an over- or under-estimation of the signal quality parameters and require additional calibration studies.

In this paper, we present a novel model-aided deep learning approach for path loss prediction, which implicitly extracts radio propagation characteristics from top-view geographical images of the receiver location. In a comprehensive evaluation campaign, we apply the proposed method on an extensive real-world data set consisting of five different scenarios and more than 125.000 individual measurements.

It is found that 1) the novel approach reduces the average prediction error by up to 53 % in comparison to ray-tracing techniques, 2) A distance of $250 - 300$ meters spanned by the images offer the necessary level of detail, 3) Predictions with a root-mean-squared error of ≈ 6 dB is achieved across inherently different data sources.

Fig. 1. Overview about different application scenarios for the proposed image-based signal strength prediction method.

I. INTRODUCTION

Radio propagation modelling is a key building block for the design of wireless communication systems and represents one the foundations for network planning [1] and network simulation [2]. Also, the ability to forecast network quality indicators for given geographical locations – e.g., along a vehicular trajectory – is an enabler for anticipatory networking [3] techniques such as opportunistic data transfer [4]. In a recent study [5], the 5G Automotive Association (5GAA) has pointed out the need to implement *predictive Quality of Service (QoS)* methods for enabling connected and autonomous driving. Moreover, knowledge about the radio propagation characteristics can be exploited to infer indicators that significantly correlate to the former. An example is the prediction of the applied transmission power based on signal strength and signal quality measurements discussed in [6].

Although conventional model-based methods are suitable for comparing the behaviour of different methods under study (e.g., resource schedulers) in abstract reference scenarios, they are not able to accurately represent the radio propagation effects in complex concrete real-world environments [7].

In this work, we present a hybrid approach which brings together model-based and data-driven methods for path loss prediction and for constructing radio environmental maps. We apply *deep learning* – which is known to achieve outstanding performances in the image classification domain – on top-view
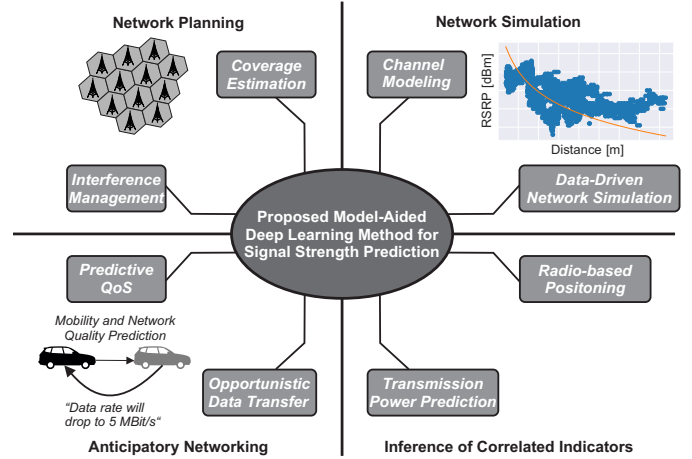
images of the receiver environment for learning geographical *radio environmental prototypes*. The latter are then utilized to forecast the received power at unobserved locations. Based on the groundwork presented in [8] where satellite images have been exploited to infer radio quality parameters, we analyze the prediction performance based on vector images obtained from OpenStreetMap (OSM) in this work. The contributions are summarized as follows:

- Presentation of a novel **model-aided deep learning method** for path loss prediction based on OSM images of the receiver environment.
- Evaluation of the proposed method on a large **real world data set**.
- The developed software is provided in an **Open Souce**[1] way.

An overview of different target applications for the proposed method is shown in Fig. 1. The remainder of the paper is structured as follows. After discussing the related work in Sec. II, we present the proposed solution approach in Sec. III. Afterwards, the applied methodology is introduced in Sec. IV and finally, the results of the performance evaluation are presented and discussed in Sec. V.

