Assessing Wireless Sensing Potential with Large Intelligent Surfaces

Vaca Rubio, Cristian Jesús; Espinosa, Pablo Ramirez; Kansanen, Kimmo; Tan, Zheng-Hua; De Carvalho, Elisabeth; Popovski, Petar

Published in:
IEEE Open Journal of the Communications Society

DOI (link to publication from Publisher):
10.1109/OJCOMS.2021.3073467

Publication date:
2021

Document Version
Publisher's PDF, also known as Version of record

Received 18 February 2021; revised 22 March 2021; accepted 9 April 2021. Date of publication 15 April 2021; date of current version 20 April 2021.

Digital Object Identifier 10.1109/OJCOMS.2021.3073467

Assessing Wireless Sensing Potential With Large Intelligent Surfaces

CRISTIAN J. VACA-RUBIO [1], PABLO RAMIREZ-ESPINOSA [1], KIMMO KANSANEN [1,2] (Member, IEEE), ZHENG-HUA TAN [1] (Senior Member, IEEE), ELISABETH DE CARVALHO [1] (Senior Member, IEEE), AND PETAR POPOVSKI [1] (Fellow, IEEE)

[1]Department of Electronic Systems, Aalborg University, 9220 Aalborg, Denmark

[2]Norwegian University of Science and Technology, 7491 Trondheim, Norway

CORRESPONDING AUTHOR: C. J. VACA-RUBIO (e-mail: cjvr@es.aau.dk)

This work was supported by the European Unions Horizon 2020 Research and Innovation Programme through the Marie Skłodowska-Curie Grant under Agreement 813999; in part by EU H2020 RISE-6G project; and in part by the Danish Council for Independent Research under Grant DFF-701700271.

A preliminary version of this work has been accepted to 15th EAI International Conference on Cognitive Radio Oriented Wireless Networks (CROWNCOM), 2020 [1].

ABSTRACT Sensing capability is one of the most highlighted new feature of future 6G wireless networks. This paper addresses the sensing potential of Large Intelligent Surfaces (LIS) in an exemplary Industry 4.0 scenario. Besides the attention received by LIS in terms of communication aspects, it can offer a high-resolution rendering of the propagation environment. This is because, in an indoor setting, it can be placed in proximity to the sensed phenomena, while the high resolution is offered by densely spaced tiny antennas deployed over a large area. By treating an LIS as a radio image of the environment relying on the received signal power, we develop techniques to sense the environment, by leveraging the tools of image processing and machine learning. Once a radio image is obtained, a Denoising Autoencoder (DAE) network can be used for constructing a super-resolution image leading to sensing advantages not available in traditional sensing systems. Also, we derive a statistical test based on the Generalized Likelihood Ratio (GLRT) as a benchmark for the machine learning solution. We test these methods for a scenario where we need to detect whether an industrial robot deviates from a predefined route. The results show that the LIS-based sensing offers high precision and has a high application potential in indoor industrial environments.

INDEX TERMS Computer vision, industry 4.0, large intelligent surfaces, machine learning, sensing.

I. INTRODUCTION

MASSIVE multiple-input multiple-output (MIMO) is one of the essential technologies in the 5th generation of wireless networks (5G) [2]. Compared with traditional multiuser MIMO systems, the base station of a massive MIMO system is equipped with a large number of antennas, aiming to further increase spectral efficiency [3]. Looking towards the 6th generation of wireless networks (6G), there are some significant breakthroughs on the design of reprogramable metamaterials [4], giving raise to new concepts such as *holographic MIMO surfaces* (*HMIMO*) [5], *large intelligent surface* (*LIS*) [6] and *reconfigurable intelligent surface* (*RIS*) [7]. While HMIMO and LIS originally refer to the use of continuous radiating surfaces where the received electromagnetic field is recorded and ultimately reconstructed [8], in practice an LIS is envisioned and regarded as a collection of closely spaced tiny antenna elements. On the other hand, RIS are composed by small passive reflectors embedded in a surface, allowing to arbitrarily modify the phase of the impinging electromagnetic waves and thus enabling a smart control of the propagation environment [9]–[11].

