

Received 6 July 2022, accepted 21 July 2022, date of publication 25 July 2022, date of current version 29 July 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3193486

APPLIED RESEARCH

An Ubiquitous 2.6 GHz Radio Propagation Model for Wireless Networks Using Self-Supervised Learning From Satellite Images

MARCO SOUSA^{1,2,3}, PEDRO VIEIRA^{1,4}, MARIA PAULA QUELUZ^{1,2},
AND ANTÓNIO RODRIGUES^{1,2}, (Member, IEEE)

¹Instituto de Telecomunicações, 1049-001 Lisbon, Portugal

²Departamento de Engenharia Electrotécnica e de Computadores, Instituto Superior Técnico, 1049-001 Lisbon, Portugal

³Celfinet-Outstanding Networks, 1495-764 Lisbon, Portugal

⁴Departamento de Engenharia Electrónica e Telecomunicações e de Computadores, Instituto Superior de Engenharia de Lisboa, 1959-007 Lisbon, Portugal

Corresponding author: Marco Sousa (marco.sousa@celfinet.com)

This work was supported by the COMPETE/Fundo Europeu de Desenvolvimento Regional (FEDER) through the Project AI4GREEN 16/SI/2019-I&DT Empresarial (Projetos Copromoção) and the International Project CELTIC-NEXT/EUREKA under Grant C2018/1-5.

ABSTRACT The performance of any Mobile Wireless Network (MWN) is dependent on the appropriate level of radio coverage, with Path Loss (PL) models being a valuable resource for its evaluation. Recently, advancements in Machine Learning (ML) and Deep Neural Networks (DNNs) have been applied to radio propagation to produce new data-driven PL models. Notoriously, these advancements have also allowed the inclusion of non-classical inputs, such as satellite images. However, data-driven PL models are often developed under the assumption that training and test data distributions are similar, which is a weak assumption in real-world scenarios. Thus, generalization (*i.e.*, the model's ability to perform on different data distributions) is a crucial aspect of data-driven PL models in the context of Mobile Network Operators (MNOs). This paper proposes a new data-driven PL model, the Ubiquitous Satellite Aided Radio Propagation (USARP) model, developed to enhance the geographical generalization capabilities of empirical PL models, by using satellite images. The USARP model considers self-supervised learning to extract general data representations of the radio environment from satellite images, improving the PL prediction Root Mean Square Error (RMSE) of the 3rd Generation Partnership Project (3GPP) PL model in the order of 9 dB, and for a data distribution distinct from the training data. Moreover, it was demonstrated the potential of the USARP model in terms of geographical and radio environment generalization. Although the generalization capabilities of ML regression algorithms are limited, the chosen USARP architecture and the use of regularization techniques had a positive impact on its geographical generalization performance.

INDEX TERMS Wireless networks, radio propagation, path loss models, satellite data, deep learning, self-supervised learning, convolutional neural networks.

I. INTRODUCTION

MNOs are continuously managing their MWNs, from initially planning the deployment of new Base Stations (BSs), to monitoring the existing network infrastructure and optimizing its performance. The network planning phase not only impacts the MNOs Capital Expenditure (CapEx) but also

The associate editor coordinating the review of this manuscript and approving it for publication was Pietro Savazzi¹.

their Operating Expense (OpEx), as the network optimization stage depends on the reliability of the network planning phase. Considering Radio Access Networks (RANs), the planning phase aims to guarantee coverage, capacity, and Quality of Service (QoS) requirements, with the least amount of investment (*e.g.*, minimizing the number of BSs). Still, the ability to estimate coverage accurately is of paramount importance in the development of successful RAN planning [1].

During the RANs optimization phases, the use of Drive Test (DT) data [2], geopositioning of network traces [3], or crowdsourcing data [4], provide accurate data to evaluate and optimize the radio coverage and the network QoS. However, during initial RAN planning phases, PL models are the primary option to estimate and evaluate coverage.

The use of PL models introduces a higher coverage estimation error than the other “signal level” data sources (e.g., DT measurements), but they are the only existent option in the RAN planning phase for coverage prediction. Different levels of PL prediction accuracy can be obtained from different PL models; however, the most accurate PL models tend to be highly computational expensive and require extensive and detailed environment data [5], which limits their practical applicability. Furthermore, the continuous advancements in ML and DNNs are providing the fundamentals for the development of new data-driven PL models [6], [7], where satellite-based data is also being considered as an additional input. The goal is to achieve higher prediction accuracy than conventional (empirical) PL models without introducing excessive computational complexity or requiring extensive environment data. Nonetheless, the generalization capability—i.e., the ability to learn from a limited volume of data and perform similarly in an out-of-distribution data—of ML or deep learning-based models is still being investigated [8]. Moreover, PL models, when calibrated with DT measurements, generally require a specific calibration for each propagation environments [9].

This paper aims to study the geographical generalization capabilities of empirical PL models, including ML/DNN based ones, towards developing ubiquitous PL models, which can be applied to multiple radio propagation environments without re-calibration or training. Therefore, a novel DNN-based model, the USARP model for PL prediction, is proposed; it uses satellite images with a self-supervised methodology to increase the PL prediction accuracy, enhancing the geographical generalization towards breaking single environment usage restraints of empirical PL models. DT data from real Long Term Evolution (LTE) networks was extensively used for the development and assessment of the USARP model.

The main contributions of this paper are summarized as follows:

- Geographical generalization analysis of empirical and ML/DNN based PL models using data distributions distinct from the initial training data.
- Proposal of a two-stage development process for DNN-based PL models using satellite images, namely: 1) use of self-supervised learning to learn radio environment representations from satellite images; 2) employment of the radio environment representations together with DT measurements, for PL prediction.
- Proposal of a new data-driven PL prediction model—the USARP model—based on the previous two-stages procedure, and its architecture optimization and

TABLE 1. Main notation used in this paper.

Notation	Definition
α	Coefficient denoting the PL dependence on distance
β	PL optimized offset value
γ	Coefficient denoting the PL dependence on frequency
n	Coefficient denoting the PL dependence on distance
$\chi\sigma$	Zero-mean Gaussian distributed random variable with standard deviation σ , describing shadow fading fluctuations
DS_{LoS}	Delay spread mean for LoS radio links
DS_{NLoS}	Delay spread mean for NLoS radio links
d_{3D}	3D distance between a BS and a User Equipment (UE)
f_c	Carrier frequency
G_{BS}	BS antenna gain
G_{UE}	UE antenna gain
h_{eff}	Radio link effective height
h_{BS}	BS antenna height above ground
h_{TBS}	Terrain height above sea level at the location of the BS
h_{UE}	UE antenna height above ground
h_{TUE}	Terrain height above sea level at the location of the UE
MPL	Calculated PL from a DT measurement
\widehat{MPL}	Predicted PL from a DT measurement
PL^{ABG}	Alpha-Beta-Gamma PL model prediction
PL^{CI}	Close In PL model prediction PL
PL^{USARP}	USARP PL model prediction
P_{RS}	Reference signal transmitted power
$RSRP$	Reference signal received power

evaluation in multiple radio environments, towards a single (multi-environment) PL model solution.

This paper is organized as follows. After the introduction, Section II overviews classical PL models and the recent work on the development of PL models using satellite images. Section III gives a brief description of the data considered in this work (satellite and DTs). Section IV explains how the useful information from satellite images is extracted for PL predictions. First, a brief background on self-supervised learning is provided. Then, the self-supervised algorithm used in the scope of this work is presented, along with the results obtained by applying it to satellite data. In Section V, the error metrics to evaluate the PL predictions are firstly defined. Then, the process leading to the development of the USARP model is presented. Section VI evaluates the PL prediction results of the USARP model, and provides a comparison to benchmark PL models. Section VII analyses the geographical generalization capability of the USARP model towards its use in multiple radio propagation environments. Finally, Section VIII presents the main conclusions and final remarks. The main notation adopted in this paper is summarized in Table 1.

II. RELATED WORK

This section overviews related work on radio propagation, notably classical PL models and PL models using satellite images. First, a classification of classical radio PL models is presented, highlighting the base structure of empirical models, which are used as a reference throughout this work. Then, the most relevant work in PL models using satellite images, and typically resorting to deep learning algorithms, is presented.

A. CLASSICAL PATH LOSS MODELS

PL models for MWNs are broadly categorized into two classes: large-scale and small-scale (or fading) models. Large-scale PL models predict the mean strength of the received signal, and small-scale models characterize the rapid fluctuations occurring in a distance of a few wavelengths or on very short time intervals [10].

A UE, with just slight motion, may experience severe signal strength oscillations, as the instantaneously received signal strength results from the contribution of several Multipath Components (MPCs) with distinct directions and random phases. This behavior is known as small-scale fading and may originate signal level fluctuations in a range of 30 dB, for distance differences comparable to the signal wavelength. Small-scale PL models attempt to predict the received signal strength under these circumstances. As the UE moves away from the BS, the local average received signal strength decreases, which is what large-scale PL models predict [10].

Depending on the modelling approach, PL models can be also classified as either deterministic or empirical; while deterministic PL models are derived from the electromagnetic theory (*e.g.*, Maxwell equations), empirical models are obtained by curve fitting from extensive DT signal strength measurements. The deterministic models may apply to various scenarios, by taking into account the reflection and diffraction laws in the PL prediction; therefore, they tend to achieve higher accuracies in the PL prediction than other modeling approaches. However, they have high computational complexity (*e.g.*, require Ray Tracing (RT) or Ray Launching (RL) techniques) and usually demand precise 3-Dimensional (3D) environment information. On the contrary, empirical PL models are mathematically tractable and do not require 3D environment data, despite tending to exhibit lower PL prediction accuracy than the deterministic counterpart [11]. Moreover, as empirical models consider all environmental impacts subjacent on the signal measurements of the respective area, they have higher accuracy in environments similar to the original measurements area [12].

For radio coverage estimation in large areas, empirical models are preferred due to their computational efficiency. The empirical PL models are mostly based on the Alpha-Beta-Gamma (ABG) or the Close In (CI) equations. The PL ABG equation, also known as Floating Intercept (FI), is dependent on the frequency and on the distance, according to [13]:

$$PL^{ABG}(f_c, d_{3D}) = 10\alpha \log_{10}(d_{3D}) + \beta + 10\gamma \log_{10}(f_c) + \chi_\sigma^{ABG} \quad (1)$$

where α and γ are coefficients denoting the dependence of PL on distance and frequency, respectively, whereas β is an optimized offset value. The variable d_{3D} is the 3D distance between the BS and the UE in meters, f_c is the carrier frequency in GHz, and χ_σ^{ABG} is a zero-mean Gaussian distributed random variable with a standard deviation σ , describing the Shadow Fading (SF) signal fluctuations. The coefficients α ,

β and γ are obtained directly from real signal measurement campaigns, fitting (1) to the measured data.

The CI PL equation is given by [13]:

$$PL^{CI}(f_c, d_{3D}) = FSPL(f_c, 1\text{ m}) + 10n \log_{10}(d_{3D}) + \chi_\sigma^{CI} \quad (2)$$

where n is the Path Loss Exponent (PLE), and the only parameter that can be used for the model calibration, the $FSPL(f_c, 1\text{ m})$ is the Free Space Path Loss (FSPL) at a BS-UE separation of 1 m and carrier frequency f_c in GHz, and χ_σ^{CI} is a zero-mean Gaussian distributed random variable with a standard deviation σ (SF).