---

[1]Available at https://github.com/jakthra/PseudoRayTracingOSM

II. RELATED WORK

Radio propagation modelling and prediction: *Model-based* methods represent the standard approach for considering radio propagation effects in network simulation [2] and network planning. Existing channel models – e.g. 3GPP TR 38.901 [9] – provide high computational efficiency and allow the comparison of different methods in a highly controlled environment. However, since fading effects are typically represented as probabilistic attenuation functions and obstacles are only modelled statistically and not explicitly, these methods are mostly not able to mimic the complex radio propagation characteristics of particular real-world scenarios [4]. Ray tracing methods [10] aim to close this gap by using detailed models of the environment to consider the dynamics of the radio link with respect to physical effects such as reflection and refraction. Although this approach is theoretically capable of providing highly-accurate representations of the radio link behaviour in concrete real-world scenarios, practical applications often lack the required amount of high-resolution data about shape and material of the obstacles. As an alternative to model-centric methods, data-driven approaches have emerged in recent years. Radio environmental maps [11] maintain network quality information – which is often acquired in a crowdsensing manner [12] – based on a grid with defined cell granularity. As discussed by [13], incomplete measurements can be compensated by *kriging*-based interpolated techniques.

Machine learning has gained immense popularity in the wireless networking domain as it allows to implicitly learn hidden interdependencies between measurable variables which are often too complex to bring together analytically. A comprehensive summary of different machine learning methods and their application for various communication applications is provided by the authors of [14]. *Deep learning* [15] has become one of the most popular learning methods after achieving impressive results in the image processing domain. Hereby, *deep* Artificial Neural Networks (ANNs), which consist of a high number of hidden layers, are iteratively trained to minimize a certain loss function. A recent trend in this domain is the embedding of expert knowledge into machine learning models. Model-aided wireless artificial intelligence [16] allows to optimize the accuracy of prediction models further and to reduce the number of required training samples. Alongside with the convergence of machine learning and wireless communications, different authors aim to improve network quality prediction in complex scenarios by replacing the traditional mathematical radio propagation models with machine learning. In [17], Enami et al. propose the Regional Analysis to Infer KPIs (RAIK) framework which utilizes geographical features (e.g., the percentage of the area covered by buildings) for enhancing the prediction quality of different network performance indicators such as Received Signal Strength (RSS). Hereby, a highly-detailed environment model that consists of buildings and trees is constructed from Light Detection and Ranging (LIDAR) information. A similar approach is implemented by the authors of [18], which further considers the number of building penetrations on the direct path between transmitter and receiver. Both approaches can
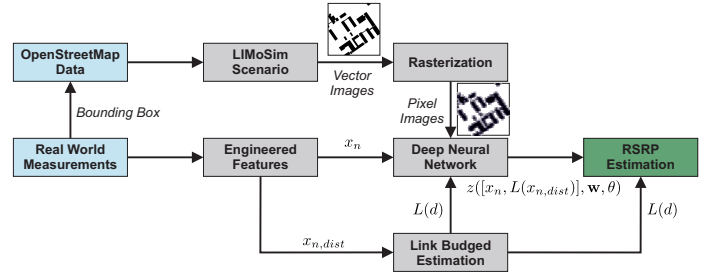

Fig. 2. Overall system architecture model of the proposed approach.

achieve a significant reduction of the resulting prediction error in comparison to traditional channel models. In contrast to these existing methods, the proposed approach utilizes the raw geographical *images* (instead of extracted features) of the environment between transmitter and receiver for deep learning-based Reference Signal Received Power (RSRP) prediction. Our general assumption and motivation for this work are that similar-looking environments will likely show similar radio propagation characteristics.

III. MACHINE LEARNING-BASED SIGNAL STRENGTH PREDICTION

The overall system architecture model for the proposed model-aided deep learning approach is shown in Fig. 2. The model architecture utilized for this work is identical to basic principles detailed in [8], however, with a few changes to 1) input parameters, i.e. features and 2) the overall model complexity. A basic path loss model is utilized to aid the learning process through a rough estimation of the link budget. The link budget consists of no learnable parameters and is based on the 3GPP UMa model [9]. The learning objective of the proposed model is thus to approximate the function $y(\cdot)$ that is capable of approximating the RSRP such that

$$\text{RSRP} = y(x_n, \mathbf{w}, \theta) + \epsilon \qquad (1)$$

where $y(\cdot)$ is the model to learn, with inputs x_n, weights $\mathbf{w}$, and hyper-parameters θ. The noise ϵ on the observed RSRP values is assumed Gaussian distributed, which fits well with the imposed log-normal distribution of large-scale fading. The model, $y(\cdot)$ is decomposed into a basic path loss model $L(\cdot)$ and a Deep Neural Network (DNN) $z(\cdot)$.