The performance analysis of LIS and RIS assisted systems has attracted considerable attention in the recent years, and many works have studied the applicability of these technologies. For instance, the use of RIS to control the signals propagation has been analyzed in the context of communications through the so-called passive beamforming [7], [12]–[14], in location and positioning systems [15]–[17], and in physical layer security [18], [19]. The combination of Deep Learning (DL) and RIS elements efficient reconfiguration has also been studied in [20]. In turn, LIS are considered as a natural extension of massive MIMO, and their potential for data transmission [6], [21] and positioning [22], [23] has been also addressed.

However, the potential of LIS could go beyond communications applications. Indeed, such large surfaces contain many antennas that can be used as sensors of the environment based on the channel state information (CSI).

Sensing strategies based on electromagnetic signals have been thoroughly addressed in the literature in different ways, and applied to a wide range of applications. For instance, in [24], a real-time fall detection system is proposed through the analysis of the communication signals produced by active users, whilst the authors in [25] use Doppler shifts for gesture recognition. Radar-like sensing solutions are also available for user tracking [26] and real-time breath monitoring [27], as well as sensing methods based on radio tomographic images (RTI) [28], [29]. Interestingly, whilst some of these techniques resort solely on the amplitude (equivalently, power) of the receive signals [26], [29], in the cases where sensing small scale variations is needed, the full CSI (i.e., amplitude and phase of the impinging signals) is required [27], [28]. Moreover, in terms of power-based radio maps generation, some indoor positioning strategies leverage the use of machine learning (ML) solutions [30] based on the Received Signal Strength (RSS) of several beforehand known anchors for localization purposes.

On a related note, ML based approaches are gaining popularity in the context of massive MIMO, mainly due to the inherent complexity of this type of systems and their sensitivity to hardware impairments and channel estimation errors. Hence, DL techniques arise as a promising solution to deal with massive MIMO, and several works have shown the advantages of ML solutions in channel estimation and precoding [13], [31]–[36]. Due to the even larger dimensions of the system in extra-large arrays, DL may play a key role in exploiting complex patterns of information dependency between the transmitted signals. Also, the popularization of LIS as a natural next step from massive MIMO gives rise to larger arrays and more degrees of freedom, providing huge amount of data which can feed ML algorithms.

Despite all the available works dealing with beyond massive MIMO and sensing, both topics have been addressed rather separately. This has motivated the present work, where the objective is to assess the potential of the combined use of DL algorithms and large surfaces for the purpose of sensing the propagation environment. To that end, the received signal along with the LIS is treated as a radio image of the propagation environment, giving raise to the use of image processing techniques to improve the performance of sensing systems beyond purely radio-based approaches. Also, we analyze the pros and cons of this image sensing proposal, comparing it to alternative solutions based on classical post-processing of the received radio signal. Specifically, the contributions of this work are summarized as follows:

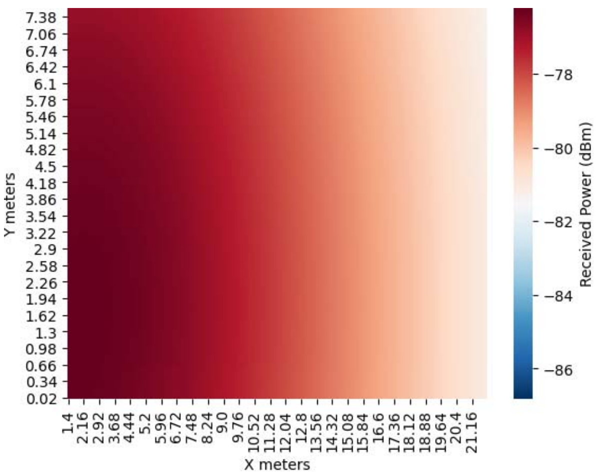
- We propose an image-based sensing technique based on the received signal power at each antenna element of an LIS. These power samples are processed to generate a high resolution image of the propagation environment that can be used to feed ML algorithms to sense large-scale events. The usage of received signal power would lead to simple deployments, since there is no need of coherent receivers.
- A ML algorithm, based on transfer learning and local outlier factor (LOF), is defined to process the radio images generated by the LIS in order to detect anomalies over a predefined robot route.
- We show the advantage of representing the radio propagation environment as an image, allowing us to use a denoising autoencoder network (DAE) for augmenting image resolution and significantly increasing the performance of the system.
- We derive a statistical test, based on the classical generalized likelihood ratio test (GLRT), to carry out the same sensing task, and perform a comparison with the ML solution in terms of generality, performance and further potential applications.