The 3GPP TR 38.901 model [14] is an example of a ABG-based PL model. Its latest version is valid for a wide range of carrier frequencies (f_c), ranging from 0.5 GHz to 100 GHz (including the entire 5th Generation (5G) spectrum), and for a limited number of propagation scenarios [15]. This PL model separates Line-of-Sight (LoS) from Non-Line-of-Sight (NLoS) propagation, with specific PL equations and parameters for each propagation condition. Moreover, it considers additional variables, such as the BS and UE heights.

Many other empirical PL models are available in the literature, from the more classical models to the new 5G compliant models. Examples of the more classical PL models include the Okumura-Hata model [16] or the Lee model [12], while the Millimetre-Wave Based Mobile Radio Access Network for Fifth Generation Integrated Communications (mmMAGIC) [17] or the NYUSIM [18] are PL models developed for the 5G.

B. SATELLITE-BASED PATH LOSS MODELS

The incorporation of satellite images in radio propagation modeling has been gradually proposed in the last years, fueled by advances in the computer vision field. In [19], the authors proposed the use of satellite images to predict radio channel parameters (PLE and SF) for a given area. The data used to train the model was supported by a deterministic PL model for an Unmanned Aerial Vehicle (UAV) scenario (with a transmitter antenna height of 300 m and a carrier frequency of 900 MHz). Accuracies of 88% and 75% in predicting the PLE and SF, respectively, were reported. The authors proposed the use of pre-trained Convolutional Neural Networks (CNNs), despite being pre-trained on an image dataset very distinct from satellite images, composed of objects, animals, vehicles, and others [20].

In [21], the authors proposed a CNN-based deep learning model for PL estimation using images with building footprints. The PL measurements to train the model were obtained using a deterministic PL model, considering a 900 MHz frequency and an antenna height of 35 m. After training the model, a Mean Square Error (MSE) of 19.52 dB was reported between the ground truth (using the deterministic PL model) and the predicted PL. The authors reported that the proposed model could adapt to modified environments; however, no results have been presented to support that claim.

In [22], a deep learning model that also considers satellite images as input was proposed to estimate LTE signal

metrics namely, Reference Signal Received Power (RSRP), Reference Signal Received Quality (RSRQ), and Signal-to-Interference plus Noise Ratio (SINR). The model was developed with real signal measurements, limited to three BSs, and was composed of a CNN to process the image data, and a Neural Network (NN) to process the radio propagation variables (*e.g.*, distance between the UE and the BS). For each training signal measurement, a satellite image (centered on the UE location) is required. The authors reported a MSE of 7.7 dB between measured RSRP and the proposed model predictions. The authors evaluated the generalization performance of the proposed model by considering training and testing data, but both datasets resulted from the same signal measurements distribution and same locations.

In [23], the authors proposed the use of satellite images to estimate PL with a deep learning model. The model was composed of the 3GPP Urban Macro (UMa) PL model and a correction term generated by a DNN. The DNN contains a CNN to extract features from satellite images, a NN to process radio propagation variables, and a final NN that estimates the PL given the features learned by the previous two modules. The proposed model was trained and evaluated using real LTE PL measurements from three BSs and two distinct carrier frequencies. The authors demonstrated that the use of satellite images provided a reduction of 0.8 dB on the RMSE considering the ground truth PL and the model predictions. Nonetheless, the limited amount of data prevented to derive considerations about the generalization capacity of the model. In [24], the same authors expanded the results of [23] (with some model adjustments) by considering a dataset containing 125000 PL measurements from five distinct environments, allowing to further evaluate the generalization capabilities of the model. The authors reported a prediction RMSE of around 6 dB for unseen locations. However, in the latter work, the proposed DNN model was used to estimate RSRP and not PL.

Overall, the proposed PL models use CNNs to extract features from images. Transfer learning has been applied in [19], but most PL models were trained end-to-end (where a model learns all the parameters of the different modules simultaneously) [21]–[24]. PL models considering only images as input have been proposed in [19], [21], [25], along with mix approaches considering both image-based features and already known radio propagation variables [22]–[24]. Moreover, in several contributions [22]–[24], real signal measurements were used to develop the propagation models. Nevertheless, the used measurements tend to be limited in number and type of environment, limiting the generalization analysis of the proposed models.

This work proposes a new DNN-based PL model using both satellite images and radio propagation variables as input, where the satellite images are used as a complementary data source to increase PL prediction accuracy. The proposed model uses pretraining to enhance its geographical generalization capabilities but, instead of transfer learning, it uses a

self-supervised paradigm, which has demonstrated promising results on several applications [26]. To the best of the author's knowledge, this work constitutes the first application of self-supervised learning to data-driven PL models. Moreover, this work is supported by data from a live network with extensive PL measurements obtained from multiple BSs, in distinct radio propagation environments. Related work has been generally supported by simulated data or limited measurements, restricted to the same geographical area. Furthermore, this work analyzes the geographical generalization capability of data-driven PL models, a topic that has received limited contributions in the related literature, culminating with the proposal of a PL model with enhanced geographical generalization capability.

III. SATELLITE AND DRIVE TEST DATA

In this section, the data that supported the development of this work is presented, comprising the description of the used satellite data and the procedures to obtain the PL from DT measurements.

A. SATELLITE DATA

In this work, the used satellite images cover an area of 194 km², encompassing a mix of urban/suburban environments, along with some areas dominated by vegetation and trees (see Fig. 1). The images were stored in Geospatial Tagged Image File Format (GeoTIFF) format files, being already georeferenced with the same coordinate system used in the DT data, have a pixel resolution of 5 m × 5 m (after sub-sampling), corresponding to Visible Satellite Images (VSIs) with three bands, setting the Red, Green, Blue (RGB) color information, and were cropped with a size of 2 km × 2 km. The reference area, depicted in Fig. 1, contains the geographical area correspondent to the DT data used throughout this work.

B. DRIVE TEST DATA

For this work, DT measurements from a live LTE network were used. The DT data, including coordinates, RSRP, and Physical Cell Identity (PCI), was obtained from 23 BSs operating with a carrier frequency of 2.6 GHz. Moreover, a binning approach [27] considering squared areas of 10 m × 10 m (bin) was carried out to preprocess the DT measurements. Therefore, for each bin, the average value of the RSRP values and the coordinates for each PCI were used, resulting in 23936 DT measurements.

Afterwards, the PL of each measurement, MPL, was computed, in dB, as:

$$MPL|_{[dB]} = P_{RS}|_{[dBm]} + G_{BS}|_{[dBi]} + G_{UE}|_{[dBi]} - RSRP|_{[dBm]} \quad (3)$$

where P_{RS} is the Reference Signal (RS) transmitted power in dBm, and G_{BS} and G_{UE} are the BS and the UE antenna gains in dBi, respectively. The G_{BS} was computed using the 3GPP antenna model [14], using the datasheet parameters of the



FIGURE 1. Reference area including urban and suburban environments along with some open areas [31].

real antenna: vertical and horizontal Half-Power Beamwidth (HPBW), Front-To-Back Ratio (FTBR), Side-Lobe Level (SLL), and the antenna maximum gain.

One of this paper objectives is to evaluate the geographical generalization capabilities of the PL models, when used in locations distinct from those where the PL models were initially calibrated or developed. So, the considered DT measurements were split in three datasets: train, validation, and generalization. From a central location of the reference area (cf. Fig. 1), DT measurements from 14 BSs were retrieved and randomly divided between the training and the validation sets, using a ratio of 80%/20%. The training set (with 11293 measurements) was used to calibrate the PL models and to develop data-driven approaches using ML algorithms, namely, Linear Regression (LR), Support Vector Regression (SVR) [28], Random Forest Regression (RFR) [29], and Light Gradient Boosting Machine (LightGBM) [30] regression. While the LR provides a linear model similar to the structure of the widely used ABG PL model, the remaining allow for exploring non-linear and more complex regression models. The validation set (2824 measurements) was used to evaluate the PL models' accuracy on similar conditions to the training data. Finally, the generalization set contains 9819 DT measurements from nine BSs, from other locations. The generalization dataset is used to evaluate the PL models' accuracy when applied to locations distinct from the training ones, providing insights into the location dependability of a PL model. The three datasets are represented geographically in Fig. 2.

The PL measurements for the train, validation, and generalization sets are depicted in Fig. 3 as a function of the 3D distance between the BS and the UE. Besides, this figure also presents the normalized histogram of the 3D distance (on the upper part of the figure) and of the PL variable (on the right side of the figure). Individually, the histograms

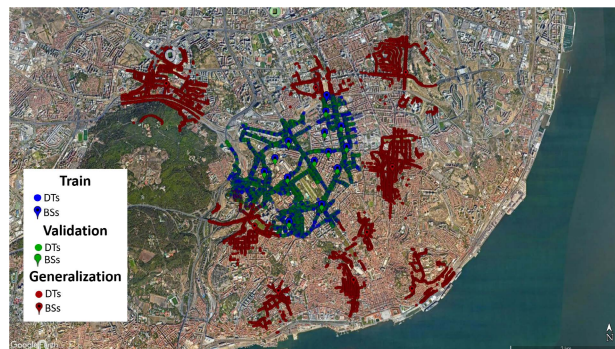


FIGURE 2. Geographical disposition of the train, validation, and generalization DT datasets (based on [31]).

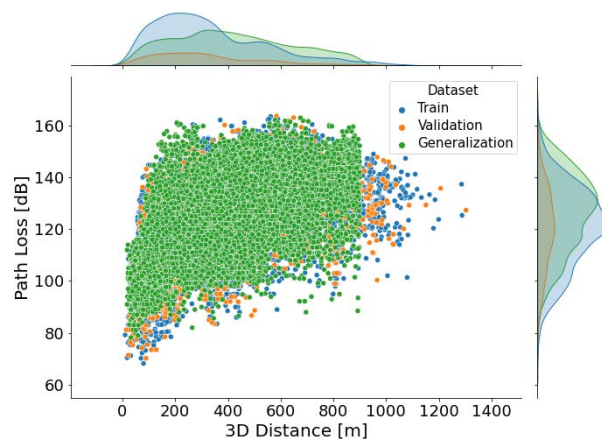


FIGURE 3. Train, validation, and generalization PL measurements as a function of the 3D distance between the BS and the UE.

of the PL values and of the 3D distances are similar for train, validation, and generalization. Moreover, the PL dispersion becomes more evident when evaluating the PL as a function of the 3D distance, *i.e.*, for a fixed distance the PL can vary substantially, which is a consequence of the distinct environments and the different radio link conditions (*e.g.*, LoS or NLoS). Note that the 3D distance information was not present on the measurements dataset but calculated using the the BS and measurement point coordinates, including terrain elevation at the respective positions.

IV. SELF-SUPERVISED LEARNING WITH SATELLITE DATA

The development of realistic PL models is influenced by the quality and quantity of PL measurements, as a low number of measurements may fail to properly enclose all the radio propagation mechanisms or adequately characterize the radio environment. However, extensive PL measurements are not always available, limiting the accuracy of the developed model. Considering deep learning PL models that use satellite images as input, if the PL measurements are limited to a few and homogeneous geographical locations, the corresponding satellite images tend to be similar, which is undesirable towards developing models with geographical generalization capacities. With limited data, a deep learning model

can easily overfit to the particularities of the used satellite images and to the specific environment corresponding to the PL measurements area. Therefore, this work proposes to split the problem of PL prediction into two parts: firstly, learn effective representations of the radio environment from satellite images (without supervision), regardless of representing areas with or without PL measurements; secondly, use them with the PL measurements to train the PL model. With this, the generalization capability of a PL model is expected to be enhanced.

This section starts by providing a background on representation learning and CNNs. Representation learning is used in this work to develop DNNs to extract features from satellite images, and CNNs are the prime NN architecture to handle images as data inputs. Then, the self-supervised methodology (a particular representation learning approach) used in this work is presented. Finally, the self-supervised methodology is applied to the satellite data described in section III-A.

