Toward EM Digital Twins: EM-Guided Deep Learning for Accurate Prediction of Wireless Power Transfer in IIoT Environments

Toward EM Digital Twins: EM-Guided Deep Learning for Accurate Prediction of Wireless Power Transfer in IIoT Environments

Elisa Augello, *Student Member, IEEE*, Giacomo Paolini, *Member, IEEE,* Diego Masotti, Senior *Member, IEEE*, and Alessandra Costanzo *Fellow, IEEE*

***Abstract*— Accurately predicting electromagnetic (EM) field propagation is fundamental for the design of wireless power transfer (WPT) and industrial Internet of Things (IIoT) systems, where sensors performance and energy availability strongly depend on the surrounding environment, particularly in harsh or cluttered conditions. Conventional full-wave simulations are precise but computationally prohibitive for large or dynamic scenarios, limiting their use for real-time cases. Models based on ray tracing rely on the far-field approximation that is often violated in industrial and automotive environments. This work presents a data-efficient deep learning framework that integrates physics-based modelling with convolutional neural networks (CNNs) to reconstruct spatial power density distributions in complex environments. A dedicated data generation pipeline, combining active sampling with an integral equation (IE)-based EM solver, enables the construction of informative training datasets from a limited number of simulations. Different optimization objective functions are investigated to improve the accuracy and physical consistency of the predicted fields. Despite being trained on a reduced dataset, the proposed CNNs accurately reproduce both global propagation patterns and localized field variations, achieving strong agreement with experimental measurements. The results demonstrate that CNN-based models can deliver fast and reliable evaluation of EM power distribution allowing optimal sensor placement in industrial environments. These capabilities represent an important first step toward developing data-driven EM digital twins for next-generation IIoT systems.**

***Index Terms*—Artificial Intelligence, Convolutional Neural Network, Internet of Things, Wireless Power Transfer**

I. Introduction

THE industrial internet of things (IIoT) has recently attracted increasing interest as a paradigm that promotes enhanced and seamless connectivity among diverse tools [1]. IIoT offers the potential to introduce a wide range of advanced functionalities such as remote monitoring, predictive maintenance, and interoperability across diverse applicative scenarios such as automotive, manufacturing, and healthcare sectors. The implementation of these functionalities is facilitated through the deployment of wireless sensor networks (WSNs) within the target environment [2]. Powering the growing number of sensor nodes with wires is increasingly impractical, limiting flexibility and scalability while adding weight and complexity. To address these limitations, wireless power transfer (WPT) and energy harvesting (EH) have emerged as transformative technologies enabling the realization of fully autonomous WSNs, removing the need for cables and bulky batteries and supporting lighter and more modular architectures [3]. These benefits are particularly relevant in constrained or harsh environments – such as manufacturing systems, automotive powertrains, and aircraft interiors – where space, temperature, and weight limit wired power. WSN performance and wireless powering efficiency depend on accurate radio channel characterization. Industrial environments are electromagnetically complex, with components of different shapes and materials affecting propagation through shielding, absorption, and interference. Accurate prediction of spatial power distribution is therefore essential for effective sensor placement, continuous power delivery, as well as estimation of energy harvesting potential and charging times [4–5]. Additionally, within a WPT framework, such knowledge enables an optimization in the co-location of both multiple radiofrequency (RF) sources – minimizing destructive interference among them – and receiving elements. Propagation is typically characterized by using models derived from extensive measurements, but these are often set-up specific or too coarse to capture complex wave interactions [6]. Ray tracing (RT)-based solvers utilize precise geometric descriptions of the environment to achieve accurate modeling of propagation phenomena [7], yet rely on the far-field approximation which is often violated in industrial or automotive environments. Within this context, artificial neural networks (ANNs) have emerged as a promising alternative due to their capacity to learn complex, high-dimensional relationships between input parameters and target outputs [8–10]. Additionally, they are key enablers for the development of digital twins (DTs), owing to their ability to learn complex, high-dimensional relationships and replicate the behavior of physical systems with high fidelity and adaptability [11]. In the context of EM propagation, DTs have gained considerable attention as virtual counterparts of physical environments, capable of reproducing intricate propagation mechanisms across diverse scenarios [12]. By harnessing the predictive power of ANNs, EM DTs offer data-driven models of real-world environments, enabling rapid, scalable, and cost-effective analysis. They support system design, performance optimization, and real-time decision-making, making them well suited for applications such as autonomous IIoT network powering and next-generation wireless communication. Preliminary results of this work were presented in [13], where ANNs were employed to directly predict the optimal sensor

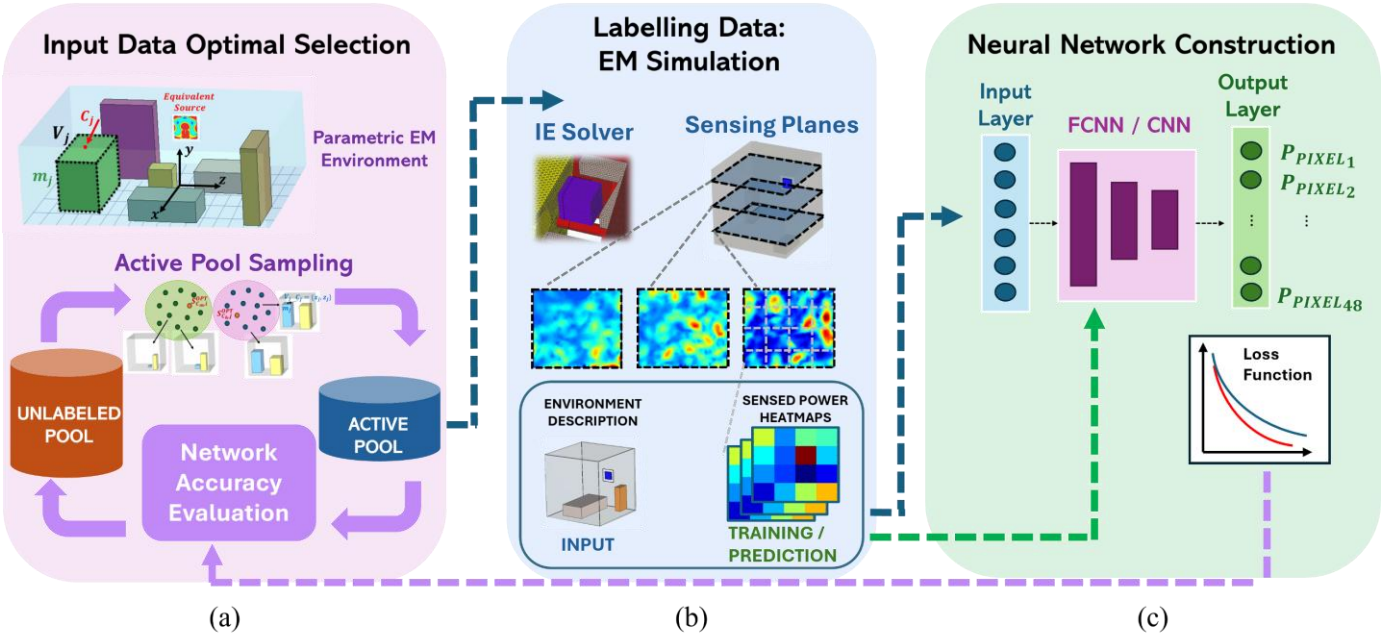

Fig. 1. Workflow for developing the NN model: (a) a parametric description of the EM environment is defined, and an optimal subset of samples is first selected based on their information content; (b) these selected samples are then labeled through EM simulations using a configuration with reduced computational complexity based on Integral Equation Solver; (c) the resulting dataset, comprising the selected inputs and their simulated labels, is used to train and validate the NN. This process is iteratively refined until satisfactory model performance is achieved.

position and the corresponding received power at that location. This work extends the methodology to predict full power distribution maps across multiple sensing planes, enabling evaluation and optimized co-location of multiple wireless sensors. Convolutional neural networks (CNNs) are used for their spatial modeling capability, validated through EM simulations and measurements. By capturing near-field effects, multipath, and boundary interactions, this approach supports effective WPT planning and multi-sensor placement in realistic IIoT scenarios. This paper is structured as follows. Section II describes the optimized data generation and selection process, outlining the case study and the datasets employed for training the neural networks (NNs). Section III details the development of the predictive modeling framework, presenting the CNN architectures and their validation using multiple loss functions. Section IV illustrates the experimental setup and compares measured data with model predictions. Finally, Section V summarizes the main findings and outlines potential directions for future research.