$$y(x_n, \mathbf{w}, \theta) = L(x_{n,dist}) + z([x_n, L(x_{n,dist})], \mathbf{w}, \theta) \qquad (2)$$

The learning objective is defined as maximizing the likelihood function through the minimization of the sum-of-squares error function between the model output, e.g. $y(x_n, ...)$ and the observed values t_n [8], [15].

The observations of the RSRP are obtained using a drive-testing approach (see Sec. IV-A). The features, i.e. the inputs to the models are (re)-defined as

$$x_n = [v, d, \Delta_{\text{lat}}, \Delta_{\text{lon}}, f_c, \mathcal{A}] \qquad (3)$$

where v is the vehicle's velocity, d is the 3D distance between the User Equipment (UE) and evolved Node B (eNB).
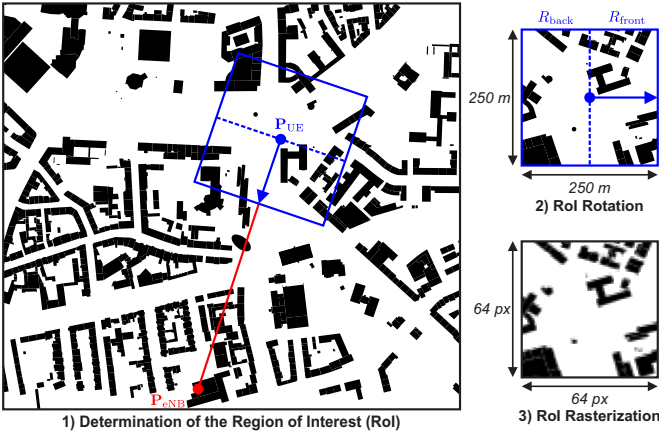
Fig. 3. Example of the different steps of the generation process of the environmental images based on the GER Urban data set. (Map data: ©OpenStreetMap contributors, CC BY-SA).

This is different from the work in [8] where the raw latitude and longitude coordinates where used. These features are reduced to the differences Δ_{lat}, and Δ_{lon} respectively. f_c is the carrier frequency in MHz. Finally, $\mathcal{A}$ denotes an image displaying the local surroundings of the UE.

A. Generation of the Environmental Images

For a given scenario, the corresponding OSM map is exported based on the latitude/longitude bounding box of the scenario. The map file is then imported by the vehicular mobility simulator Lightweight ICT-centric Mobility Simulation (LIMoSim) [19] which converts the raw OSM data into a simulation scenario. We utilize this uncommon approach since LIMoSim provides an integrated engine for automatically exporting vector graphic figures of the environment. The actual generation of the environmental images that contain the Region of Interest (RoI) is then performed in a step-wise process:

1) **RoI determination**: We define the RoI as a quadratic box of width w centered around the UE location $\mathbf{P}_{\text{UE}}$ which points towards the eNB position $\mathbf{P}_{\text{eNB}}$.
2) **RoI rotation**: It can be assumed that different types of regions within the RoI provide different types of information for the signal strength prediction process since they are affected by different radio propagation effects. For example, the front region R_{front} facing the eNB position is highly impacted by obstacle shadowing while the back region R_{back} more likely corresponds to multipath propagation effects. For allowing the neural networks to implicitly learns these impact factors, the rotation of the images must be normalized. For a given direction vector $\mathbf{v} = \mathbf{P}_{\text{eNB}} - \mathbf{P}_{\text{UE}}$, all elements of the image are rotated around the angle $\phi = -\operatorname{atan2}(\mathbf{v}.y, \mathbf{v}.x)$.
3) **RoI rasterization**: Finally, the vector image is rasterized in order to allow the further processing with the deep learning pipeline. Each pixel of the image corresponds to one input neuron of the neural network.