A. BACKGROUND

Nowadays, the volumes of produced data are ever increasing, making the manual task of extracting valuable information a huge challenge. An alternative is to automatically extract features from the raw data, for which representation learning has been used successfully throughout the years, particularly in computer vision tasks. The goal of representation learning is to extract a set of general representation features that can be used to increase the performance of downstream tasks, such as data regression [32].

1) SELF-SUPERVISED REPRESENTATION LEARNING

The use of pre-trained models is common in the computer vision field; these are trained for specific tasks in large datasets (*e.g.*, ImageNet [20]) and fine-tuned to new tasks. Firstly, the NN parameters learned from large datasets provide a good initialization of the NN, allowing a faster convergence. Secondly, the hierarchical NN features learned from models using large datasets can prevent overfitting, particularly if the final task has a small dataset.

However, large-scale datasets are expensive and time-consuming when labeling is required, and many problems do not have large enough datasets. This problem is mitigated with self-supervised methods that learn visual features from unlabeled images. Generally, a transformation (*e.g.*, an image rotation) is applied to the unlabeled images and a NN is trained to predict the properties of the transformation. These transformations are known as pretext tasks [33]. Thus, the NN is trained by learning the objective function of the pretext tasks, and new feature representations are discovered in this process. The learned NN parameters associated with the feature representations are carried to other tasks, typically supervised ones, where the available data might be more limited [34].

In self-supervised learning, several pretext tasks have been proposed, designed so that features of the training images have to be captured by a CNN to solve the pretext tasks.

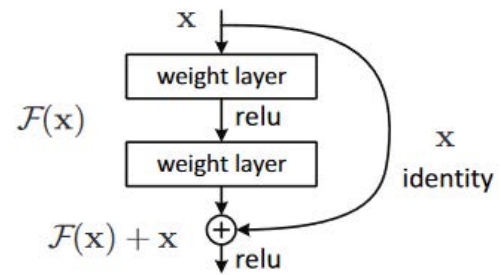


FIGURE 4. Two-layer residual block [36].

At the same time, the pretext task generates a label for each image, according to the applied transformation, making this a supervised problem. According to the taxonomy proposed in [34], pretext tasks can be classified as generation-based, context-based, free semantic label-based, and cross-modal-based. The pretext tasks belonging to the class of generation-based involve image or video generation, forcing the learned features to be relevant for this purpose. Context-based tasks require the learned features to describe context similarity between images, the spatial structure within an image, or the temporal structure for video data, besides others. The free semantic label-based tasks require the automatic generation of semantic labels to train the NN. Finally, cross-modal-based tasks intend to train the NN by verifying if two different input data channels correspond to each other (*e.g.*, video and audio correspondence) [34].

2) CONVOLUTIONAL NEURAL NETWORKS

CNNs have been used in most of computer vision tasks, such as semantic segmentation, object detection, or image classification, achieving state-of-the-art results. This success is tightly associated with the CNNs architecture, which has several advantages over other deep learning architectures, such as the use of local connections [35].

In this work, a particular CNN architecture was used: the ResNet [36]. Its architecture addresses some of the problems associated with deeper NNs (*e.g.*, gradient vanishing) with the introduction of residual connections, as depicted in Fig. 4; these connections, represented by the identity shortcut connection in Fig. 4, propagates the input of a given layer, X , to subsequent layers, which helps the training process of deeper networks, leading to more accurate models. The output of a residual block is the sum of the weight layers mapping function, $\mathcal{F}(X)$, and the respective input, X .

The ResNet architecture has been widely used in several computer vision tasks, and multiple ResNet-based architectures have been proposed (*e.g.*, [37]). In addition, a recent work [38] demonstrated that the original work of the ResNet matches recent state-of-the-art models when using advanced training and scaling methodologies. Therefore, in this work a ResNet architecture—the ResNet50 [36]—was selected to develop the CNN for processing the satellite images.

The ResNet50 is composed of a total of 48 convolution layers, one max-pooling layer, and one average pooling

layer. This particular architecture exploits the benefits of deeper architectures without being too computational complex. The ResNet50 implementation provided in [39] was used.

B. SELF-SUPERVISED MODEL

Several self-supervised learning models have been proposed in the recent literature, most requiring a pretext task to be solved. Within the scope of this paper, the image representations of the radio environment, learned from satellite images, should be relevant to discern the several factors that influence radio PL. Such factors are the existence (or not) of obstructions (*e.g.*, buildings), areas with vegetation, the width of the streets, among others.

According to the pretext task taxonomy provided in [34], the context-based is the most appropriate group of pretext tasks for this work. These can be set for the CNN to predict the relative positions of two patches from the same image, as in [40]. Another pretext task is to predict the rotation angle applied to an image or to recognize the correct order of a sequence of shuffled patches from the same image, also known as puzzle tasks [41], [42]. To accomplish these pretext tasks, CNNs need to learn spatial context information, such as the shape of the objects and the relative positions of different parts of an object [34]. However, a recently proposed methodology for self-supervised learning, called as Bootstrap Your Own Latent (BYOL) [43], has achieved exciting results, outperforming previous models. The main goal of BYOL is to learn image representations that can then be used for downstream tasks, which in the scope of this work is to develop a model for predicting radio path loss. The BYOL architecture includes two NNs, the online and target networks, as depicted in Fig. 5.

The BYOL uses image augmentation to produce two additional views from the original image; by applying random transformations to the input images (*e.g.*, color transformations) it enriches the training data, reduces overfitting, and improves the model generalization [44]. BYOL considers the following transformations for augmenting images: random cropping, left-right flip, color jittering, color dropping, Gaussian blurring, and solarization [43]; then, each additional view is used as input to the online and target networks. While the online network is constituted by an encoder, a projector, and a predictor, the target network contains an encoder and a projector (see Fig. 5). In the original work, the encoder is implemented using a ResNet network (other architectures can be used), and the projectors and the predictor are implemented using Multi-Layer Perceptrons (MLPs). The overall network is trained to minimize the MSE loss between the normalized predictions of the online network ($q_{\theta}(z_{\theta})$) and the target network projections (z'_{ξ}); the weights of the target network result from an exponential moving average of the weights of the online network. After training the global network, the encoder from the online network can be used to generate image representations in downstream tasks.

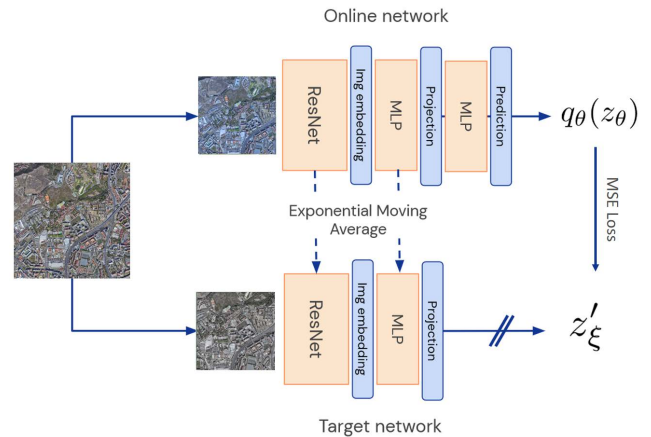


FIGURE 5. BYOL architecture (adapted from [43]).



FIGURE 6. Example of a satellite image used for training, corresponding to an area of $2 \text{ km} \times 2 \text{ km}$ [31].

C. SATELLITE SELF-SUPERVISED LEARNING

In the following, the ResNet50 architecture was used as the encoder for the BYOL self-supervised learning approach, considering the reference area described in section III-A. Therefore, 500 images, with a resolution of 400×400 pixels (each corresponding to an area of $2 \text{ km} \times 2 \text{ km}$) were randomly cropped from the satellite image of the reference area. Fig. 6 presents an example of an image used for training. The BYOL network was trained during 360 epochs, with a learning rate of 3×10^{-4} .

Fig. 7 depicts the MSE between the normalized predictions of the online network and the target network projections, as a function of the epoch number. From this figure, it can be stated that, despite some variability, the network training loss converged.

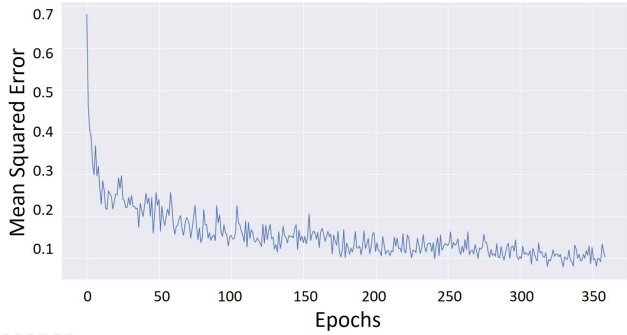


FIGURE 7. BYOL training loss by epoch using the satellite images.

The general success of deep learning, and particularly of CNNs, is achieved at the cost of low interpretability, which is still an active and open research question. However, a simple approach to gain intuition about the trained CNNs is to represent the feature maps of the convolutional layers for a given input image. A feature map is the output of a single filter of a convolutional layer.

Fig. 8 depicts four feature maps (from the first convolutional layer), when the image in Fig. 6 is used as input for the ResNet50 model. Comparing the satellite image with the four feature maps, it can be concluded that each feature map represents different information from the original image; it is also noted that information representing roads, buildings, and open areas is preserved, which is a valuable information for the development of a PL model.

In the next section, the trained ResNet50 model is used to extract features from the satellite images and support the PL predictions.

V. UBIQUITOUS SATELLITE AIDED RADIO PROPAGATION

This section describes the proposed USARP model for PL estimation. Firstly, the data inputs of the USARP model are presented alongside the model base architecture. Then, the model base architecture is optimized to maximize the geographical generalization capabilities of the model. The section ends by presenting the final architecture of the proposed USARP model.

A. USARP INPUTS

The inputs of the USARP model are satellite images, the BS and UE locations, and variables describing the BS to UE radio link, namely the 3D distance in logarithmic scale, $\log_{10}(d_{3D})$, and the radio link effective height, h_{eff} , defined as:

$$h_{eff} = (h_{TBS} + h_{BS}) - (h_{TUE} + h_{UE}) \quad (4)$$

where h_{BS} and h_{UE} are the BS and the UE antenna heights above ground, respectively, h_{TBS} and h_{TUE} are the terrain heights above sea level at the location of the BS and the UE, accordingly.

The inclusion of the satellite images as input of the USARP model is performed as follows:

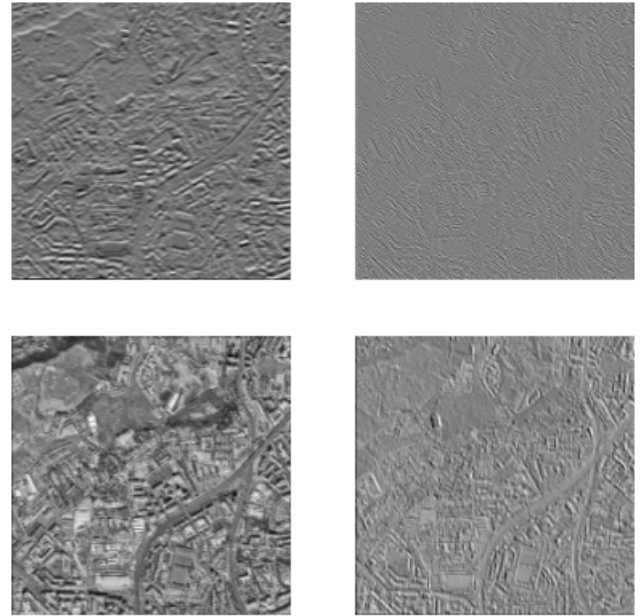


FIGURE 8. Example of feature maps from ResNet50 after trained with the satellite images.