II. Dataset Mapping of the EM Environments

Accurate prediction of power propagation in enclosed industrial or automotive settings using NNs requires a representative dataset of input-output mappings. The considered environment (Fig. 1a) is a simplified industrial enclosure with metallic walls, multiple internal obstacles, and an RF source operating at a specified frequency to wirelessly power embedded sensors. Dataset construction, however, is inherently expensive: each simulated sample requires a full-wave EM solution, while measurement campaigns demand significant setup effort, making large-scale acquisition impractical. To address this, we rely on EM simulations combined with an optimized data generation strategy (Fig. 1). Candidate configurations are generated through parametric modeling of the environment, while an active sampling algorithm identifies the most informative ones. The selected samples are then labeled via an IE-based EM solver, yielding high-fidelity power density distributions in a more efficient way compared to plain full-wave simulations [14]. The EM simulation results are used to assign a label to each selected environment, corresponding to the power distribution, as shown in Fig. 1(b). This curated dataset is subsequently employed to train NN models (Fig. 1(c)), enabling rapid and accurate EM field prediction across diverse deployment scenarios. The following subsections describe in detail the simulation methodology and the smart sampling procedure used to build the training set.

A. Optimal Data Selection

A single generic EM environment is represented by means of a set of descriptive variables, which are varied and sampled to generate a large set of environments, referred to as the candidate set. The generic element of the matrix describing the candidate set is referred to as $S_{i,j,k}$ with $i = 1, \dots, N_V$ descriptive variables, $j = 1, \dots, N_O$ number of obstacles present in the environment and $k = 1, \dots, N_E$ total number of environments. A generic environment therefore is described by a matrix of $N_V \cdot N_O$ −dimension, composed of descriptive variables that define the geometry and material properties of each obstacle within the scenario. By sampling these descriptive variables within their admissible ranges, a broad and diverse set of feasible environments is generated, which constitutes the candidate set. The descriptive variables chosen for this work include volume sides $V = \{x_l, z_l, h_l\}$, centers $C = \{x_c, z_c\}$ and material type m, as depicted in Fig. 2(a) and following the procedure detailed in [15], but the framework can be easily extended to include more complex geometries and configurations for enhanced environmental characterization. Each environment is simulated through full-wave EM

simulation to obtain the corresponding power density distribution associated with that topology. To reduce the labelling effort, the number of $S_{i,j,k}$ sets to be EM simulated is carefully selected to remain representative of real-world variability while minimizing the overall computational burden. To address this challenge, an active learning algorithm inspired by the one in [16] is applied to the candidate set. To initialize the process, K-means clustering is applied, which is a well-established algorithm that groups similar candidate samples – representing comparable EM environments – into M clusters. The corresponding cluster centroids are added to the active pool as the initial selection, which is subsequently refined through iterative updates. For each cluster C_m with $m = 1, \dots, M, \ll N_E$, one optimally selected sample is chosen and added to the active pool by jointly considering its diversity with respect to the already selected samples and its representativeness within its own cluster. We refer to an environment which is a sample of the candidate set belonging to cluster C_m as $S_{i,j,p}$ with $p = 1, \dots, N_{C_m}$ and N_{C_m} number of samples in the cluster. Two figures of merit (FoM) are defined: representativeness G_R and diversity G_D. The representativeness of a sample $S_{i,j,p}$ is expressed as:

$$G_R(S_{i,j,p}) = \frac{1}{N_{C_m} - 1} \times \sum_{q=1}^{N_{C_m}} \left\| S_{i,j,p} - S_{i,j,q_{(q \neq p)}} \right\| \quad (1)$$

with $p, q = 1, \dots, N_{C_m}$ elements in the same cluster. It accounts for the similarity between a sample with the others in the same cluster. The diversity metric is defined as:

$$G_D(S_{i,j,p}) = \min_{o=1,\dots M} \left\| S_{i,j,p} - S_{i,j,o}^{OPT} \right\| \quad (2)$$

with $p = 1, \dots, N_{C_m}$ samples in the cluster C_m and $o = 1, \dots, M$ index scanning the optimal samples already selected in other different clusters C_o with $o \neq m$. It accounts for the minimum distance between a candidate sample in the cluster C_m and the optimally selected samples from other clusters that have already been included in the active pool. The overall selection criterion which extracts the optimal sample for the cluster C_m combines representativeness and diversity through a trade-off optimization defined as

$$S_{i,j,p}^{OPT} = \arg\max[G_D(S_{i,j,p}) - G_R(S_{i,j,p})] \quad (3)$$

with $p = 1, \dots, N_{C_m}$. This ensures that the chosen EM environments are sufficiently distinct from one another while remaining representative of their respective cluster. The procedure can be iteratively applied, where the optimally selected M samples from the current active pool serve as the initialization set for the next one. At the end of each iteration n, the newly selected configuration $M^{(n)}$ are labelled through EM simulation and used to train the ANN. The process is repeated until the model achieves the desired prediction accuracy, and the results of these iterations are reported in Section IV.

B. EM Simulation of the Selected Dataset

Choosing an appropriate EM simulator is critical for WSN design. Full-wave simulators can be prohibitively slow for large domains, while ray tracing methods are faster but rely on far-field assumptions. In realistic environments, sensors often operate in the near- or mid-field of antennas, where far-field models, such as those based on the Friis formula, fail to capture true RF interactions, including mutual coupling, obstacle effects, and complex multipath in cluttered settings. To address these challenges, the present methodology leverages an integral equation (IE)-solver, following the approach outlined in [17] and further described in [5]. By employing a surface-based meshing strategy rather than volumetric discretization, the IE solver maintains the accuracy of traditional full-wave frequency-domain simulations while substantially reducing computational requirements. The solver decomposes the EM field generated by a given current distribution into near-field and far-field contributions. The near-field component is computed using the method of moments (MoM), after which an approximate evaluation is performed for the far-field region. This hybrid approach combines high precision with computational efficiency, making it suitable for large and complex EM problems [18]. An equivalent source model of the RF emitter avoids fine meshing in the full environment by first characterizing the source in a bounded volume and then applying its effect as an impressed current in the larger scenario.