An example which illustrates the different processing steps in a concrete scenario is shown in Fig. 3.

B. Deep Neural Network

The model, and function $z(\cdot)$ consists of DNN building blocks and methodologies. The model utilizes a set of convolutional layers to process the image input, and a set of dense fully-connected layers to process the remainder of the features. In short, three sets of sub-functions are utilized within the DNN, which transforms a set of features into valid predictions.

$$z(x_n, \mathbf{w}, \theta) = f[g(x_{n,A}, \mathbf{w_g}, \theta_{\mathbf{g}}), h(x_{n,\notin A}, \mathbf{w_h}, \theta_{\mathbf{h}}), \mathbf{w_f}, \theta_{\mathbf{f}}] \tag{4}$$

Here the function $g(\cdot)$ denotes a convolutional neural network, which consists of 4 layers. Each layer uses a set of sequential operations. 1) 2D convolutions, 2) Rectified Linear Unit (ReLU) activation, 3) Batch Normalization and 4) Max Pooling. The function $h(\cdot)$ is a dense neural network and is termed *Feature NN*. The NN takes all features except that of the image as input. A set of fully-connected layers and a ReLU activation function is used in the NN. The function $f(\cdot)$ is termed *Output NN*, and is a set of fully-connected linear layers that adds the latent features provided by the output of $g(\cdot)$ and $h(\cdot)$. The parameters for each sub-function and layer can be found in Tab. I.

IV. METHODOLOGY

In this section, we give an overview of the evaluation scenarios, the performance metrics and the training of the Deep Learning model.

A. Data Sets and Scenarios

For the following performance evaluation, we use the aggregated measurements of two large data sets which have initially been acquired in [8] (Danish data set DK, single Mobile Network Operator (MNO)) and [20] (German data set GER, three MNOs). The resulting data set consists of more than 125.000 individual vehicular RSRP measurements in five different scenarios with different building densities:

- **DK Campus** (57586 samples): Campus area of the Technical University of Denmark
- **GER Campus** (8579 samples): Campus area of the TU Dortmund University
- **GER Urban** (11921 samples): Inner city ring of Dortmund
- **GER Suburban** (27152 samples): Suburban region close to campus of the TU Dortmund University
- **GER Highway** (20662 samples): German highway A45 with a maximum speed of up to 150 km/h

B. Error Metric

For the evaluation of the prediction performance, we focus on the Root Mean Square Error (RMSE) which is defined as

$$\text{RMSE} = \sqrt{\frac{\sum_{i=1}^{N} (\tilde{y}_i - y_i)^2}{N}}. \tag{5}$$

with $\tilde{y}_i$ being the current prediction, y_i being the current true value and N being the number of samples.

TABLE I
HYPER-PARAMETERS FOR THE DEEP NEURAL NETWORK MODEL.

Parameter	Value
Weight decay	8e-4
Learning rate	1e-3
Filters	[32, 32, 10, 1]
Kernel size	[(5,5), (3,3), (3,3), (2,2)]
Max pooling	[2, 2, 2, 2]
Feature NN layer size	[32, 32]
Output NN layer size	[16, 16]
Image augmentation angle	20
Image size	64 px × 64 px
Batch size	12

C. Training the model

The proposed method is implemented using the framework PyTorch accelerated using a GTX 1080 Ti GPU. The implementation is available open-source. The well-known Adam optimizer is used for training the model through backpropagation principles. A so-called learning rate scheduler on plateau is utilized, with a patience parameter of 20 epochs, before lowering the learning rate with a factor of 10. In other words, if the test performance has not seen any improvements for 20 epochs, the learning rate is lowered. Early stopping is enforced when the learning rate is $< 1e - 7$. So-called mini-batch training is utilized [15], splitting the size of the data set into smaller batches. We denote a single epoch as iterating over all mini-batches in the data set.

Each image is supplied as a `.png` image, with a fixed resolution. Data augmentation is applied to the images to improve the generalization of the model and reduce the overfitting during training. A random transformation (random rotation of ± 20 degrees) is applied to the input images every epoch.