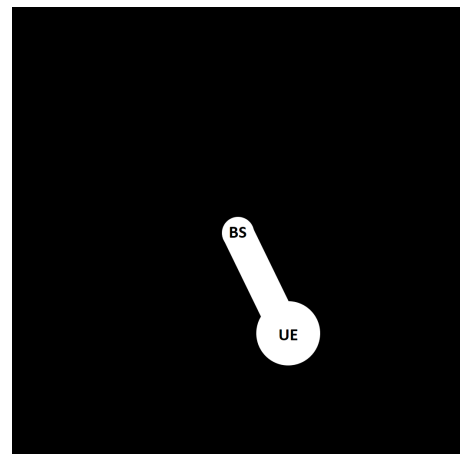


FIGURE 9. Example of an ROI mask with the BS location on the smallest white circle, the UE location on the center of the largest white circle, and the area corresponding to the direct link between both ends.

- The satellite images are centered in the BS.
- A ROI mask is produced to identify the BS and UE locations and the direct radio link region.
- The overlap between the satellite image and the ROI mask keeps only the areas of the satellite images that are relevant for the PL prediction: the BS and UE locations, and the direct radio link region.

The ROI mask works as an attention mechanism, in the satellite image, retaining the key locations to estimate the PL. Fig. 9 presents an example of the ROI mask. In the ROI mask, the radius of the circle identifying the UE location was defined according to the Delay Spread (DS) of a radio signal in an UMa environment and the frequency of the DT measurements (2.6 GHz). According to [14], the DS mean

(in a logarithmic scale) for LoS radio links in an UMa environment is given by:

$$DS_{LoS}[dB] = -6.955 - 0.0963 \log_{10}(f_c) \quad (5)$$

where f_c is the carrier frequency in GHz. For NLoS radio links, the DS mean is given by [14]:

$$DS_{NLoS}[dB] = -6.28 - 0.204 \log_{10}(f_c) \quad (6)$$

The distribution of DS is further characterized by a standard deviation of 0.66 for LoS and 0.39 for NLoS [14]. Considering a reference DS based on one standard deviation from the mean, that accounts approximately for 68% of the DS distribution ($\mu_{DS} \pm \sigma_{DS}$), the corresponding DS distance is 318 m for NLoS and 139 m for LoS. Accordingly, the radius of the circle characterizing the UE locations was defined with the value of 318 m. The radius of the circle surrounding the BS location is smaller (empirically defined as half of the UE circle radius), as the surrounding area of a BS is usually unobstructed, not conditioning the PL, and therefore has less impact on the PL than the surrounding of the UE. The direct link between BS and UE was defined as having a width of half of the UE circle radius, to identify the existence of propagation obstacles between the BS and the UE.

The geo-referencing between the DT measurements and the satellite images, using an ROI mask, is enabled using the GeoTIFF image format with the same coordinate system as the DTs measurements [45]. Furthermore, any error associated with the location of the UE or in the pixel association is mitigated, as the ROI mask also considers the neighboring pixels of the UE, as previously described.

Overall, the training of the USARP model requires DT measurements (see Section III-B) and the respective satellite image data. Firstly, satellite images centered on a BS location were generated for each BS reported in the DT data (see Fig. 6 as an example of such images); then, the ROI masks were created for each pair of BS and UE locations (see Fig. 9).

B. USARP BASE ARCHITECTURE

This section presents the base architecture of the USARP model and explains how its inputs (satellite image, ROI mask, and radio link variables) are considered.

Fig. 10 depicts the base architecture of the USARP model, which was inspired in [46], where the author presented a filter approach for focusing the attention of CNNs on an ROI. The ROI Filter implementation corresponds to an element-wise multiplication between the feature maps resulting from the convolution applied to the satellite image and the ROI mask. This process acts as a hard attention mechanism by discarding the image features that do not belong to the respective ROI. Moreover, after the element-wise multiplication of the satellite image by the ROI mask, the resulting image is rotated so that the UE is at zero degrees relatively to the north direction. Thus, the USARP model is invariant to the direction between the BS and the UE. Furthermore, the image is cropped to

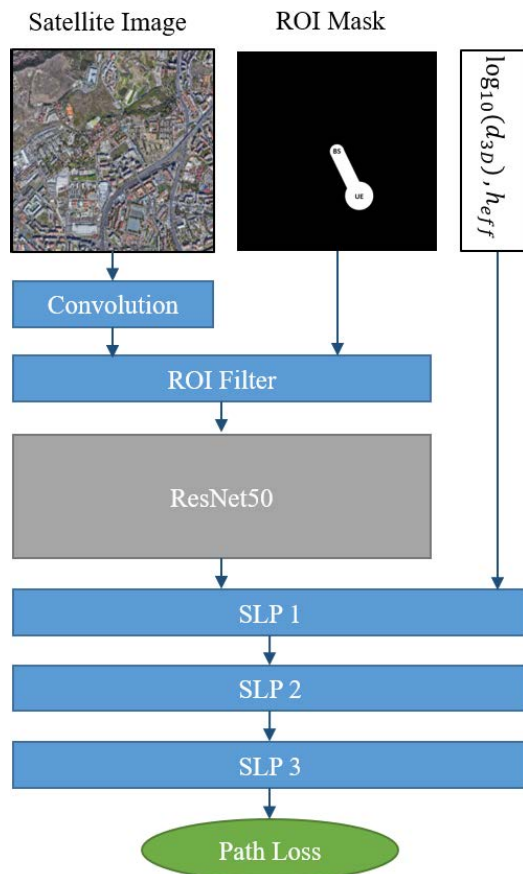


FIGURE 10. Base architecture of the USARP model for PL predictions using satellite images.

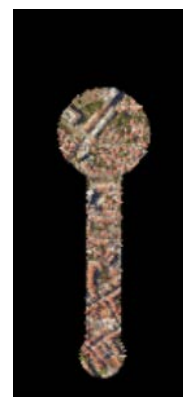


FIGURE 11. Example of the output of the ROI Filter layer.

enclose the ROI mask. Fig. 11 presents an example of the output of the ROI Filter layer.

After, the CNN (ResNet50), trained using the self-supervised learning approach (as described in section IV-C), is used to extract features from the ROI Filter output. The vector of features resulting from the ResNet50, and the radio link variables ($\log_{10}(d_{3D})$ and h_{eff}), are then concatenated. At this point, the resulting feature vector is the input of an MLP network composed of three Single-Layer Perceptrons (SLPs).

Each SLP includes a batch normalization and a non-linear activation (Parametric Rectified Linear Unit (PReLU)). The output of the last layer is the PL prediction at the UE position.

C. USARP ARCHITECTURE OPTIMIZATION

In this section, the base architecture of the USARP model (cf. Fig. 10) was optimized, with the goal of developing a PL model able to generalize to data distributions distinct from the training data distribution. Accordingly, the contributions of the radio link variables, $\log_{10}(d_{3D})$ and h_{eff} , and the satellite images were evaluated by applying some modifications to the base architecture of the USARP model, namely the following:

- The addition of a linear layer that outputs the PL based on the radio link variables and on the features extracted from the satellite images.
- The disabling of the ResNet50 parameters update during training.
- The removal of the convolutional layer having the satellite images as input.

The proposed modifications were evaluated using the training dataset for training purposes, and the validation and generalization datasets to measure the ability of each modification to generalize to new data distributions. For the PL predictions assessment, three error metrics were used, namely the Root Mean Square Error (RMSE), the Mean Absolute Error (MAE), and the Explained Variation Score (EVS). The PL prediction error vector is defined as:

$$e = MPL - \widehat{MPL} \tag{7}$$

where MPL , and \widehat{MPL} are, respectively, the PL ground truth vector and the predicted PL vector. The MAE and RMSE metrics are computed as:

$$MAE = \frac{1}{N} \sum_{i=0}^{N-1} |e_i| \tag{8}$$

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=0}^{N-1} e_i^2} \tag{9}$$

where e_i is the i -th element of e and N is the length of e .

The EVS, which measures the proportion of variation accounted for in a given set of predictions, is computed according to:

$$EVS = 1 - \frac{\text{Var}(e)}{\text{Var}(MPL)} \tag{10}$$

where $\text{Var}(\cdot)$ is the variance function.

All experiments were conducted using fixed training parameters: 50 epochs, a learning rate of 0.5, a batch size of 20, and the MSE as loss function. Also, the model parameters corresponding to the epoch with the lowest validation error were retained for comparison.

The addition of a final linear layer, which acts as a simple linear regression, to predict the PL based on the radio link

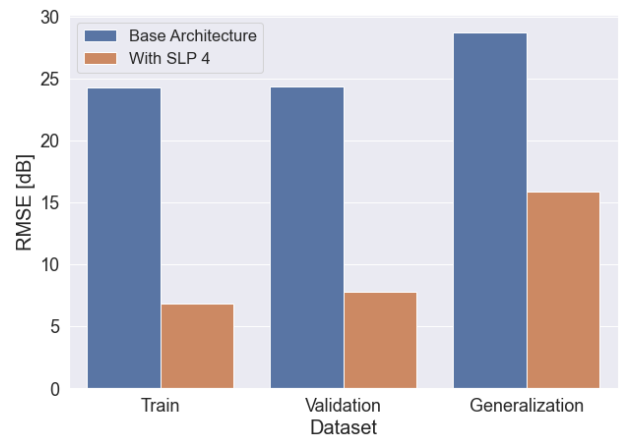


FIGURE 12. PL RMSE of the base architecture and the modified architecture (with the addition of the SLP 4) in the training, validation, and generalization datasets.

variables and the satellite-based features, enforces two constraints: the independency between the contributions of the radio link variables and the satellite-based features, to the PL prediction; the linear dependence of the predicted PL from the radio link variables. These constraints, known to be valid according to the FSPL theory, were not guaranteed in the base architecture. Nonetheless, in related works [23], [24], radio link variables and image-based variables are commonly concatenated before using a NN to estimate the PL. The final linear layer modification corresponds to the introduction of SLP 4 having, as input, the output of SLP 3 and the radio link variables.

The initial and modified architectures were compared using the RMSE between the PL ground truth and the respective PL predictions for each dataset (cf. Fig. 12), allowing to conclude that even without optimizing the training parameters, the modified architecture (using the SLP 4) achieves a lower error than the base one (without SLP 4). Thus, the modified architecture with the addition of a final linear layer is more efficient in solving the PL prediction problem.

The second modification to the base architecture was to disable the update of the ResNet50 parameters during training. Particularly in the computer vision domain, DNN models are commonly trained in one initial task before fine-tuning them into a second task, which is known as transfer learning [47]. In this work, the ResNet50 model was initially trained with extensive satellite images using self-supervised learning (cf. Section IV-C) before being integrated into the USARP model architecture. However, in the PL prediction problem, the images used in the self-supervised stage and the PL prediction stage are from the same source, contrary to typical transfer learning scenarios. Therefore, updating the ResNet50 parameters during the USARP model training may limit the range of radio environment representations already learned. Accordingly, the update of ResNet50 parameters was disabled during the training to assess its impact on the generalization performance.

Finally, the impact of removing the convolutional layer that directly processes the satellite images was also evaluated, as it could be leading to overfitting by over-represent environment properties already included on the training data.

The error metrics of the validation and generalization datasets, for the studied architecture elements, are presented in Table 2; the first and the second rows of this table correspond, respectively, to the base architecture and to the addition of a final linear layer (SLP 4). All error metrics show that the modified architecture achieves a better generalization than the base one; therefore, it is more representative of the PL prediction problem.