C. A Case Study Environment

The reference environment shown in Fig. 2(a) consists of a square Perfect Electric Conductor (PEC) enclosure ($L_{env} = 50$ cm) containing a fixed RF source and multiple sensing planes, on which the IE solver computes the power density distribution. A 2.45 GHz patch antenna, representative of typical WSN applications, is used as the illuminator. To generate dataset variability, two internal obstacles ($N_O = 2$) are parametrically varied in size, location, and material properties to generate the dataset. Their volumes V_j are sampled over the

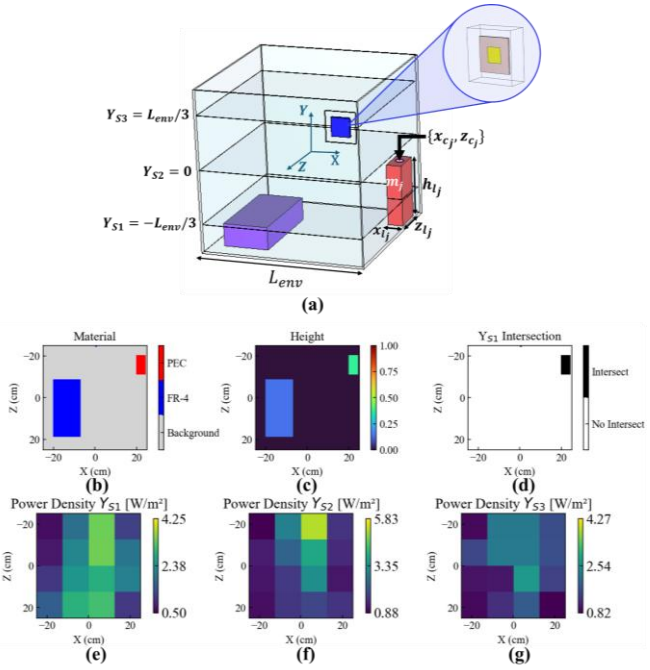

Fig. 2. (a) Schematic representation of the reference enclosed environment in presence of the equivalent source; (b) the environment is described by: a material channel, (c) a normalized height channel, (d) the coordinates of the obstacles intersection with the sensing planes, and by the power density maps on the sensing planes (e) Y_{s1}, (f) Y_{s2}, (g) Y_{s3}.

ranges [8–8000] cm^3, with obstacles ranging from 0.3λ to 3λ, where λ is the wavelength of the operative frequency, while the corresponding centers $C_j = \{x_{C_j}, z_{C_j}\}$ cover the various positions in the entire environment. Two possible materials are allowed for these volumes, a non-ideal insulator such as FR-4 ($\varepsilon_r = 4.3, \tan\delta = 0.025$), and PEC. Starting from the initial candidate with $N_E = 10000$, active pools are iteratively extracted using (3), optimally selecting M=100 samples for n=8 iterations, resulting in a total of 800 samples. For the samples selected in the active pools, $G_R = [0.2,0.5]$ and $G_D = [0.2,0.6]$. Since both metrics are computed in a normalized space, extreme values close to zero or to the maximum distance would correspond to redundant samples or outliers. The description of the EM environment is given with both tabular features and images, to exploit spatial correlations in addition to obstacle properties. Fig. 2 shows an overview of the procedure chosen to describe the input sets. $S_{i,j,k}$ represents the generic sample descriptive variable, with $i = 1, \dots, N_O$ obstacles, $j = 1, \dots, N_V$ and k refers to a specific environment, $N_V = 6$ are the descriptive variables of each obstacle ($N_O = 2$, for the present case), that are organized in tabular data and inform about the obstacles' geometry, the associated material, and their location. These descriptors are then used to generate a 2D rendering of the environment. The XZ-plane is discretized into $N_w = 64$ pixels, and a value is assigned to each pixel to produce the image that is used as the input to the NN. Fig. 2(b-d) show the resulting five input channels: i) the obstacles material type, represented by a binary encoding (Fig. 2(b)), therefore to each pixel filled by an obstacle, a value corresponding to the material is assigned; ii) a normalized height channel, defined to give information about the Y-direction (Fig. 2(c)), (iii)-(v)) the intersections of the obstacles with the three sensing planes: in Fig. 2(d) the one corresponding to Y_{s1} is shown only, for brevity. Therefore, the set of inputs describing one environment, in the form of images, has dimensionality $(N_w, N_w, N_c) = (64,64,5)$, where N_w is discretization of the features encoding and N_c is the number of input channels used. The primary objective of NN modeling is to predict the power density distribution across the sensing planes, as a function of the above-described possible environments. Due to the complexity of the problem and the limited size of the dataset, the discretization used to capture the power-density distribution is chosen as a compromise between spatial accuracy and the computational efficiency required for the learning tasks. Consequently, each pixel in the power density representation corresponds to a square of side approximately equal to one wavelength, and the average power density within each pixel is assigned to the associated output map. For the present case, three maps (one for each sensing plane) of 4×4 pixels each are provided as the corresponding network output. An example of this discretization for the three reference planes is shown in Fig. 2(g-i), corresponding to an output flattened dimensionality of $N_P = 48$, where each of the 16 pixels in the three sensing planes contributes to one power density value.

III. Neural Network Architecture

Traditional analytical models and FCNNs often struggle to represent the spatial dependencies and complex EM interactions present in realistic environments. In contrast, CNNs naturally exploit spatial features, enabling more accurate prediction of power density maps. In this section, CNN architectures are developed and trained to reconstruct received power distributions within a confined environment. Multiple loss functions are evaluated to assess their impact on learning stability, accuracy, and generalization. Performance is also compared with a baseline FCNN using the same inputs, confirming the advantages of convolution-based learning for EM propagation modeling.

A. Loss Functions

To evaluate and optimize model performance, several loss functions are implemented, each emphasizing different aspects of prediction accuracy. Networks are trained to predict power density across three sensing planes, aiming to capture both magnitude and spatial distribution of the EM field. The baseline mean squared error (MSE) penalizes large deviations, promoting overall amplitude accuracy, but treats all pixels equally and ignores spatial structure. To account for field nonuniformity and enhance sensitivity to low-power regions, the maximum relative error (MRE) normalizes point-wise errors by the maximum field intensity. While MSE and MRE quantify local errors, they do not explicitly ensure structural consistency. For this, the structural similarity index measure (SSIM) [19] encourages preservation of spatial patterns and smooth variations, improving reconstruction of continuous field maps. Additionally, a custom max-pixel weighted (MPW) loss combines global MSE with a local term emphasizing the highest-intensity pixel and its position. A weighting parameter balances global and local accuracy, ensuring physically coherent power maps. Further analytical details are provided in Appendix I.

B. Convolutional Neural Network Details

The proposed hybrid model combines a CNN for spatial feature extraction with an FCNN feedforward stage to process both visual and tabular data. Its hyperparameters – including convolutional layers, kernel sizes, filters, activations, and dropout rates – are systematically optimized to minimize validation error for each loss function. Fig. 3 illustrates the CNN architecture optimized with the MPW loss, while Table I lists optimal parameters for the other loss functions. The CNN inputs, describing the training environments, are provided in the form of $(N_w, N_w, N_c) = (64,64,5)$ tensors, as detailed in Section II. The convolutional layers progressively extract spatial features from the input maps by applying localized filters, also called kernels, which slide over the input images to capture local variations, edges, and correlations among neighboring pixels. Each kernel performs a weighted sum over its receptive field, allowing the network to detect spatial patterns at different scales and orientations, effectively learning features that are relevant for predicting the EM power distribution. By increasing the depth of the representation in this way, the network can simultaneously capture multiple aspects of the environment. Since each kernel has its own distinct set of learnable weights, it specializes in different structures of input data images. The total number of filters, N_{Fl} in the l-convolutional layer is optimized by the minimization of the chosen loss functions and determines the types of feature maps to be extracted. These maps are obtained by sliding the corresponding kernels over the inputs. For the