To evaluate the capabilities for sensing of LIS, we consider a simple problem of route anomaly detection in an indoor industrial scenario. Hence, we analyze the feasibility of this proposal to determine whether a robot has deviated from its predefined route, and compared it with the here derived statistical solution.

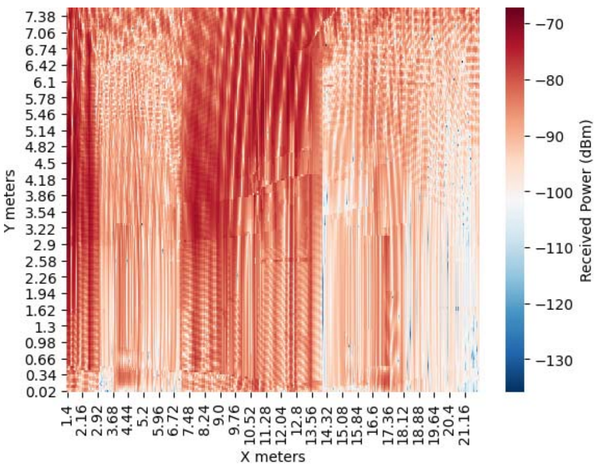
The reminder of the paper is organized as follows. Section II introduces the concept of sensing based on radio images. Then, Section III presents the problem of robot deviation detection in an industrial setup. The classical solution based on hypothesis testing is derived and characterized in Section IV and the proposed ML algorithm is detailed in Section V. With the ML solution presented, the model validation is carried out in Section VI, whilst simulated results are discussed in Section VII. Finally, some conclusions are drawn in Section VIII.

II. RADIO IMAGE-BASED LIS SENSING

In a wireless context, a LIS could be described as a structure which uses electromagnetic signals impinging in a determined scatterer in order to obtain a profile of the environment. That is, we can use the resulting signal of the superposition of all the involved paths that imping into every of the antenna elements conforming the surface. Then, the power of the resulting superimposed signal is used to obtain a high resolution image of the propagation environment. Note the LIS elements are using the CSI information as envelope detectors, as no phase estimation is needed but the received signal power. Using this approach, the complexity of the multipath propagation is reduced to using information represented as an image. This provides a twofold benefit: *i)* the massive number of elements that compose the LIS leads to



(a) LoS, noiseless.



(b) Real scenario, noiseless.

FIGURE 1. Radio images for line-of-sight (LoS) and Industry scenarios.

an accurate environment sensing (i.e., high resolution image), and *ii)* it allows the use of computer vision algorithms and ML techniques to deal with the resulting images.

As an illustrative example, an LIS is deployed in a wall along a 22×8 m physical aperture, containing antenna elements separated $\frac{\lambda}{2}$ cm while an arbitrary robot is transmitting a sensing signal. Fig. 1 shows the LIS radio images obtained from different propagation environments under this setup. Specifically, Fig. 1(a) corresponds to an LoS propagation (no scatterers), whilst Fig. 1(b) is obtained from an industrial scenario with a rich scattering. Note that, in the case in which different scatterers are placed, their position and shapes are captured by the LIS and represented in the image. Beyond that, LIS-based imaging does not need of a previous calibration period as well as no scatterers need to be modelled to be captured in the radio image, contrary to other wireless image reconstruction techniques that rely on the received signal power, such as RTI [28], [29]. To the best of the authors' knowledge, this is the first time that LIS image based wireless sensing is proposed in the literature.