The third row of Table 2 shows the errors obtained by disabling the update of the ResNet50 parameters during training (but keeping the newly added SLP 4). Although the performance on the validation set decreases, it increases on the generalization set. So, the use of the ResNet50 with the parameters learned with the BYOL algorithm, as opposed to allowing them to be updated contributes to mitigating satellite image overfitting, resulting in lower generalization errors. The use of a self-supervised algorithm, as the BYOL, enables the incorporation of a wider variety of satellite images, as the existence of DT measurements for the respective areas of the satellite images is not required. Therefore, more representations of radio environments are learned, and the generalization capabilities of the USARP model are increased.

Finally, the last row of Table 2 corresponds to an architecture without the convolution layer that precedes the ROI Filter layer (cf. Fig. 10), but retaining the previous architectural modifications. This architecture achieves the highest performance in the generalization dataset in all metrics. As before, the higher generalization performance is achieved by degrading the validation performance.

Overall, having the radio link variables as input of the last SLP, disabling the update of the ResNet50 parameters during training, and removing the convolutional layer preceding the ROI Filter layer, leads to the highest generalization.

D. USARP MODEL

This section starts by defining the final architecture of the USARP model based on the previous analysis. Then, the hyperparameters of the USARP model are optimized, and regularization methods to further improve the generalization capacity of the proposed model are introduced.

1) FINAL ARCHITECTURE

The architecture analysis presented in Section V-C is reflected on the final USARP model architecture depicted in Fig. 13. Comparing the final architecture with the base architecture (see Fig. 10), the initial convolution layer was removed, as discussed. The ResNet50 produces an output vector with a dimension of 2048, but the three SLPs allow to reduce its dimension. Each SLP includes a batch normalization layer and a PReLU activation. The SLP 4 was added to the base architecture, to enforce the linear association between the radio link variables and the predicted PL.

TABLE 2. Prediction error metrics for the modified architectures of the USARP model in the validation and generalization datasets.

	Dataset	RMSE [dB]	MAE [dB]	EVS
Base Architecture	Val	24.34	19.69	-0.10
	Gen	28.74	23.05	-0.42
With SLP 4 Layer	Val	7.81	6.13	0.75
	Gen	15.89	12.39	-0.04
ResNet parameters update disabled	Val	8.50	6.67	0.70
	Gen	15.39	12.06	0.01
Without Conv. Layer	Val	9.16	7.21	0.65
	Gen	13.88	10.85	0.18

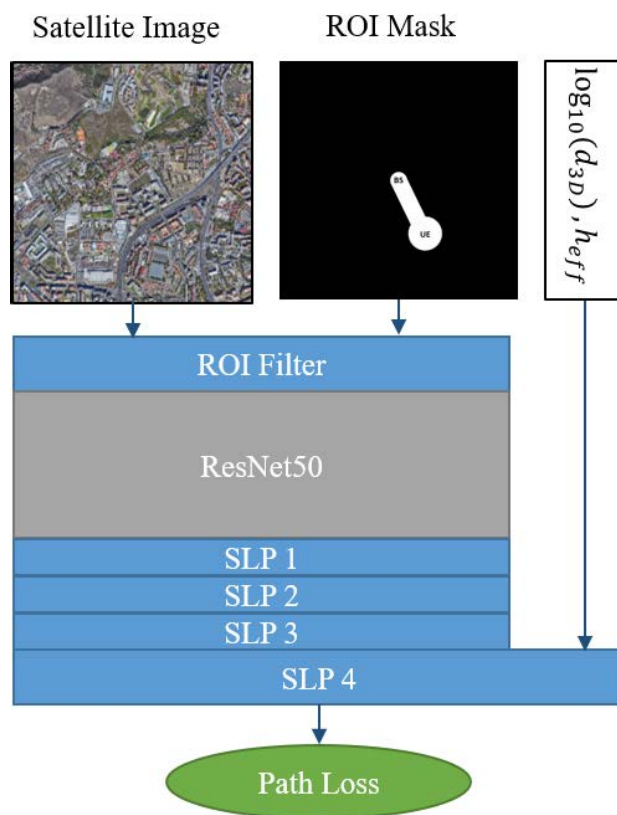


FIGURE 13. Proposed architecture for the USARP model using satellite images.

Formally, the PL predictions of the USARP model, obtained by the SLP 4, can be decomposed as:

$$PL^{USARP}(x, r, s) = f_1(x, r) + f_2(s) \quad (11)$$

where $x \in \{0, \dots, 255\}^{3 \times M \times N}$ is an RGB satellite image with a dimension of $M \times N$ pixels, $r \in \{0, 1\}^{M \times N}$ is a binary image representing the ROI mask, and $s \in \mathbb{R}^2$ is a vector containing the distance and the effective height of the radio link. Functions $f_1(x, r)$ and $f_2(s)$ represent the independent

contributions to the PL predictions of the image-based inputs, (x, r) , and the radio link variables, (s) , respectively.

According to the proposed architecture in Fig. 13, $f_1(x, r)$ is implemented by the ROI Filter layer, the ResNet50, SLPs 1, 2, and 3, and the nodes of the SLP 4 that process the SLP 3 output. Similarly, $f_2(s)$ is implemented by the nodes of the SLP 4 that process the radio link variables and the bias term of the SLP 4. The PL dependence on the radio link variables, $f_2(s)$, is computed by:

$$f_2(s) = \omega_0 + \sum_{i=1}^2 \omega_i s_i \quad (12)$$

where $\omega \in \mathbb{R}^{P+3}$ is a vector with the SLP 4 parameters and P is an hyperparameter of the USARP model indicating the number of output nodes in the SLP 3. Therefore, $f_2(s)$ follows the structure of the widely used ABG PL model. The PL dependence on the image-based variables, $f_1(x, r)$, is calculated as part of the SLP 4, as follows:

$$f_1(x, r) = \sum_{i=3}^{P+2} \omega_i f_4(f_3(T(x \odot s)))_i \quad (13)$$

where $T(\cdot)$ rotates the input image [39], aligning the BS and UE locations vertically (the BS location is always centered in the satellite image), $f_3(\cdot) \in \mathbb{R}^{2048}$ is the output of ResNet50 according to the implementation [39] used in this work, $f_4(\cdot) \in \mathbb{R}^P$ is the output of the MLP (composed of SLP 1, SLP 2, and SLP 3), and \odot is the element-wise product operator. Additionally, each SLP is given by:

$$\text{SLP}_j = g(\text{BN}(W^j u_j + b_j)), \quad j \in \{1, 2, 3\} \quad (14)$$

where $g(\cdot)$ is the PReLU activation, $\text{BN}(\cdot)$ is the batch normalization function, W^j is the j -th layer parameter vector, u_j is the j -th layer input vector, and b_j is the bias parameter of the j -th layer. Finally, $\text{BN}(\cdot)$ is given by (for a batch \mathfrak{B}) [48]:

$$\text{BN}(u') = \gamma \odot \frac{u' - \mu_{\mathfrak{B}}}{\sigma_{\mathfrak{B}}} + \beta \quad (15)$$

where γ and β are the scale and the shift parameters (learned during training), $\mu_{\mathfrak{B}}$ is the sample mean, and $\sigma_{\mathfrak{B}}$ is the sample standard deviation of batch \mathfrak{B} , for $u' \in \mathfrak{B}$.

2) REGULARIZATION AND HYPERPARAMETER TUNING

DNNs can approximate very complex functions due to their large number of parameters and expressiveness. However, they can easily overfit and provide poor generalization. Therefore, regularization techniques have been proposed; one of such techniques is the use of dropout during the DNN training [49]. The dropout consists of randomly ignoring nodes during the DNN training, which prevents single neurons from becoming too specialized, and neighboring neurons too dependent on each other. In the USARP architecture, the dropout was applied to the SLPs 1, 2, and 3.

Another regularization technique, widely used even before DNNs, is the L2 regularization [50]; it adds a penalty to the loss function, which penalizes the magnitude of the learned

TABLE 3. Hyperparameter search space for the USARP model.

Hyperparameter	Minimum	Maximum
Number of Image Features	1	10
Dropout Probability	30%	70%
Regularization Rate	2	100
Learning Rate	0.1	0.001
Number of Epochs	5	50

model parameters. The loss function, $L(w)$, of the USARP model, is given by:

$$L(w) = \frac{1}{N} (MPL - u_4 w)^2 \quad (16)$$

where N is the number of PL measurements, MPL is the measured PL, and u_4 is the input of the SLP 4. Further, the L2 regularization penalty, $L_{\text{reg}}(w)$, is computed only for the parameters corresponding to the output features of SLP 3:

$$L_{\text{reg}}(w) = \lambda \sqrt{\sum_{i=3}^{P+2} w_i^2} \quad (17)$$

where λ is the regularization rate. The rationale is that no overfitting would be associated with the radio link variables but possibly to the image-based variables. Thus, the total training loss is given by:

$$\tilde{L}(w) = L(w) + L_{\text{reg}}(w). \quad (18)$$

Then, a set of hyperparameters of the USARP model were optimized, namely the number of the output nodes of the SLP 3, the dropout probability, the regularization rate, the learning rate, and the number of epochs. The number of output nodes of the SLP 3 represents the final number of image-based features to estimate the PL (first row of Table 3). The dropout probability establishes the probability of ignoring network nodes during training, while the regularization rate, λ , (see (17)) enforces the magnitude of the L2 regularization penalty. Then, the learning rate and the number of epochs governs the network training process.

For the optimization of hyperparameters, an open-source optimization framework, called as Optuna [51], was used. This optimization framework firstly requires defining the search space for each hyperparameter, which is presented in Table 3. The Optuna framework allows the use of various sampling methods for the defined search space. In this work, the Tree-structured Parzen Estimator (TPE) sampling method was used [52], which efficiently explores the hyperparameter search space towards the optimal configuration; also 200 trials (iterations searching the optimal configuration) were conducted. The resulting best hyperparameter configuration is presented in Table 4.

VI. RESULTS

This section starts with the performance assessment of several empirical PL models (considered as benchmark), on the

TABLE 4. Optimal hyperparameter values for the USARP model.

Hyperparameter	Value
Number of Image Features	6
Dropout Probability	54.23%
Regularization Rate	74.57
Learning Rate	0.093
Number of Epochs	20

validation and generalization datasets, establishing the baseline performance for the remainder of the section. Then, the performance of the USARP model is presented and compared with the baseline approaches. Finally, the results of ablations studies performed on the USARP model are analysed.

A. BENCHMARK MODELS

The DT train dataset, presented in section III-B, was used to train data-centric PL models, while the DT validation dataset measures the respective PL prediction performance. Additionally, the DT generalization dataset (see Fig. 2) was used to estimate the PL models' performance in distinct environments (but still similar to the training environments).

Firstly, to gauge the performance of non-calibrated empirical PL models, the 3GPP TR 38.901 model [14] was used to estimate the PL corresponding to the locations of the DT validation data. This model has distinct equations for LoS and NLoS, requiring the classification of each of the considered DT measurements accordingly. The LoS classification was performed deterministically, using terrain and 3D building information [53]. Afterwards, the 3GPP model was applied, and an RMSE of 20.96 dB was obtained, which is within the values reported in [8]. The 3D building data was limited to the train and validation areas, preventing an evaluation of the 3GPP PL model on the locations of the generalization DT data. Nonetheless, the generalization RMSE of the 3GPP PL model is expected to be within the same order of magnitude of the RMSE obtained for the validation dataset, taking into account the similarities of the radio environment.