example of Fig. 3 using the MPW loss function, three convolutional layers have been constructed with the following characteristics: in the first layer 16 filters with kernel size of 5×5 are applied independently over the inputs, producing 16 distinct 2D feature maps; the second layer employs 16 filters with kernel size of 3×3 to further refine these representations, again producing 16 feature maps; finally, the third layer uses 32 filters of kernel size 3×3, resulting in 32 feature maps. Each convolution is followed by batch normalization, which stabilizes and accelerates the training process normalizing the feature distributions, and by a nonlinear activation function that enables the network to model complex relationships between the spatial input patterns and the corresponding EM power distributions. At the end of the convolutional stage, an adaptive average pooling layer reduces the spatial resolution of the 32 feature maps from the last layer to the chosen resolution, the 4×4 pixel grid, producing a compact feature descriptor of size $32 \times 4 \times 4$. This downsampling technique replaces each pooling region – a small local neighborhood of pixels – with the average value of its elements, providing a summarized representation of the local features. By adapting the pooling regions to produce a fixed output size, the layer ensures consistent feature dimensions regardless of the input size, while preserving the most representative spatial information extracted by the convolutional filters. The resulting spatial descriptor is subsequently concatenated with the tabular input vector, which encodes detailed non-image features of the environment. This fusion produces a unified latent representation that serves as the input to the fully connected (FC) stage. The latter comprises two hidden layers of 128 neurons each, followed by batch normalization, nonlinear activation, and dropout regularization to enhance generalization and training stability. The final output layer consists of 48 neurons, containing the pixel-wise predicted power density across the three sensing planes. As mentioned above, this procedure has been carried out for various loss functions, in addition to the MPW, and Table 1 summarizes the hyperparameters obtained for each of them, as a trade-off between model complexity and generalization capability. The different loss functions favor slightly different NN configurations, reflecting the distinct sensitivity of each metric: using MSE and MRE, the best performance is achieved with convolutional representation employing a larger number of filters with respect to the MPW case, suggesting how both metrics benefit from increased feature extraction capacity, to capture the detailed power amplitude variations. SSIM and MPW, which both emphasize a structural consistency – favor a deeper convolutional backbone but with a lower number of filters. The dropout rate, which is the probability of a neuron being deactivated during training, and the weight decay, which serves as a regularization technique penalizing large weights, remain moderate but are fundamental to prevent overfitting, given the reduced size of the dataset.

TABLE I
OPTIMIZED HYPERPARAMETERS FOR LOSS FUNCTIONS

	MSE	MRE	SSIM	MPW
Convolutional Layers	2	2	3	3
Convolutional Filters	[32,64]	[64,128]	[16,16,32]	[16,16,32]
Convolutional Kernels	[3,3]	[3,5]	[5,3,3]	[5,3,3]
Weight Decay	10^{-4}	10^{-4}	$4 \cdot 10^{-5}$	$4 \cdot 10^{-5}$
Dropout Rate	0.2	0.1	0.08	0.09
Activation Function [20]	ReLu	GeLu	Leaky ReLu	Leaky ReLu
Fully Connected Layers	[256,128]	[256,256]	[256,256]	[128,128]

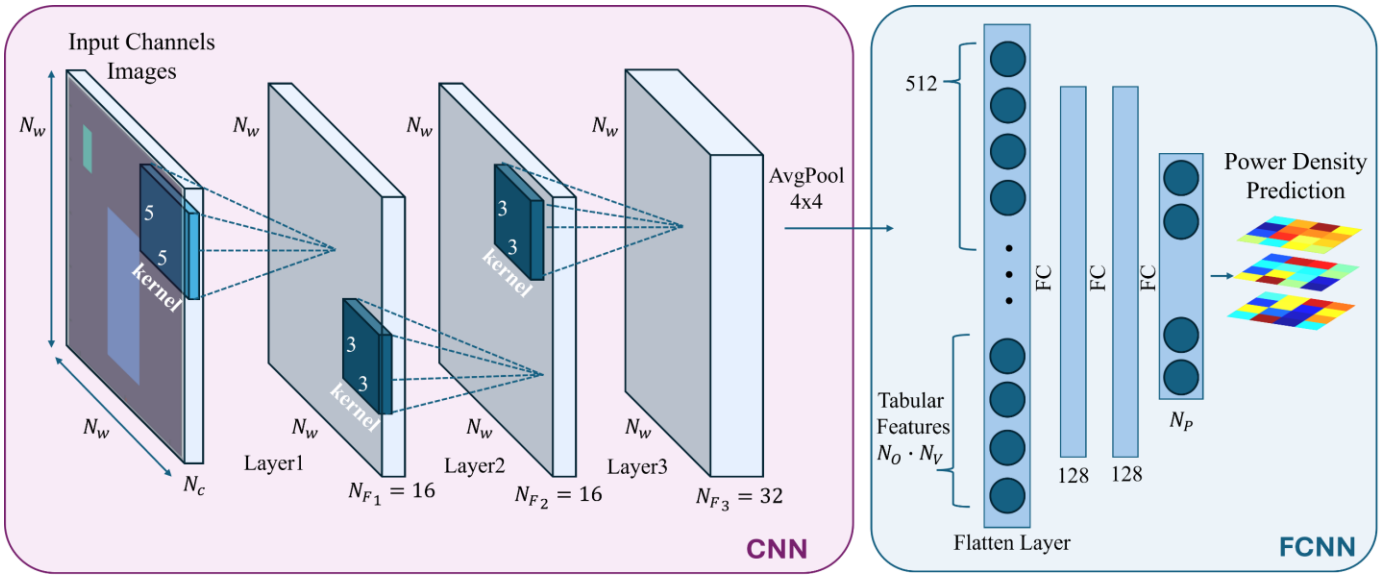

Fig. 3. CNN schematic architecture, corresponding to the optimal configuration identified when the MPW loss function is used: the input tensor is processed by three convolutional layers, and the resulting feature descriptor is combined with tabular inputs and passed through two fully connected layers to produce power density values across the three planes.

IV. Results

To quantitatively assess the performance of the proposed models and evaluate the influence of the selected loss functions, the training and testing phases were carefully monitored. Each environment EM simulation requires approximately one hour of computational time on an Intel® Xeon® CPU E5-2620 v4 @ 2.10 GHz with 64 GB of random-access memory (RAM). Nonetheless, selecting the most informative samples using an active learning strategy significantly decreases the computational effort while maintaining model performance. Each network is trained on a dataset consisting of 800 samples, partitioned into 80% and 20%, for testing and training, respectively, ensuring a balanced representation of the parameter space, with a training time of approximately 2 minutes for all the loss functions. Prior to training, all tabular features were standardized using standard normalization to improve numerical stability and convergence behavior. Model optimization was carried out using the Adam optimizer with a fixed learning rate of 0.001 and a mini-batch size of 64, empirically identified as providing the most stable convergence across all tested loss functions. The accuracy of the regression task is assessed using the Root Mean Squared Error (RMSE), which provides a direct physical interpretation of the prediction error in terms of power density, quantifying how closely the NN predictions match the corresponding reference values obtained from EM simulations, effectively measuring the average magnitude of the prediction error across all output samples. During training, the RMSE is computed at the end of each epoch – defined as a complete iteration over the entire training dataset – on both the training and validation sets, allowing monitoring of the model's convergence and generalization. This per-epoch evaluation shows how the prediction error evolves over time, indicating whether the network is learning the underlying patterns or starting to overfit. The RMSE is computed spanning all the expected power density outputs y_{ij}^{NN} during the training and validation procedure as:

$$RMSE = \sqrt{\frac{1}{N_S N_P} \sum_{i=1}^{N_S} \sum_{j=1}^{N_P} (y_{ij}^{EM} - y_{ij}^{NN})^2}, \qquad (4)$$

where N_s is the number of the chosen training or validation sets and N_P=48 is the number of output pixels. y_{ij}^{EM} is the actual value for the i-th sample and j-th pixel resulting from the EM simulator and y_{ij}^{NN} is the correspondent predicted value from the NN. Fig. 4 reports the evolution of the training and validation losses for the considered loss functions. In every case, the networks exhibit smooth convergence after approximately 50 epochs, with validation losses stabilizing around an RMSE of about 2 W/m², confirming consistent learning behavior across the different training objectives. Among the considered loss metrics, the MPW function achieves the lowest training loss, indicating a stronger capability to minimize the objective function on the training set. However, this advantage does not directly translate to superior generalization, as the MRE-based model attains the lowest validation and test errors. This suggests that while MPW effectively captures high-intensity regions – where higher power densities dominate the objective – it may overfit those regions at the expense of overall balance across the spatial