V. RESULTS

In this section, we present and discuss the results of achieved by applying the proposed method on real-world measurement data.

A. Setup Configuration, Training and Parameter Selection

The hyper-parameters used for the resulting Deep Learning model can be found in Table I. The hyper-parameters found are the results of an extensive search (over 500 experiments) using Bayesian optimization techniques [15] utilizing all of the available data. A significant reduction in model complexity compared to [8] is achieved. For instance, a reduction of ≈ 170 filters in the initial convolutional layers is achieved. The fully-connected layers are also reduced from 200 neurons to only 32. Furthermore, the number of convolutional layers are reduced from 6 to 4.

The method proposed in this paper is an extension of the work in [8]. A comparison of the obtained performance can be seen in Fig. 4. A similar level of performance is achieved, even with a significant reduction in model complexity. In comparison to ray-tracing techniques, the average RMSE is reduced by 53%.
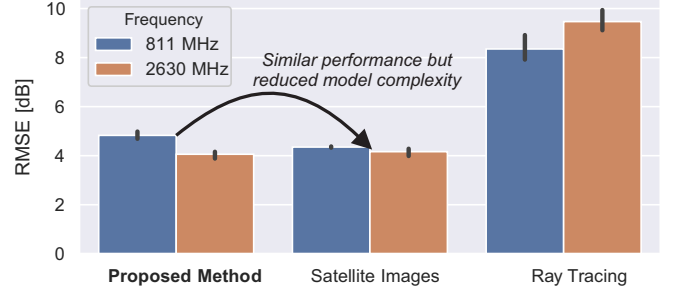

Fig. 4. Performance of the proposed method compared to the use of satellite images and ray tracing for the `DK Campus` data set [8].
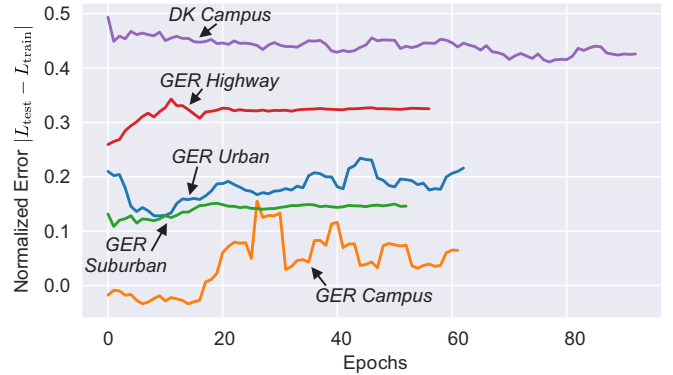

Fig. 5. The difference between the training and test error over training epochs for each data subset.

B. Cross-scenario Performance

The cross-scenario performance is evaluated in terms of RMSE. Hereby, one of the scenarios is used as the test set, while the remaining data sets compose the training set of the machine learning model. The difference between the training and test loss, utilizing the cross-scenario approach can be seen in Fig. 5. The training and test loss is used as an indicator for the achieved generalization across data sets. If the difference is zero, the trained model is well-tuned for the problem and thus the unseen data. It can be observed that the `GER Campus` subset is well generalized if the remainder of the data subsets are used for training. This includes the use of an inherent difference data source, i.e. the `DK Campus` subset. The generalization achieved for both the `GER Suburban` and `GER Urban` scenarios are similar, however with a decrease in generalization performance compared to the `GER Campus` subset. The generalization achieved across data sources is visualized by the difference in training and test error of the `DK Campus` subset. In other words, the model is trained on a collection of subsets that all originate from the same data source. However, evaluated on a subset of data with a different data origin.