Secondly, four ML regression-based algorithms were considered to develop data-driven PL models based on the DT training dataset: LR, SVR [28], RFR [29], and LightGBM [30] regression. These algorithms have, as input, the 3D distance between the BS and the UE locations in logarithmic scale, $\log_{10}(d_{3D})$, and the BS effective height, h_{eff} (cf. (4)). Each regression algorithm predicts the PL, \widehat{MPL} , as a function of the 3D distance and effective height:

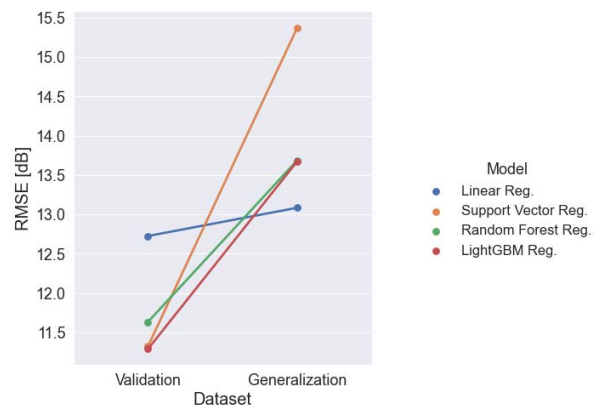
$$\widehat{MPL} = f(\log_{10}(d_{3D}), h_{\text{eff}}) \quad (19)$$

The four data-driven PL models were developed according to the methodology presented in [8].

The overall PL prediction performance of the 3GPP and the data-driven PL models is presented in Table 5, using the error metrics from section V-C in the validation (Val) and the generalization (Gen) datasets. This table shows that, for the validation dataset, all the data-driven PL models outperform the 3GPP PL model, in all error metrics. Considering

TABLE 5. The 3GPP and data-driven PL models' performance for the validation and the generalization datasets.

	Dataset	RMSE [dB]	MAE [dB]	EVS
3GPP	Val	20.96	17.57	0.22
Linear	Val	12.73	10.25	0.32
	Gen	13.08	10.47	0.31
SVR	Val	11.32	8.58	0.46
	Gen	15.37	12.30	0.06
RFR	Val	11.63	9.16	0.43
	Gen	13.69	11.06	0.25
LightGBM	Val	11.29	8.91	0.48
	Gen	13.67	10.95	0.25

**FIGURE 14.** PL RMSE of the linear regression and data-driven PL models in the validation and the generalization datasets.

the validation dataset, the LightGBM-based model achieves the lowest RMSE, and the highest EVS, while the SVR-based model attain the lowest MAE. Notably, although the LR model obtained the highest prediction errors on the validation dataset, it showed the best performance on the generalization dataset. The LR, which is mathematically similar to the ABG equation that supports most empirical PL models, demonstrates identical performance on the validation and on the generalization datasets. Additionally, the non-linear regression algorithms (SVR, RFR, and LightGBM) have a significant performance degradation between the validation and generalization datasets. This comparison is further depicted in Fig. 14 for the RMSE metric. This figure shows that the performance of the data-driven models on the validation dataset is higher than the performance obtained when new data distributions are used. Therefore, unless the trained PL model is intended for making PL predictions on the same area of the training data, LR is preferable over the ML regression algorithms.

B. USARP MODEL

The USARP model was trained using the hyperparameters from Table 4. The resulting performance is presented

in Table 6, showing that this model surpasses the performance of all baseline models, in all metrics and datasets. Notably, it improves the generalization performance in all metrics relatively to the best baseline model, corresponding to the LR.

Fig. 15 allows a direct comparison between the considered models in terms of the resulting RMSE of the PL predictions. It can be stated that the LR still presents the lowest performance degradation between the validation and the generalization datasets. However, the superior expressiveness of the USARP model allows that even with a higher performance degradation (validation to generalization datasets), it still overcomes the LR by almost 1 dB difference in terms of RMSE.

In Fig. 16, the PL predictions of the USARP model are compared with the DT PL measurements, for the generalization dataset; the diagonal red line represents the predictions of an ideal model. From that reference, it can be stated that the USARP model follows the tendency of the measured PL. Also, the PL predictions between 110 dB and 140 dB demonstrate a higher standard deviation, possibly due to the higher volume of PL measurements in that range. Nonetheless, the USARP PL predictions are balanced between overestimating and underestimating the observed PL; the average error between the predicted and the measured PL is -0.01 dB, and the median error is 0.19 dB.

Although the PL measurements used in this work and in other related works are naturally distinct and from different experimental areas, it is still valuable to compare the order of magnitude of the attained PL prediction accuracy without failing to describe the data measurements setup. But more importantly, the adopted methodologies should be compared. For instance, in [21], the authors obtained an RMSE of 4.42 dB between ground truth PL values and the proposed model predictions, a CNN-based model using images with buildings footprints. However, the prediction error was estimated based on PL values obtained by a deterministic PL model, which could lead to a different performance when using real PL measurements. Furthermore, the generalization of the model has not been evaluated. In addition, the proposed model requires one image per PL prediction, while the USARP model requires only one satellite image per BS to make the PL predictions.

In [23], a satellite-based DNN PL model achieved an RMSE around 4 dB using a dataset of real PL measurements on the 2600 MHz band; these measurements were obtained from a single propagation environment (an university campus) using three BSs. The authors also reported an RMSE around 8.5 dB between the 3GPP TR 38.901 model predictions and the PL measurements. This error analysis resulted from PL measurements geographically adjacent to the training data. Therefore, considering the validation set used for the USARP model development, the 3GPP TR 38.901 and the USARP obtained an RMSE of 20.96 dB and 10.71 dB, respectively. The model proposed in [23] and the USARP model reduce the 3GPP model RMSE to approximately half of its error, even though the model proposed

TABLE 6. The USARP model PL error metrics in the validation and the generalization DT data.

	Dataset	RMSE [dB]	MAE [dB]	EVS
USARP	Val	10.71	8.53	0.52
	Gen	12.34	9.56	0.35

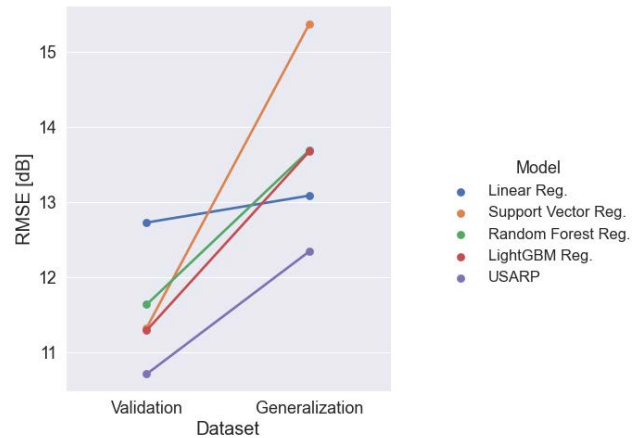


FIGURE 15. PL RMSE of the linear regression, data-driven, and USARP PL models in the validation and the generalization datasets.

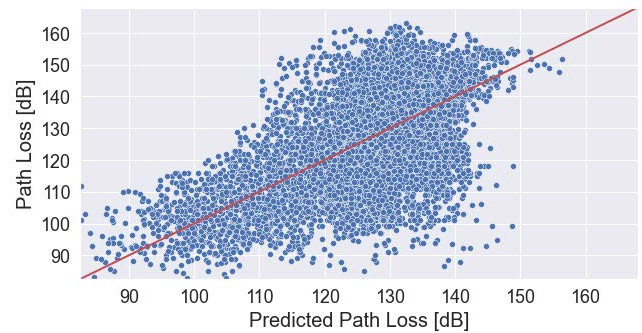


FIGURE 16. PL measurements as a function of the USARP model PL predictions in the generalization dataset.

in [23] was trained end-to-end specifically in a single (and very particular) environment. On the contrary, the USARP model was developed in urban and suburban environments and preconized the generalization, over validation, accuracy.

C. USARP ABLATION STUDIES

This section applies ablation studies to evaluate the contribution of each input type, within the USARP architecture, to the PL prediction. First, the satellite images were replaced by matrices with all elements set to zero and with the same dimensions as the satellite images; secondly, the ROI mask images were replaced by matrices of the same size, filled with ones, and, finally, the radio link variables were set to zero. The corresponding PL performance in the validation and generalization datasets, for each ablation, is presented in Table 7.

When the ablation of the satellite image is applied, it also blocks the information flow from the ROI mask. Thus, in practice, the ablation of the satellite image corresponds to only use the radio link variables as input. Comparing with the regular USARP model, the satellite-based inputs improve the RMSE of the PL predictions by more than 2 dB on the validation dataset, and around 1 dB in the generalization dataset. The remaining error metrics report a similar behavior, where the satellite-based inputs contribute with higher performance gains in the validation dataset than in the generalization dataset. Moreover, an RMSE gain of 1.03 dB and 3.28 dB is obtained for the generalization and validation datasets, respectively. In [23], the authors reported gains of just 0.8 dB for the RMSE, from using satellite images.

In the second ablation, which targeted the ROI mask, all error metrics are severely affected in both datasets. Thus, the extraction of relevant information characterizing both the UE and the BS locations is enhanced by using the ROI mask.

In the final ablation, the radio link variables were set to zero. As presented in Table 7, the radio link variables have the highest contribution to the performance of the USARP model. It results from the chosen architecture (*cf.* SLP 4 in Fig. 13) that incorporates known fundamentals of radio propagation.

VII. EXTENDING THE USARP MODEL TO MULTIPLE RADIO ENVIRONMENTS

This section shows the potential of using the USARP model for the PL prediction in multiple radio propagation environments. The supporting data —satellite and DT— are firstly introduced; the main results of the USARP model performance evaluation over multiple radio propagation environments are then presented and analysed.

A. DATA

The ResNet50 CNN used in the USARP model was trained using the BYOL (as in section IV-C) with 1000 new satellite images, each one corresponding to a $2 \text{ km} \times 2 \text{ km}$ geographical area. These images were randomly obtained from previously identified areas, encompassing rural, suburban, and urban environments to incorporate, in the ResNet50, representations of distinct radio environments.

Fig. 17 depicts an example of a rural environment satellite image used in the self-supervised training of the ResNet50, while Fig. 18 and Fig. 19 depict the suburban and urban areas, respectively. Afterwards, the ResNet50 was trained during 500 epochs, with a learning rate of 1×10^{-3} .

The DT data used to extract the PL measurements (as detailed in section III-B) was obtained from 85 distinct BSs from different radio propagation environments. Moreover, only measurements on the 800 MHz band were contemplated, as this band is widely selected regardless of the environment, from rural to urban locations, due to its lower PL. Fig. 20 exhibits the association between the PL and the 3D distance of the whole DT measurements, and the environment

TABLE 7. PL error metrics of the input ablations for the USARP model in the validation and generalization datasets.

Dataset	Ablation	RMSE [dB]	MAE [dB]	EVS
Val	None	10.71	8.53	0.52
	Satellite	13.99	11.19	0.25
	ROI	52.50	43.24	-0.03
Gen	Radio Link	98.30	97.13	0.03
	None	12.34	9.56	0.35
	Satellite	13.37	10.51	0.27
	ROI	81.12	58.27	-0.14
	Radio Link	104.05	102.74	-0.16



FIGURE 17. Example of rural satellite image used for the ResNet50 training using the BYOL [31].



FIGURE 18. Example of suburban satellite image used for the ResNet50 training using the BYOL [31].

(rural, suburban, and urban) associated with each PL measurement. This classification was obtained by considering a conversion of population density to radio environment provided by [54] and a population density map [55]. Overall, from a total of 6066 PL measurements, 3275 correspond to rural, 1300 to suburban, and 1491 to urban locations.