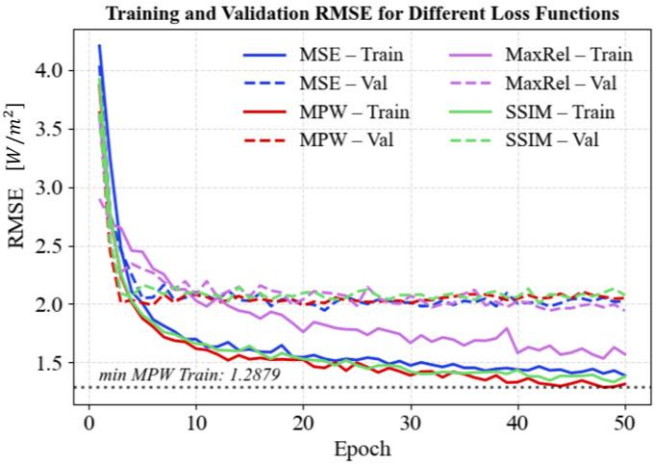

Fig. 4. Overall CNN performance showing the evolution of RMSE on training and testing sets across 50 epochs for the different loss functions. Convergence is reached within 50 epochs, with test RMSE around 2 W/m^2 for all loss functions.

domain. Conversely, the MRE objective, which normalizes errors by the true magnitude, promotes a more uniform relative accuracy across pixels of varying intensity, leading to slightly improved performance on unseen data. A more detailed analysis was conducted by computing the per-pixel RMSE, which quantifies the local reconstruction accuracy of each pixel in the predicted power map. Unlike the global RMSE (4), which aggregates the error across all outputs, this metric highlights spatial variations in prediction performance across the grid. The results are reported in Fig. 5 and show that the MSE and SSIM objectives achieve the lowest per-pixel RMSE, indicating superior spatial consistency and a better ability to reproduce fine variations in the power distribution. Conversely, the MPW objective, despite reaching the lowest overall training loss, exhibits slightly higher pixel-level error dispersion, as its magnitude-weighted formulation emphasizes regions of higher

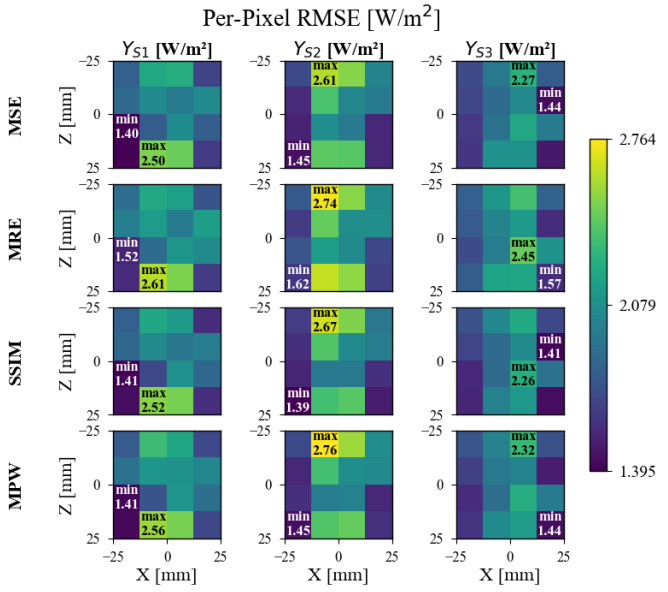

Fig. 5. Overall CNN performance in terms of per-pixel RMSE across the predefined sensing planes for the different loss functions, highlighting the spatial distribution of prediction errors.

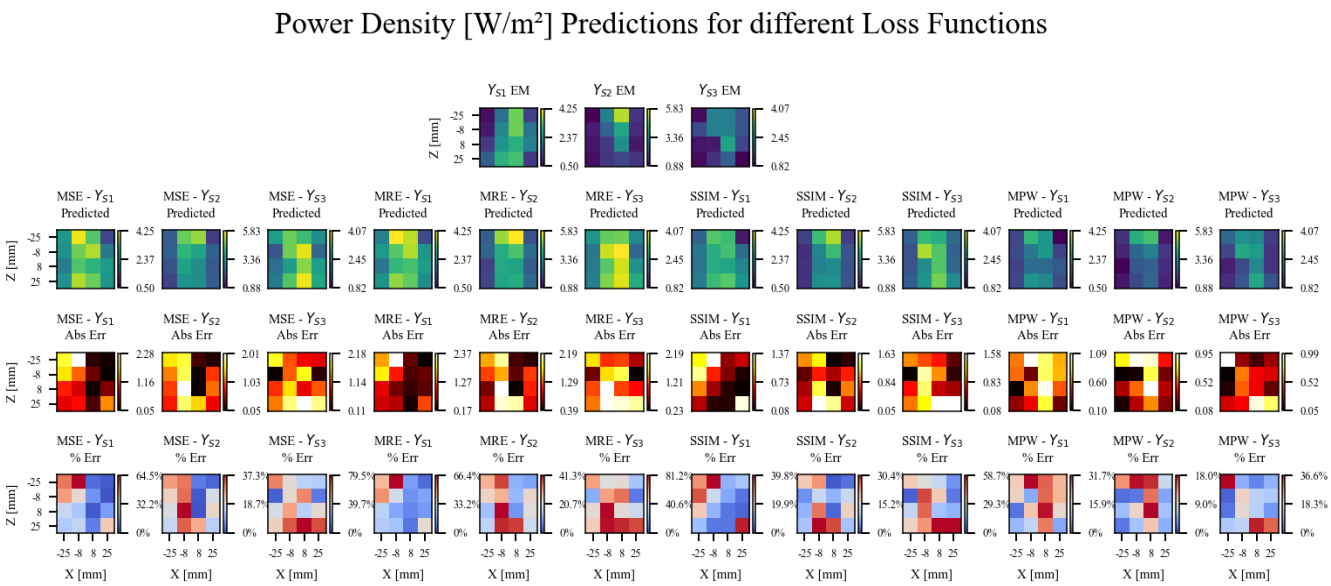

Fig. 6. Prediction results of the trained CNN showing the comparison between EM-simulated and predicted power density across multiple sensing planes. The network was trained using different loss functions, including MSE, MRE, SSIM, and MPW. In this case, MSE and MRE exhibit higher errors, while SSIM and MPW achieve better results, more accurately capturing the spatial propagation patterns.

power intensity during optimization. The maximum local errors are consistently observed in the pixels closest to the transmitting source, where the power density reaches its peak values and exhibits an extended range of variability. Overall, these results confirm that losses emphasizing structural fidelity, such as SSIM, achieve a better compromise between global accuracy and spatial detail preservation in EM field reconstruction. Fig. 6 illustrates the power density predictions obtained from CNN models trained with the different loss functions for the environment shown in Fig. 2(a). The prediction performance for each sample, corresponding to the j-th pixel, is evaluated in terms of absolute error between the EM-simulated and the NN-predicted power at that pixel. To account for the dynamic range of power distribution, the error is also expressed as a percentage, calculated by normalizing the absolute difference with respect to the maximum power detected in the sample. It can be observed that both the MSE- and MRE-trained networks effectively capture the overall propagation behavior; however, they tend to systematically overestimate the absolute power levels throughout the domain, leading to higher field amplitudes even in low-exposure regions. Conversely, the SSIM- and MPW-based models produce more physically consistent reconstructions. The custom MPW objective exhibits the best overall performance in this case, yielding the lowest absolute and percentage errors among all tested configurations. Notably, the SSIM and MPW losses demonstrate enhanced robustness in scenarios where maximum power regions appear in atypical spatial locations, effectively capturing propagation patterns that other loss functions often fail to reproduce. Conversely, it has been observed that in more typical propagation scenarios MSE and MRE loss functions tend to provide slightly more precise predictions of the power density at each pixel. Figure 7 presents another sample, corresponding to a scenario where the source is partially obstructed by a bulky object, significantly modifying the expected propagation pattern. The 3D environment and its 2D projection are shown in Fig. 7(a) and Fig. 7(b), respectively, while the MPW-predicted power distribution is displayed in Fig. 7(c). In this more complex setting, the MPW-trained network achieves the best performance, accurately capturing the altered spatial