III. SYSTEM MODEL AND PROBLEM FORMULATION

We stick to a simple baseline problem in order to analyze, for first time in the literature, the sensing potential of LIS. To that end, let consider an industrial scenario where a robot is supposed to follow a fixed route. Assume that, due to arbitrary reasons, it might temporarily deviate from the predefined route and follow an alternative (undesired) trajectory. The goal is to be able to detect whether the robot is following the correct route or not, based on the sensing signal transmitted by the target device.

In order to perform the detection, we assume that a LIS, consisting of M antenna elements, is placed along a wall. The sensing problem reduces to determine, from the received signal at each of the M LIS elements, whether the transmission has been made from a point at the desired route or from anomalous ones. Formally, if we define the correct trajectory as the set of points in space $\mathbf{P}_c \in \mathbb{R}^{N_p \times 3} = (\mathbf{p}_1, \ldots, \mathbf{p}_{N_p})$, and the received complex signal from an arbitrary point, $\mathbf{p}_k \in \mathbb{R}^{1\times 3}$, as $\mathbf{y}_k \in \mathbf{C}^{M\times 1}$, then the problem reduces to estimating whether $\mathbf{p}_k \in \mathbf{P}_c$ based on $\mathbf{y}_k$. Note that this formulation can be generalized to any anomalous detection based on radio-waves in an arbitrary scenario.

The complex baseband signal received at the LIS from point $\mathbf{p}$ (the subindex is dropped for the sake of clarity in the notation) is given by

$$\mathbf{y} = \mathbf{h}x + \mathbf{n}, \tag{1}$$

with x the transmitted (sensing) symbol, $\mathbf{h} \in \mathbb{C}^{M\times 1}$ the channel vector from point $\mathbf{p}$ to each antenna-element, and $\mathbf{n} \sim \mathcal{CN}_M(\mathbf{0}, \sigma^2\mathbf{I}_M)$ the noise vector. Moreover, we consider a static scenario where the channel $\mathbf{h}$ only depends on the user position, neglecting the impact of time variations.

In order to reduce deployment costs, and because we are interested in sensing large scale variations, we consider the received signal amplitude (equivalently, power). This assumption may lead to cheaper and simpler system implementations, avoiding the necessity of performing coherent detection.

IV. STATISTICAL APPROACH: LIKELIHOOD RATIO TEST

A. DECISION RULE

Let us consider that the system is able to obtain several samples from each point $\mathbf{p}_k$ belonging to the correct route $\mathbf{P}_c$ during a training phase. Then, once the system is trained, the problem can be tackled from a statistical viewpoint by performing a generalized hypothesis test, as shown throughout this section.

To start with, let us assume that the value of σ^2 in (1) is perfectly known and, without loss of generality, that $x = 1$. Since we are considering only received powers, the signal at the output of the i-th antenna detector is given by

$$w_i = \|y_i\|^2 = \|h_i + n_i\|^2, \tag{2}$$

where y_i, h_i and n_i for $i = 1, \ldots, M$ are the elements of $\mathbf{y}$, $\mathbf{h}$ and $\mathbf{n}$, respectively. When conditioned on h_i, w_i follows a Rician distribution (in power) [37], and due to the statistical independence of the noise samples, the joint conditioned probability density function (PDF) of the vector $\mathbf{w} = (w_1, \ldots, w_M)^T$ is given by the product of the individual PDFs.

Consider also that, during the training phase, N_t samples of $\mathbf{w}_0$, namely $\mathbf{w}_{0,j}$ for $j = 1, \ldots, N_t$, are obtained from a correct (trained) point $\mathbf{p}_0$. The samples $\mathbf{w}_{0,j}$ are then jointly Rician distributed with vector parameter[1] $\mathbf{g}_0 = (\|h_{0,1}\|^2, \ldots, \|h_{0,M}\|^2)^T$. Then, from $\mathbf{w}_{0,j}$, the system obtains an estimation $\widehat{\mathbf{g}}_0 = (\widehat{g}_{0,1}, \ldots, \widehat{g}_{0,M})^T$ of $\mathbf{g}_0$.