FIGURE 19. Example of urban satellite image used for the ResNet50 training using the BYOL [31].

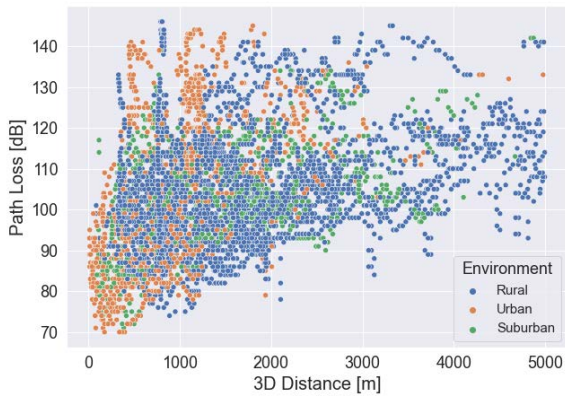


FIGURE 20. PL measurements as a function of the 3D distance between the BS and the UE with radio environment classification.

B. RESULTS

The potential of the USARP model for predicting PL in multiple radio propagation environments was assessed using the DT data (presented in the previous section). The DT data was randomly divided into training and validation datasets while maintaining the proportion of rural, suburban, and urban measurements in both datasets. Moreover, the training dataset accounted for 80% of the PL measurements, and the validation dataset with the remaining 20%.

The USARP model was trained with the training dataset using the hyperparameters shown in Table 4, except for the number of epochs (set to 100). The number of epochs was increased to obtain broader conclusions about the potential of the USARP model by evaluating the model performance along the training iterations.

Fig. 21 depicts the RMSE loss for the training and validation datasets obtained by the USARP model as a function of the number of epochs. The training loss of the USARP model is represented with the blue line, while the orange line corresponds to the USARP validation error. The training error gradually decreases as the number of epochs increases, while the validation loss follows the training loss trend,

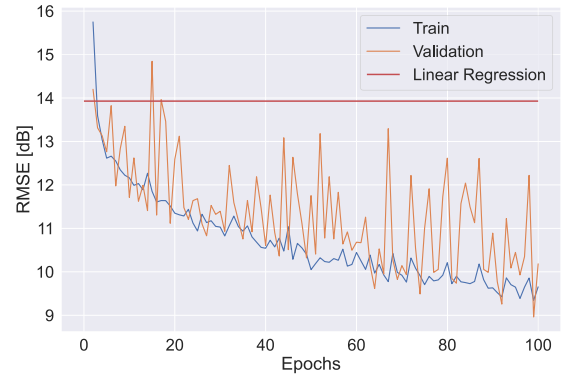


FIGURE 21. PL RMSE of the USARP model in the training and validation datasets, as a function of the epoch number, and the PL RMSE of the linear regression PL model in the validation dataset.

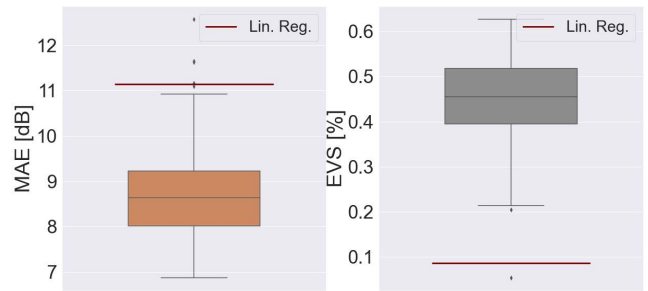


FIGURE 22. PL MAE and EVS distributions of the USARP model in the validation dataset and the respective errors of the linear regression in the same dataset (red lines).

despite exhibiting higher variability. Furthermore, the training dataset was used as fitting data for a linear regression algorithm (as in section VI-A), and the error metrics were calculated on the validation dataset. In Fig. 21, the horizontal red line corresponds to the validation RMSE obtained by the linear regression. It may be concluded that only in the worst cases (particular epochs), the USARP does not provide a lower error than the linear regression. Overall, the USARP model reaches higher accuracy in PL predictions and, in the best case, the difference to the linear regression is substantial, extending to more than 4 dB for the RMSE. Note that linear regression was selected for a reference comparison as it is the baseline model with higher generalization capacity, therefore the most trustworthy PL model (*cf.* section VI-A).

The remaining error metrics were also considered to evaluate the USARP model. Fig. 22 depicts a box-plot graph representing the MAE and EVS distributions of the USARP model in the validation dataset. The horizontal red lines denote the MAE and EVS values of the linear regression, in the validation dataset. The USARP model error distributions consistently present values outperforming the ones obtained by the linear regression PL model. For completeness, the statistics defining the previous box-plot representations (minimum, percentiles 25, 50, and 75, and the maximum) for each error metric are displayed in Table 8, along with the corresponding linear regression error metrics.

TABLE 8. Error metrics of the linear regression PL model and the respective statistics for the USARP model in the validation dataset.

	RMSE [dB]	MAE [dB]	EVS
Linear Regression	13.94	11.15	0.09
Minimum	8.96	6.87	0.21
25th Percentile	10.48	8.02	0.39
50th Percentile	11.24	8.65	0.46
75th Percentile	12.08	9.24	0.52
Maximum	13.96	10.93	0.63

TABLE 9. Error metrics of the general USARP model and the specialized linear regression models for each radio environment.

		RMSE [dB]	MAE [dB]	EVS
Urban	Lin. Reg.	15.37	12.63	0.31
	USARP	11.26	8.96	0.63
Suburban	Lin. Reg.	8.74	7.24	0.26
	USARP	6.26	4.59	0.63
Rural	Lin. Reg.	13.41	10.61	0.02
	USARP	8.50	6.68	0.61

Considering the median statistic of the USARP error distributions, it can be stated that this model improves the RMSE, MAE, and EVS of the linear regression, in 2.70 dB, 2.50 dB, and 0.37, respectively. Altogether, under the same setup regarding training and validation data representing multiple propagation environments, the USARP model clearly surpasses the linear regression. Additionally, section VI-B demonstrated that the USARP model has a higher generalization capacity than the linear regression, which in turn surpasses all the ML-based algorithms.

The potential of the USARP model for widespread geographical use, considering multiple radio propagation environments, was further evaluated. Firstly, the USARP model parameters with the lowest RMSE on the validation dataset were considered. Secondly, linear regression was used specifically for each propagation environment, where only the training data for the respective environment was used. Therefore, the USARP model was trained with data from all radio environments, while three linear regression models were obtained, specifically for each environment. Table 9 exhibits, for each radio environment, the error metrics obtained on the respective validation datasets, for each model.

The potential of the USARP model is emphasized by the lower error metrics when compared to the environment-specific linear regression models. Therefore, the USARP model has a high potential to be used in multiple propagation environments, given its generalization capacity and ability to surpass environment-specific PL models.

VIII. CONCLUSION

This paper proposes the USARP model for PL predictions, improving the geographical generalization capabilities of empirical PL models, including ML/DNN based, towards an ubiquitous PL model.

Firstly, it was shown that the performance of regression-based ML algorithms decreases significantly for locations not considered on the training, although belonging to similar propagation environments. In this context, the linear regression (the base of empirical PL models) is the most robust approach considering the geographic generalization performance. Therefore, the use of satellite images and DNN algorithms provides an opportunity to enhance the geographic generalization performance of data-driven PL models. Consequently, this paper proposes to split the problem of PL estimation using satellite images in two steps: 1) use of self-supervised learning to learn radio environment representations from satellite images; 2) employment of the radio environment representations together with DT measurements, for PL prediction. This approach allows the development of robust satellite image representations, notably from locations without DT data, contributing to the geographical generalization of the model.

Then, the USARP model, based on a DNN architecture, was proposed with focus on the generalization performance. Notwithstanding, the USARP model still exceeds the baseline methods in the validation performance, but it also surpasses the generalization performance of the baseline methods. In the generalization dataset, the USARP model attained an RMSE of 12.34 dB, 1 dB lower than the RMSE resulting from the linear regression-based model, and 3 dB and 2 dB lower than the RMSE resulting from the SVR and the RFR based models, respectively. Furthermore, the ablation studies performed on the USARP architecture revealed that the satellite-based inputs improve the RMSE of the PL predictions by more than 3 dB on the validation dataset, and around 1 dB in the generalization dataset, improving on previously reported values in the literature [23].

Finally, the potential of the USARP model for multiple radio propagation environments was shown. In fact, the USARP model can achieve a higher prediction accuracy than linear regression models specialized for each environment.

Overall, the USARP model enhances the geographical generalization capacities of empirical PL models, supported by an appropriated architecture, with regularization methods, and by successfully exploiting data from satellite images in a self-supervised approach.

Future work is in motion to extend the USARP model for multiple radio frequencies and develop new approaches to learn even more insightful representations of the radio environment from satellite images.

ACKNOWLEDGMENT

The authors would like to thank the Instituto de Telecomunicações (IT) and to Celfinet for the support and contributions to this work.

REFERENCES

- [1] R. Borralho, A. Mohamed, A. U. Quddus, P. Vieira, and R. Tafazolli, "A survey on coverage enhancement in cellular networks: Challenges and solutions for future deployments," *IEEE Commun. Surveys Tuts.*, vol. 23, no. 2, pp. 1302–1341, Jan. 2021.