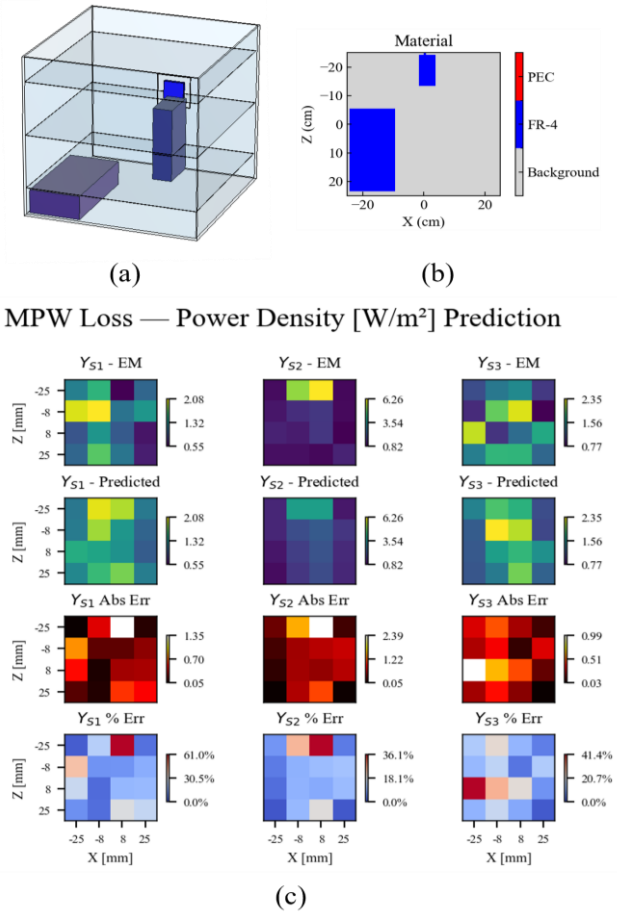

Fig. 7. (a) 3D environment with the source partially obstructed by a bulky obstacle and (b) its corresponding 2D projection. (c) Predicted power density distribution obtained using the MPW-trained network.

distribution of the power density. For clarity, only the MPW predictions are shown, as they effectively reproduce the overall propagation behavior. Importantly, only a few points in the map exhibit higher errors, demonstrating that the network preserves high accuracy across most of the domain, even in the presence of substantial obstacles.

A. Fully Connected Neural Network

As a preliminary approach, FCNNs were used to estimate spatial power distributions from EM simulations, but CNNs consistently outperformed them. This is because FCNNs operate on tabular features, whereas CNNs exploit the spatial structure of image-based inputs, capturing local patterns and correlations that FCNNs cannot. Consequently, CNNs achieve higher accuracy, particularly in configurations with strong spatial dependencies. MAE comparisons using an 80–20 training-testing split confirm this trend, with CNNs showing lower errors across all loss functions, especially MSE and MPW. FCNNs trained with MSE tend to produce homogenized predictions, while MPW emphasizes high-power regions but still yields larger errors. The FCNN architectures correspond to those listed in Table 1, and their structures also form the feedforward stage of the CNN models. The learning curves further substantiate these observations, showing systematically lower MAE values for CNNs across all loss functions, as shown in Fig. 8. The improvement is especially evident for the MSE and MPW loss functions. FCNNs trained with MSE loss tend to homogenize the predicted power distribution, producing spatially averaged patterns that fail to represent complex variations, which consequently leads to larger errors. The MPW loss, being a weighted form of MSE, better emphasizes regions of higher power and spatial contrast but still leads to a large error. CNNs and FCNNs were further compared by tracking learning performance over successive iterations of the active pool sampling algorithm (Fig. 9). Models were trained on progressively larger subsets ranging from 200 to 800 samples, with evaluations from iteration $n=2$ to $n=8$ using a fixed test set for consistency. At small sample sizes, CNNs initially show higher errors due to greater complexity and more parameters. As the dataset grows, CNNs leverage spatial correlations and local dependencies, reducing errors faster. By the final iteration, CNNs consistently achieve lower MAE, demonstrating superior ability to model spatial propagation and reconstruct detailed, physically meaningful power distributions compared to FCNNs.

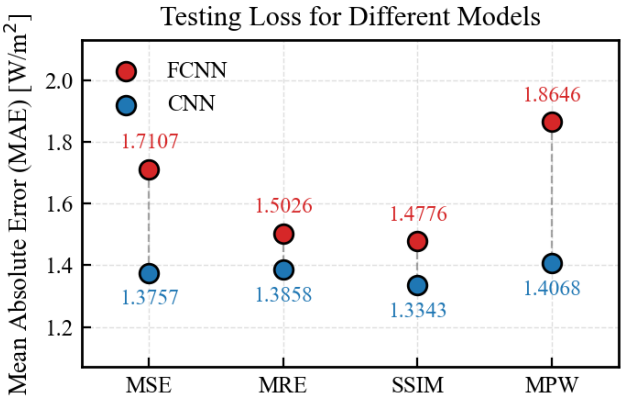

Fig. 8. Mean Absolute Error for the different loss functions comparing FCNN and CNN.

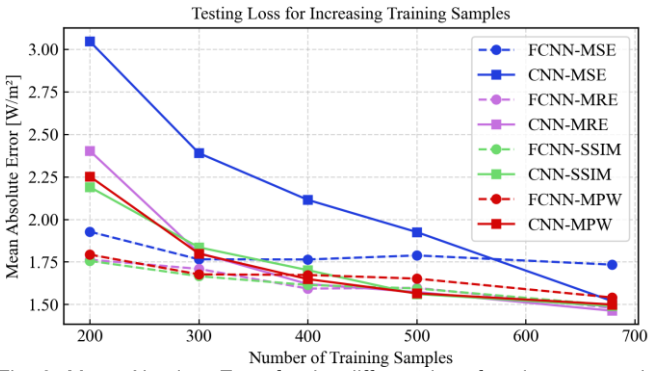

Fig. 9. Mean Absolute Error for the different loss functions comparing FCNN and CNN for increasing number of training samples.