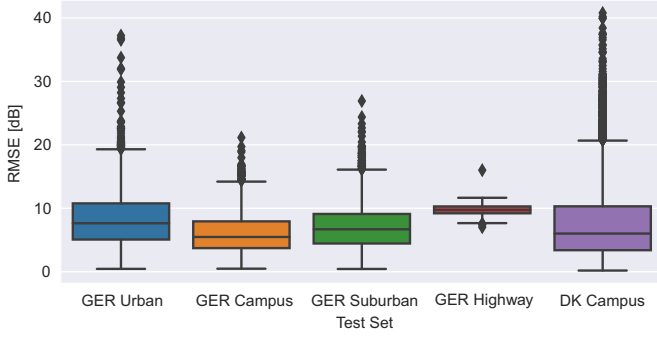The cross-scenario results for all subsets can be found in

Fig. 6. Cross-scenario prediction performance. For each test set, the training set is composed of the aggregation of all remaining data sets.

Fig. 6. It shows the performance of the proposed approach evaluated on each individual subset and trained on the remainder of the available subsets. In other words, the performance of the `GER Urban` scenario based on a model trained on all other subsets excluding the `GER Urban` scenario. The best generalization, also in terms of predictive performance, is achieved on the `GER Campus` subset of 6.3 dB RMSE. Furthermore, the `GER Campus` predictions offer a significant reduction in the standard deviation of $\sigma = 3.6$ dB compared to the `DK Campus` scenario, $\sigma = 6.0$ dB. The highway subset achieves the worst predictive performance of 9.7 dB, however, does have a reduced $\sigma = 1.2$ dB.

A heatmap of `RSRP` at 2630 MHz can be observed in Fig. 7. The model is trained on all data points in the `DK Campus` scenarios. A grid of features is generated for all locations on the map. The model is then evaluated with respect to all generated features. The resulting predictions show no indication of severe overfitting. There is observed no isolated areas where the predictions are non-feasible in terms of `RSRP` magnitude. The range of predicted `RSRP` values, span from -80 to -140 dBm, with a strong increase in signal strength observed near the eNB location.

C. Comparison of Image Types

Due to the simplicity of the images and the reduced complexity of the model, significant studies of how distance and local variability is embedded in the images have been conducted. The initial intuition is that the model is learning a correction related to large-scale fading impairments using the images for deducing local variability. However, the magnitude of the large-scale fading is a complex interaction of the objects (e.g. buildings and other) in the environment. Thus, instead of providing the model with an image spanning only the local area of the measurement, a full-size image can instead be given. The performance of doing so can be observed in Fig. 8. The so-called full-size image is an image in which the receiver and transmitter are both localized. In other words, if the antenna separation increases the area spanned by the image increases. The performance is compared to so-called regular images, in which the area spanned is kept constant at 250 meters around the measurement position. A similar
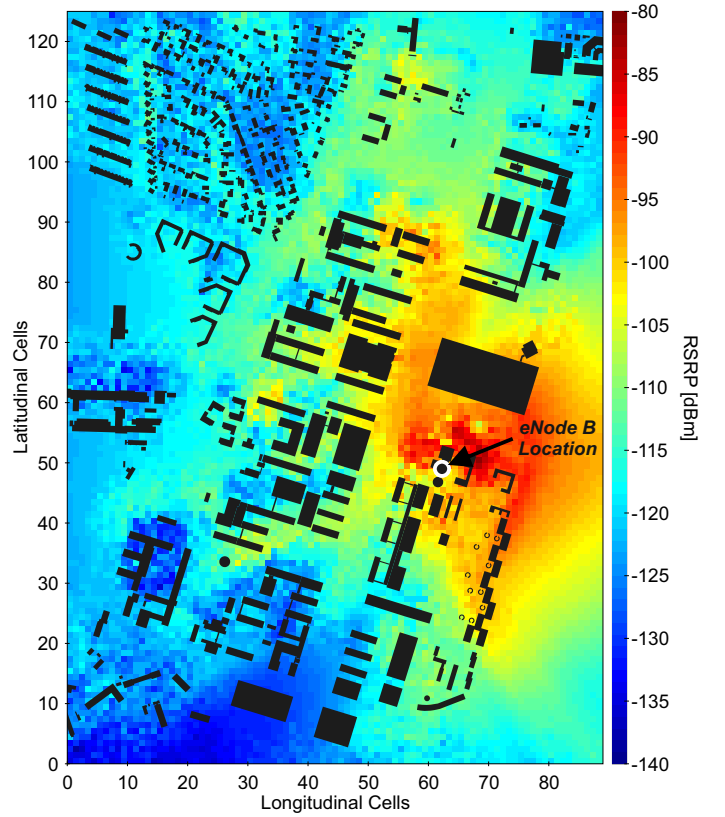