- [2] M. Sousa, A. Martins, and P. Vieira, "Self-optimization of low coverage and high interference in real 3G/4G radio access networks," *Inst. Eng. Lisbon Academic J. Electron. Telecommun. Comput.*, vol. 3, no. 1, p. 12, 2018. [Online]. Available: <http://journals.isel.pt/index.php/IETC/article/view/39>
- [3] R. Borralho, D. Duarte, A. Quddus, and P. Vieira, "Developing a LTE localization framework using real network data towards RAN optimization through context knowledge," in *Proc. 23rd Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, Oct. 2020, pp. 1–6.
- [4] W. Fan, Y. Peng, Z. Yuan, P. Chen, C. Hu, and X. Zhang, "Sensing and monitoring for cellular networks: A crowdsourcing platform from mobile smartphones," in *Proc. IEEE Int. Conf. Data Sci. Data Intensive Syst.*, Dec. 2015, pp. 472–473.
- [5] C. Phillips, D. Sicker, and D. Grunwald, "A survey of wireless path loss prediction and coverage mapping methods," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 1, pp. 255–270, 1st Quart., 2013.
- [6] N. Moraitis, L. Tsipi, and D. Vouyioukas, "Machine learning-based methods for path loss prediction in urban environment for LTE networks," in *Proc. 16th Int. Conf. Wireless Mobile Comput., Netw. Commun. (WiMob)*, Oct. 2020, pp. 1–6.
- [7] S. Aldossari and K.-C. Chen, "Predicting the path loss of wireless channel models using machine learning techniques in mmWave urban communications," in *Proc. 22nd Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, Nov. 2019, pp. 1–6.
- [8] M. Sousa, A. Alves, P. Vieira, M. P. Queluz, and A. Rodrigues, "Analysis and optimization of 5G coverage predictions using a beamforming antenna model and real drive test measurements," *IEEE Access*, vol. 9, pp. 101787–101808, 2021.
- [9] G. Mao, B. D. O. Anderson, and B. Fidán, "WSN06–4: Online calibration of path loss exponent in wireless sensor networks," in *Proc. IEEE Globecom*, Nov. 2006, pp. 1–6.
- [10] T. S. Rappaport, *Wireless Communications—Principles and Practice*. Upper Saddle River, NJ, USA: Prentice-Hall, 1996.
- [11] J. Huang, Y. Liu, C.-X. Wang, J. Sun, and H. Xiao, "5G millimeter wave channel sounders, measurements, and models: Recent developments and future challenges," *IEEE Commun. Mag.*, vol. 57, no. 1, pp. 138–145, Jan. 2019.
- [12] O. O. Erunkulu, A. M. Zungeru, C. K. Lebekwe, and J. M. Chuma, "Cellular communications coverage prediction techniques: A survey and comparison," *IEEE Access*, vol. 8, pp. 113052–113077, 2020.
- [13] S. Sun, T. S. Rappaport, S. Rangan, T. A. Thomas, A. Ghosh, I. Z. Kovacs, I. Rodriguez, O. Koymen, A. Partyka, and J. Jarvelainen, "Propagation path loss models for 5G urban micro- and macro-cellular scenarios," in *Proc. IEEE 83rd Veh. Technol. Conf. (VTC Spring)*, May 2016, pp. 1–6.
- [14] *Study on Channel Model for Frequencies From 0.5 to 100 GHz (Release 16)*, document TS 38.901, 12, 3GPP, version 16.1.0., 2019. [Online]. Available: <https://portal.3gpp.org>
- [15] C.-X. Wang, J. Bian, J. Sun, W. Zhang, and M. Zhang, "A survey of 5G channel measurements and models," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 4, pp. 3142–3168, 4th Quart., 2018.
- [16] M. Hatay, "Empirical formula for propagation loss in land mobile radio services," *IEEE Trans. Veh. Technol.*, vol. VT-29, no. 3, pp. 317–325, Aug. 1980.
- [17] *Measurement Results and Final mmMAGIC Channel Models*, document ICT-671650, 5GPP, mmMAGIC Project, version 2.0., 2017.
- [18] T. S. Rappaport, G. R. Maccartney, M. K. Samimi, and S. Sun, "Wideband millimeter-wave propagation measurements and channel models for future wireless communication system design," *IEEE Trans. Commun.*, vol. 63, no. 9, pp. 3029–3056, Sep. 2015.
- [19] H. F. Ates, S. M. Hashir, T. Baykas, and B. K. Gunturk, "Path loss exponent and shadowing factor prediction from satellite images using deep learning," *IEEE Access*, vol. 7, pp. 101366–101375, 2019.
- [20] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
- [21] S. P. Sotiropoulos, S. K. Goudos, and K. Siakavara, "Deep learning for radio propagation: Using image-driven regression to estimate path loss in urban areas," *JCT Exp.*, vol. 6, no. 3, pp. 160–165, Sep. 2020.
- [22] 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.
- [23] 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.
- [24] J. Thrane, B. Sliwa, C. Wietfeld, and H. L. Christiansen, "Deep learning-based signal strength prediction using geographical images and expert knowledge," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Taipei, Taiwan, Dec. 2020, pp. 1–6.
- [25] X. Zhang, X. Shu, B. Zhang, J. Ren, L. Zhou, and X. Chen, "Cellular network radio propagation modeling with deep convolutional neural networks," 2021, *arXiv:2110.01848*.
- [26] X. Yang, X. He, Y. Liang, Y. Yang, S. Zhang, and P. Xie, "Transfer learning or self-supervised learning? A tale of two pretraining paradigms," 2020, *arXiv:2007.04234*.
- [27] Y. Egi, E. Eyceyurt, I. Kostanic, and E. Carlos Otero, "An efficient approach for evaluating performance in LTE wireless networks," in *Proc. World Congr. Comput. Sci., Comput. Eng., Appl. Comput.*, 2017, pp. 1–7.
- [28] A. J. Smola and B. Schölkopf, "A tutorial on support vector regression," *Statist. Comput.*, vol. 14, no. 3, pp. 199–222, Aug. 2004, doi: [10.1023/B:STCO.0000035301.49549.88](https://doi.org/10.1023/B:STCO.0000035301.49549.88).
- [29] L. Breiman, "Random forests," *Mach. Learn.*, vol. 45, no. 1, pp. 5–32, 2001, doi: [10.1023/A:1010933404324](https://doi.org/10.1023/A:1010933404324).
- [30] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, and T.-Y. Liu, "LightGBM: A highly efficient gradient boosting decision tree," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 3146–3154.
- [31] *Google Earth Pro*, Image 2022 TerraMetrics, Data SIO, NOAA, U.S. Navy, NGA, GEBCO, Google, Dublin, Ireland, 2022.
- [32] Q. Li, B. Li, J. M. Garibaldi, and G. Qiu, "On designing good representation learning models," 2021, *arXiv:2107.05948*.
- [33] I. Misra and L. van der Maaten, "Self-supervised learning of pretext-invariant representations," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 6706–6716.
- [34] L. Jing and Y. Tian, "Self-supervised visual feature learning with deep neural networks: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 11, pp. 4037–4058, Nov. 2021.
- [35] Z. Li, W. Yang, S. Peng, and F. Liu, "A survey of convolutional neural networks: Analysis, applications, and prospects," 2020, *arXiv:2004.02806*.
- [36] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," 2015, *arXiv:1512.03385*.
- [37] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," 2016, *arXiv:1608.06993*.
- [38] I. Bello, W. Fedus, X. Du, E. D. Cubuk, A. Srinivas, T.-Y. Lin, J. Shlens, and B. Zoph, "Revisiting ResNets: Improved training and scaling strategies," 2021, *arXiv:2103.07579*.
- [39] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, and A. Desmaison, "PyTorch: An imperative style, high-performance deep learning library," in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2019, pp. 8024–8035. [Online]. Available: <http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf>
- [40] C. Doersch, A. Gupta, and A. A. Efros, "Unsupervised visual representation learning by context prediction," 2015, *arXiv:1505.05192*.
- [41] D. Kim, D. Cho, D. Yoo, and I. S. Kweon, "Learning image representations by completing damaged jigsaw puzzles," 2018, *arXiv:1802.01880*.
- [42] C. Wei, L. Xie, X. Ren, Y. Xia, C. Su, J. Liu, Q. Tian, and A. L. Yuille, "Iterative reorganization with weak spatial constraints: Solving arbitrary jigsaw puzzles for unsupervised representation learning," 2018, *arXiv:1812.00329*.
- [43] J. B. Grill, F. Strub, F. Altché, C. Tallec, P. Richemond, E. Buchatskaya, C. Doersch, B. A. Pires, Z. Guo, M. G. Azar, and B. Piot, "Bootstrap your own latent: A new approach to self-supervised Learning," *Proc. Adv. Neural Inf. Process. Syst.*, 2020, pp. 21271–21284.
- [44] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," *J. Big Data*, vol. 6, no. 1, pp. 1–48, Dec. 2019.
- [45] N. Ritter and M. Ruth, "The GeoTiff data interchange standard for raster geographic images," *Int. J. Remote Sens.*, vol. 18, no. 7, pp. 1637–1647, May 1997.
- [46] S. Eppel, "Setting an attention region for convolutional neural networks using region selective features, for recognition of materials within glass vessels," 2017, *arXiv:1708.08711*.
- [47] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 27, Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2014. [Online]. Available: <https://proceedings.neurips.cc/paper/2014/file/375c71349b295fbc2dcdca9206f20a06-Paper.pdf>

- [48] A. Zhang, Z. C. Lipton, M. Li, and A. J. Smola, "Dive into deep learning," 2021, *arXiv:2106.11342*.
- [49] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, Jan. 2014. [Online]. Available: <http://jmlr.org/papers/v15/srivastava14a.html>
- [50] A. Y. Ng, "Feature selection, L_1 vs. L_2 regularization, and rotational invariance," in *Proc. 21st Int. Conf. Mach. Learn. (ICML)*, New York, NY, USA, 2004, p. 78, doi: [10.1145/1015330.1015435](https://doi.org/10.1145/1015330.1015435).
- [51] T. Akiba, S. Sano, T. Yanase, T. Ohta, and M. Koyama, "Optuna: A next-generation hyperparameter optimization framework," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, New York, NY, USA, Jul. 2019, pp. 2623–2631, doi: [10.1145/3292500.3330701](https://doi.org/10.1145/3292500.3330701).
- [52] J. Bergstra, R. Bardenet, Y. Bengio, and B. Kégl, "Algorithms for hyperparameter optimization," in *Advances in Neural Information Processing Systems*, vol. 24, J. Shawe-Taylor, R. Zemel, P. Bartlett, F. Pereira, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2011.
- [53] A. Alves, M. Sousa, P. Vieira, M. P. Queluz, and A. Rodrigues, "A new 3D beamforming antenna model for 5G propagation modeling based on real data," in *Proc. 23rd Int. Symp. Wireless Pers. Multimedia Commun. (WPMC)*, Oct. 2020, pp. 1–6.
- [54] ANACOM. *Dimensioning of the Network*. Accessed: Jan. 23, 2022. [Online]. Available: <https://www.anacom.pt/render.jsp?categoryId=383691>
- [55] *Gridded Population of the World, Version 4 (GPWv4): Population Density, Revision 11*, Columbia Univ., Center Int. Earth Sci. Inf. Netw., Palisades, NY, USA, 2018, doi: [10.7927/H49C6VHW](https://doi.org/10.7927/H49C6VHW).



MARCO SOUSA received the B.Sc. degree in electronic and telecommunications engineering from the Universidade da Madeira (UMa), Portugal, in 2014, and the M.Sc. degree in electronic and telecommunications from the Instituto Superior de Engenharia de Lisboa (ISEL), Portugal, in 2017. He is currently pursuing the Ph.D. degree in electrical and computer engineering (PDEEC) with the Instituto Superior Técnico (IST), Technical University of Lisbon, Portugal.

Since 2015, he has been working with the Research Department, CELFINET, a Portuguese technological consultancy company. He is also the Team Leader of the research activities related to mobile network configuration management and optimization. His main research interests include mobile network optimization, radio access network coverage and interference, and applied machine learning/deep learning algorithms for mobile networks.



PEDRO VIEIRA was born in Lisbon, Portugal, in 1974. He received the B.Eng., M.Sc., and Ph.D. degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), Technical University of Lisbon, Portugal, in 1997, 2003, and 2008, respectively. Since 1997, he has been with the Electronics, Telecommunications and Computer Engineering Department, Lisbon Polytechnic Institute (ISEL). He is currently an Adjunct Professor with the ISEL, and a Researcher

at the Instituto de Telecomunicações, where he is researching aspects of wireless communications, including radio propagation, radio network planning, and optimization and SON systems. He is also a Senior Engineer registered in the Portuguese Engineering Order (OE), and an OE Telecommunications Specialist, since 2008. He is also leading the research activities with CELFINET, a Portuguese technological consultancy company, where he is engaging in applied research to create new methodologies, tools, and algorithms for the mobile communications business.



MARIA PAULA QUELUZ received the B.S. and the M.S. degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), University of Lisbon, Portugal, and the Ph.D. degree from the Catholic University of Louvain, Louvain-la-Neuve, Belgium. She is currently an Assistant Professor with the Department of Electrical and Computer Engineering, IST, and a Senior Research Member of the Instituto de Telecomunicações, Lisbon, Portugal. Her main

research interests include image/video quality assessment, image/video processing, and wireless communications.



ANTÓNIO RODRIGUES (Member, IEEE) received the B.S. and M.S. degrees in electrical and computer engineering from the Instituto Superior Técnico (IST), Technical University of Lisbon, Lisbon, Portugal, in 1985 and 1989, respectively, and the Ph.D. degree from the Catholic University of Louvain, Louvain-la-Neuve, Belgium, in 1997. Since 1985, he has been with the Department of Electrical and Computer Engineering, IST, where he is currently an Associate Professor. He is also

a Senior Research Member at the Instituto de Telecomunicações, Lisbon. His current research interests include mobile and satellite communications, wireless networks, modulation, coding, and multiple access techniques.

...