V. Experimental Validation

To experimentally validate the predictive capability of the proposed models and the optimized EM simulation technique, a controlled setup was constructed to replicate the propagation environment. The enclosure was assembled using aluminum walls and obstacles to reproduce realistic reflection and scattering conditions which in the simulation corresponded to PEC. The EM simulations retrieved with aluminum instead of PEC did not show significant differences. A patch antenna operating at 2.45 GHz was connected to a signal generator and positioned on one side of the enclosure at the prescribed location, serving as the transmitting source. On the opposite side, a monopole antenna acted as a sensing element, used to measure the received power levels at multiple spatial locations inside the box. The signal strength received was recorded using a spectrum analyzer, capturing the average power detected over a fixed acquisition period (15-seconds window) at each of the 48 predefined measurement points distributed across three sensing planes. These measurements enabled the reconstruction of the experimental power density maps within the enclosure. The complete experimental setup is illustrated in Fig. 10, together with a highlighted view of one of the three measurement planes. One of the testing samples was selected for reproduction, with obstacles reconstructed to match the geometry of the numerical model. Since both obstacles are assumed to be metallic, they were covered with aluminium foil to emulate their conductive

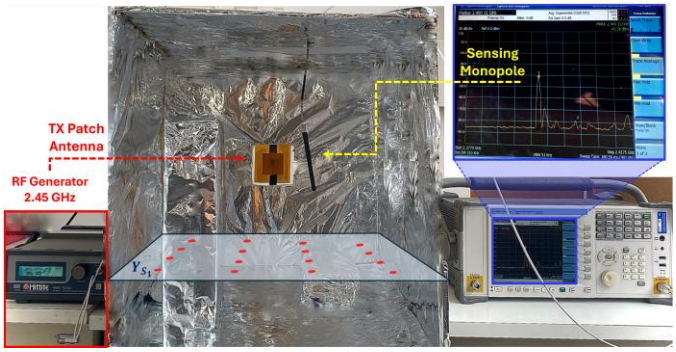

Fig. 10. Measurement setup replicating the metallic enclosure environment. The transmitting patch antenna is connected to an RF generator, and the received field strength is captured using a monopole probe connected to a spectrum analyzer.

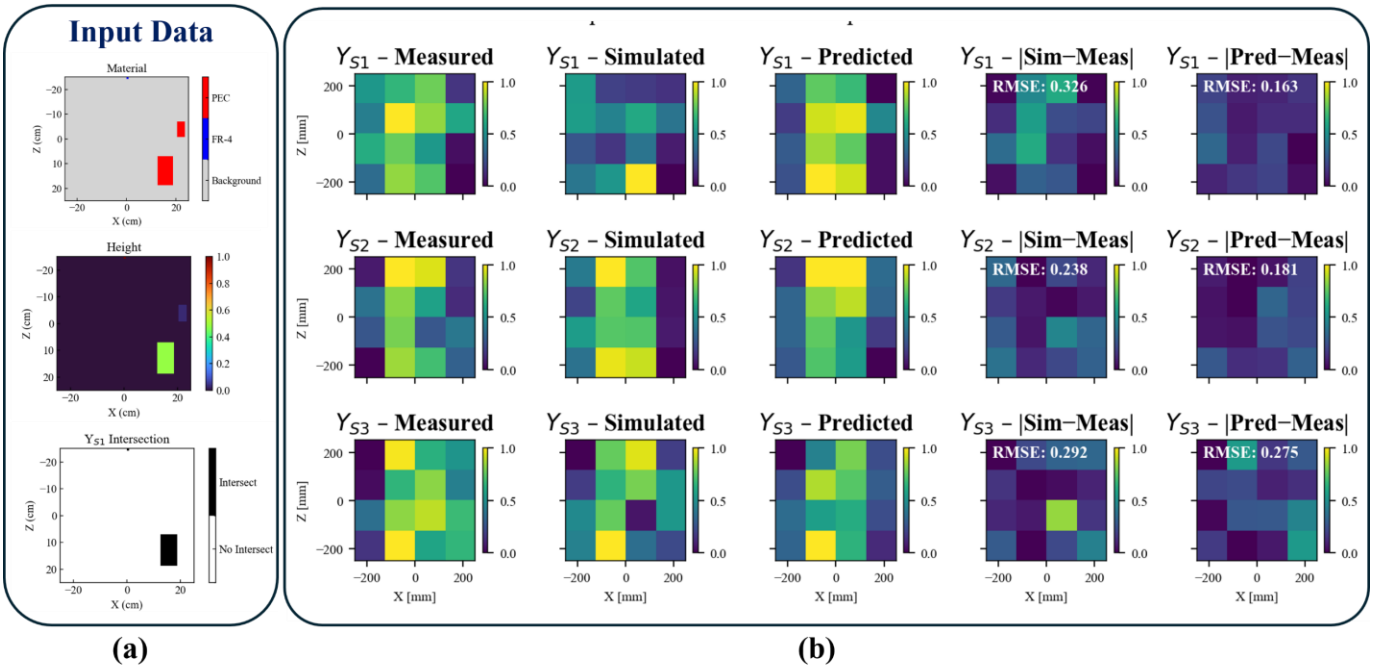

Fig. 11. (a) Channel representation of sample reproduced experimentally and (b) comparison between measured, simulated and predicted results normalized with respect to their maxima.

behaviour. The selected input sample is shown in Fig. 11(a), together with a comparison of the measured, simulated, and predicted power distributions (Fig. 11(b)). To ensure a fair comparison across these datasets, each measured power map was normalized with respect to its maximum value, emphasizing the relative spatial distribution of the EM power rather than absolute magnitude. Overall, the predicted power maps demonstrate strong qualitative agreement with both measured and simulated distributions, effectively reproducing the main propagation lobes and attenuation regions, as confirmed by the normalized RMSE values reported in the overall map of Fig. 11(b). The observed discrepancies between the measured and predicted results are mainly due to the real material used for the obstacles and enclosure in the experimental validation and are reflected as well with respect to the predicted results. Furthermore, while the predicted power density is averaged over each pixel area, in the experimental setup it is recorded at a single point located approximately at the center of each pixel. This mismatch in spatial sampling inherently introduces deviations, particularly in regions adjacent to obstacles, where the simulated fields include zero-valued areas that reduce the average power within those pixels. In any case, the NN provides a coherent reconstruction of the power distribution, effectively capturing the dominant propagation patterns within the environment. Its convolutional architecture naturally produces smooth spatial outputs, emphasizing large-scale trends while mitigating sensitivity to local variations or peaks. Importantly, this predictive capability enables the determination of optimal sensor locations without requiring direct measurements of the environment, allowing for efficient and non-intrusive sensor placement in complex scenarios.

VI. Conclusion

This work has proposed a method based on CNNs for predicting the best mutual positioning of RF power sources and battery-less sensors for WPT in enclosed and complex environments, where the far-field approximation is not applicable. Such conditions are common in industrial sites or within machine components equipped with multiple wireless, battery-free sensors that are difficult to access, as often encountered in the IIoT. A data-efficient pipeline, combining active sampling with IE-based EM simulations, was developed to generate representative training datasets while minimizing computational effort. Experimental validation within a controlled metallic enclosure further confirmed the reliability of the proposed models. While the dataset used in this work was intentionally compact to limit computational cost, expanding it to encompass a broader range of geometries, materials, and frequencies – and integrating measured data with simulated samples – could further enhance the model's accuracy and robustness. Ultimately, the proposed CNN-based framework establishes a powerful and scalable approach for fast, data-driven estimation of EM field propagation in complex environments. Its ability to provide near real-time predictions can be instrumental in optimizing sensor placement, ensuring reliable wireless power delivery, and improving overall network efficiency in dense industrial scenarios. By bridging physics-based modeling and deep learning, this approach paves the way toward intelligent, adaptive digital twins of complex EM environments capable of supporting the design and optimization of next-generation wireless power and communication systems.