Fig. 7. Resulting RSRP radio environmental map for the `DK campus` scenario generated by the proposed method. (Map data: ©OpenStreetMap contributors, CC BY-SA).

average performance is observed. The distribution of the RMSE evaluated per batch is shown as a kernel density estimation for both cases. The distribution of the RMSE using the full-size images is noticeably different. More specifically, the number and range of the outliers (i.e. predictions with a high magnitude of error) is increased. A RMSE of 6.3 dB is observed for using the *regular* images compared to 7.7 dB for using the *full size* images.

A model was trained using images spanning different distances. An increase in distance results in more area covered by the image. The results of adjusting this distance can be seen in Fig. 9. The best performing model was obtained using images spanning a distance of 250 meters, and similar predictive performance was obtained using images spanning a distance of 300 meters.

VI. CONCLUSION

In this paper, we presented a novel deep learning-based approach utilizing simple geographical images and expert knowledge for improving signal strength predictions in unseen locations. The approach is validated on a comprehensive collection of measurement campaigns. Latent features describing radio characteristics can be implicitly learned from geographical images, spanning a constant distance ($250-300$ meters across) rotated towards the transmitter location. These images contain only information on buildings in the local environment o the
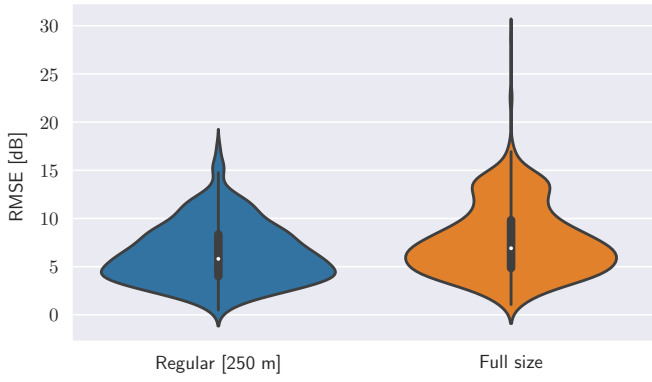
Fig. 8. Comparison of different image variants for the path loss estimation process. *Regular* images are receiver-centric and only contain the UE location. For the *Full size* variant, the images contain the whole transmission path between UE and eNB. Evaluated on the `GER Campus` subset, trained on the remainder.
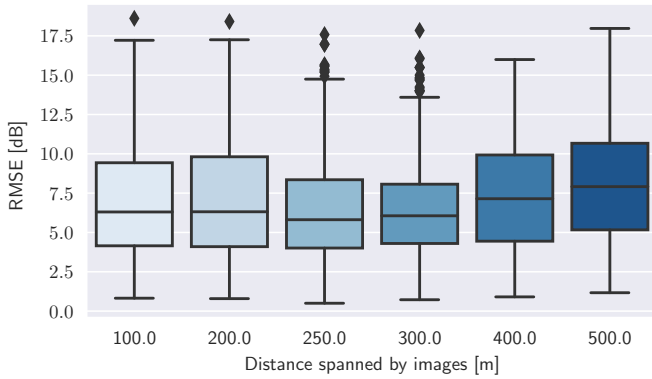

Fig. 9. Comparison of different spanning distances for the receiver-centric images. Evaluated on the `GER Campus` subset, trained on the remainder.

UE. The proposed method is assisted by expert knowledge, ensuring optimal training conditions and improved prediction accuracy. It is shown that the proposed approach is effective in predicting signal strength parameters in terms of predicting the RSRP for unseen locations. Specifically, this results in an RMSE of ≈ 6 dB over inherently different measurement campaigns. In future work, we want to extend the proposed method by integrating height profile information for enabling 3D-based signal strength prediction.

ACKNOWLEDGMENT

REFERENCES

[1] 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.

[2] E. R. Cavalcanti, J. A. R. de Souza, M. A. Spohn, R. C. d. M. Gomes, and A. F. B. F. d. Costa, "VANETs' research over the past decade: Overview, credibility, and trends," *SIGCOMM Comput. Commun. Rev.*, vol. 48, no. 2, pp. 31–39, May 2018.

[3] N. Bui, M. Cesana, S. A. Hosseini, Q. Liao, I. Malanchini, and J. Widmer, "A survey of anticipatory mobile networking: Context-based classification, prediction methodologies, and optimization techniques," *IEEE Communications Surveys & Tutorials*, 2017.

[4] B. Sliwa, R. Falkenberg, T. Liebig, N. Piatkowski, and C. Wietfeld, "Boosting vehicle-to-cloud communication by machine learning-enabled context prediction," *IEEE Transactions on Intelligent Transportation Systems*, Jul 2019.

[5] 5GAA, "White paper: Making 5G proactive and predictive for the automotive industry," 5G Automotive Association, Tech. Rep., Jan 2020.

[6] R. Falkenberg, B. Sliwa, N. Piatkowski, and C. Wietfeld, "Machine learning based uplink transmission power prediction for LTE and upcoming 5G networks using passive downlink indicators," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*, Aug 2018, pp. 1–7.

[7] B. Sliwa and C. Wietfeld, "Data-driven network simulation for performance analysis of anticipatory vehicular communication systems," *IEEE Access*, vol. 7, pp. 172 638–172 653, Nov 2019.

[8] J. Thrane, D. Zibar, and H. L. Christiansen, "Model-aided deep learning method for path loss prediction in mobile communication systems at 2.6 GHz," *IEEE Access*, vol. 8, pp. 7925–7936, 2020.

[9] 3GPP, "3GPP TR 38.901 - 5G; Study on channel model for frequencies from 0.5 to 100 GHz," 3rd Generation Partnership Project (3GPP), Tech. Rep. 38.901, Dec 2019, version 16.1.0.

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

[11] T. Pögel and L. Wolf, "Optimization of vehicular applications and communication properties with connectivity maps," in *2015 IEEE 40th Local Computer Networks Conference Workshops (LCN Workshops)*, Oct 2015, pp. 870–877.

[12] V. Raida, P. Svoboda, M. Lerch, and M. Rupp, "Crowdsensed performance benchmarking of mobile networks," *IEEE Access*, vol. 7, pp. 154 899–154 911, 2019.

[13] X. Wang, M. Umehira, B. Han, P. Li, Y. Gu, and C. Wu, "Online incentive mechanism for crowdsourced radio environment map construction," in *ICC 2019 - 2019 IEEE International Conference on Communications (ICC)*, 2019, pp. 1–6.

[14] J. Wang, C. Jiang, H. Zhang, Y. Ren, K. Chen, and L. Hanzo, "Thirty years of machine learning: The road to pareto-optimal wireless networks," *IEEE Communications Surveys Tutorials*, pp. 1–1, 2020.

[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[16] 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 Vehicular Technology Magazine*, vol. 14, no. 3, pp. 60–69, Sep. 2019.

[17] R. Enami, D. Rajan, and J. Camp, "RAIK: Regional analysis with geodata and crowdsourcing to infer key performance indicators," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, April 2018, pp. 1–6.

[18] U. Masood, H. Farooq, and A. Imran, "A machine learning based 3D propagation model for intelligent future cellular networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*, Dec 2019, pp. 1–6.

[19] B. Sliwa, M. Patchou, and C. Wietfeld, "Lightweight simulation of hybrid aerial- and ground-based vehicular communication networks," in *2019 IEEE 90th Vehicular Technology Conference (VTC-Fall)*, Honolulu, Hawaii, USA, Sep 2019.

[20] B. Sliwa and C. Wietfeld, "Empirical analysis of client-based network quality prediction in vehicular multi-MNO networks," in *2019 IEEE 90th Vehicular Technology Conference (VTC-Fall)*, Honolulu, Hawaii, USA, Sep 2019.

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