Appendix I: Loss Functions

Let $P^{EM} \in \mathbb{R}^{N_P}$ be the ground-truth power vector from EM simulations and $P^{NN} \in \mathbb{R}^{N_P}$ the NN prediction ($N_P = 48$). The baseline MSE is defined as:

$$MSE = \frac{1}{N_S N_P}\sum_{i=1}^{N_S}\sum_{j=1}^{N_P}\left(P_{i,j}^{EM} - P_{i,j}^{NN}\right)^2 \quad (5)$$

with $P_{i,j}^{EM}$ denoting the ground-truth target for the i-th output, and $j = 1, \dots, N_P$ representing the total number of pixels of the i-th sample, with $i = 1, \dots, N_S$ and N_S number of samples. The MRE is normalized with respect to the maximum power density value in the sample is defined as

$$MRE = \frac{1}{N_S N_P}\sum_{i=1}^{N_S}\sum_{j=1}^{N_P}\left(\frac{P_{i,j}^{EM} - P_{i,j}^{NN}}{\max_j P_{i,j}^{EM}}\right)^2. \quad (6)$$

reducing the bias toward high-intensity regions. To better preserve the spatial structure of the EM field, the Structural Similarity Index (SSIM) [19] is adopted as a complementary loss metric. SSIM compares two images by jointly assessing their luminance, contrast, and structural similarity, yielding a value between 0 and 1, with 1 indicating a perfect match. Unlike purely pixel-wise metrics, it accounts for spatial context, making it more effective in evaluating smooth variations and dominant patterns in power maps. For each sample, SSIM loss is defined as:

$$\mathcal{L}_{SSIM_i} = 1 - SSIM\left(P_j^{NN}, P_j^{EM}\right), \quad (7)$$

so that lower values correspond to better structural agreement. Finally, the MPW loss is defined, to enforce the error on the pixel with highest power:

$$MPW = (1-\gamma)\cdot MSE + \gamma\left(MSE_{max} + D_{pixel}\right). \quad (8)$$

Here $\gamma \in [0,1]$ balances the contribution between global accuracy and local spatial alignment and through empirical tuning, the optimal value was determined to be $\gamma = 0.2$. For each i-th sample let $j^* = \arg\max_j y_{i,j}^{EM}$ denote the index of the pixel with maximum ground-truth power. The error at the maximum power pixel is then:

$$MSE_{max} = 1/N_S \sum_{i=1}^{N_S}\left(P_{i,j^*}^{EM} - P_{i,j^*}^{NN}\right)^2 \quad (9)$$

and pixel-distance penalty measures the Euclidean distance between the predicted and true positions of maximum power, with each pixel j assigned 2D coordinates $r_j = (x_j, z_j)$ on the discretized output grid. The spatial deviation is defined as:

$$D_{pixel} = 1/N_S \sum_{i=1}^{N_S}\left\| r_{i,j^*}^{EM} - r_{i,j^*}^{NN}\right\|_2. \quad (10)$$

This additional constraint penalizes models that correctly predict the intensity pattern but misplace its spatial peak, thus promoting more physically consistent outputs.

ACKNOWLEDGMENT

This study was carried out partly within the Spoke 13 of the MOST (Sustainable Mobility National Research Center) and received funding from the European Union under the NextGenerationEU programme (National Recovery and Resilience Plan (NRRP) – Mission 4 Component 2, Investment 1.4 – D.D. 1033 17/06/2022, CN00000023) and partly by the European Union under the NextGenerationEU programme under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3 (CUP: J33C22002880001), partnership on "Telecommunications of the Future" (PE00000001 – program "RESTART").

REFERENCES

[1] H. Boyes *et al*., "The industrial internet of things (IIoT): An analysis framework," *Comput. Ind.*, vol. 101, pp. 1–12, May 2018, doi: 10.1016/j.compind.2018.04.015.

[2] W. J. Fleming, "New automotive sensors—A review," *IEEE Sens. J.*, vol. 8, no. 11, pp. 1900–1921, Nov. 2008, doi: 10.1109/JSEN.2008.2006452.

[3] S. A. Huda, *et al*., "Wireless power transfer in wirelessly powered sensor networks: A review of recent progress," *Sensors*, vol. 22, no. 8, Art. no. 2952, Apr. 2022, doi: 10.3390/s22082952.

[4] G. Paolini *et al*., "RF-powered low-energy sensor nodes for predictive maintenance in electromagnetically harsh industrial environments," *Sensors*, vol. 21, no. 2, pp. 386–403, Jan. 2021, doi: 10.3390/s21020386.

[5] E. Augello, *et al*., "Optimizing sensors placement and wireless power transfer in harsh electromagnetic environments," in *Proc. IEEE Int. Symp. Meas. Netw.* (M&N), Rome, Italy, 2024, pp. 1–6, doi: 10.1109/MN60932.2024.10615906.

[6] U. Bas and S. C. Ergen, "Ultra-wideband channel model for intra-vehicular wireless sensor networks beneath the chassis: From statistical model to simulations," *IEEE Trans. Veh. Technol.*, vol. 62, no. 1, pp. 14–25, Jan. 2013, doi: 10.1109/TVT.2012.2215969.

[7] L. Tian *et al*.,"Semi-deterministic radio channel modeling based on graph theory and ray-tracing," *IEEE Trans. Antennas Propag.*, vol. 64, no. 6, pp. 2475–2486, Jun. 2016.

[8] H. M. El Misilmani, *et al*., "A review on the design and optimization of antennas using machine learning algorithms and techniques," *Int. J. RF Microw. Comput.-Aided Eng.*, vol. 30, no. 2, e22356, 2020, doi: 10.1002/mmce.22356

[9] M. Luo and K. Huang, "Prediction of the electromagnetic field in metallic enclosures using artificial neural networks," *Prog. Electromagn. Res.*, vol. 116, pp. 171–184, 2011, doi: 10.2528/PIER11031101.

[10] A. Seretis and C. D. Sarris, "Toward physics-based generalizable convolutional neural network models for indoor propagation," *IEEE Trans. Antennas Propag.*, vol. 70, no. 6, pp. 4112–4126, Jun. 2022, doi: 10.1109/TAP.2021.3138535.

[11] D. Piromalis and A. Kantaros, "Digital twins in the automotive industry: The road toward physical–digital convergence," *Appl. Syst. Innov.*, vol. 5, no. 4, p. 65, 2022, doi: 10.3390/asi5040065.

[12] L. Li, *et al*., "Digital twins of electromagnetic propagation environments for live 5G networks—Part I: Channel acquisition, EM simulation, and verification," *IEEE Trans. Antennas Propag.*, vol. 73, no. 4, pp. 2053–2064, Apr. 2025, doi: 10.1109/TAP.2024.3494260.

[13] E. Augello, *et al*., "Modular artificial neural networks for wireless power transfer optimization in sensor-driven industrial IoT," in *Proc. IEEE Wireless Power Technol. Conf. Expo* (WPTCE), 2025, pp. 1–5.

[14] A. Costanzo, *et al*., "RF/baseband co-design of switching receivers for multiband microwave energy harvesting," *Sensors Actuators A Phys.*, vol. 179, pp. 158–168, Jun. 2012, doi: 10.1016/j.sna.2012.02.005.

[15] E. Augello, *et al*., "An efficient approach for a digital twin model of harsh EM environments," in 2024 IEEE Int. Symp. Antennas Propag. & INC/USNC-URSI Radio Sci. Meeting (AP-S/INC-USNC-URSI), Firenze, Italy, Jul. 2024, pp. 1–2.

[16] Z. Liu, *et al*., "Pool-based unsupervised active learning for regression using iterative representativeness-diversity maximization (iRDM)," *Pattern Recognit. Lett.*, vol. 142, pp. 11–19, 2021, doi: 10.1016/j.patrec.2020.11.019.

[17] M. Shanawani, *et al*., "Efficient simulation method for wireless power transfer," in *2019 IEEE Int. Conf. RFID Technol. Appl.* (RFID-TA), Pisa, Italy, 2019, pp. 71–75, doi: 10.1109/RFID-TA.2019.8892029.

[18] K. Bleszynski, *et al*., "A fast integral equation solver for electromagnetic scattering problems," in *Proc. IEEE Antennas Propag. Soc. Int. Symp.* and URSI Nat. Radio Sci. Meeting, Seattle, WA, USA, vol. 1, pp. 416–419, 1994.

[19] Z. Wang, *et al*., "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004, doi: 10.1109/TIP.2003.8198611.

[20] J. Lederer, "Activation functions in artificial neural networks: A systematic overview," *arXiv preprint*, arXiv:2101.09957, 2021.