Once trained (evaluation phase), the LIS receives another set of samples $\mathbf{w}_k$ for $k = 1, \ldots, N_v$ from an arbitrary point $\mathbf{p}$. Therefore, the objective is, based on $\mathbf{w}_{0,j}$ and $\mathbf{w}_k$, to determine whether $\mathbf{p} = \mathbf{p}_0$ or not. To that end, we formulate a binary hypothesis test as

$$\begin{aligned} H_0 &: \widehat{\mathbf{g}} = \widehat{\mathbf{g}}_0 \\ H_1 &: \widehat{\mathbf{g}} \neq \widehat{\mathbf{g}}_0, \end{aligned} \tag{3}$$

where $\widehat{\mathbf{g}} = (\widehat{g}_i, \ldots, \widehat{g}_M)^T$ is the channel vector estimated from $\mathbf{w}_k$. The test is hence formulated based on the GLRT, but replacing the knowledge of the null hypothesis by its estimated counterpart, i.e.,

$$\begin{aligned} \log\Lambda = &\sum_{k=1}^{N_v}\sum_{i=1}^{M} \log I_0\left(\frac{2\widehat{g}_{0,i}\sqrt{w_{i,k}}}{\sigma^2}\right) + \sum_{k=1}^{N_v}\sum_{i=1}^{M} \frac{\widehat{g}_i - \widehat{g}_{0,i}}{\sigma^2} \\ &- \sum_{k=1}^{N_v}\sum_{i=1}^{M} \log I_0\left(\frac{2\widehat{g}_i\sqrt{w_{i,k}}}{\sigma^2}\right) \underset{H_1}{\overset{H_0}{\gtrless}} \eta, \end{aligned} \tag{4}$$

where $w_{i,k}$ denote the i-th entry of $\mathbf{w}_k$. Replacing the true value of $\mathbf{g}_0$ by its estimation introduces a non-negligible error in the test that has to be considered in the threshold design, as we will see in the following subsections.

B. ESTIMATOR FOR G

In conventional likelihood ratio tests, the estimation of the involved parameters is carried out through maximum likelihood. However, since in our problem the distribution of the received power signal $\mathbf{w}_k\ \forall\ k$ is a multivariate Rician, the maximum likelihood estimation implies solving a system of M non-linear equations [38]. This may lead to a considerable computational effort taking into account the large number of antennas (M) that characterizes the LIS. To circumvent this issue, we proposed a suboptimal — albeit unbiased — estimator based on moments.

Since $\mathbb{E}[\mathbf{n}\mathbf{n}^H] = \mathbf{I}_M$, the estimation of the channel at each antenna element can be solved separately. Then, we can estimate g_i in both the training and evaluation phases as

$$\widehat{g}_{0,i} = \frac{1}{N_t}\sum_{j=1}^{N_t} w_{0,i,j} - \sigma^2, \quad \widehat{g}_i = \frac{1}{N_v}\sum_{k=1}^{N_v} w_{i,k} - \sigma^2, \tag{5}$$

where $w_{0,i,j}$ are the elements of $\mathbf{w}_{0,j}$. It is easily proved that the estimators in (5) are unbiased with normally distributed error for relatively large number of samples.

C. THRESHOLD DESIGN

Although the asymptotic properties of $\log\Lambda$ have been well studied in the literature (see, e.g., [39]), these general results are not valid in our case because *i)* we are replacing the true value of $\mathbf{g}_0$ by its estimation, and *ii)* we are using moment-based estimators instead of the optimal one. A more general result, which is the starting point of our derivation, is that the limiting distribution of $-2\log\Lambda$, under the null hypothesis, is given by [40, eq. (4.3)]

$$-2\log\Lambda|_{H_0} \xrightarrow{p} (\widehat{\mathbf{g}}_0 - \widehat{\mathbf{g}})^T N_v \mathbf{J}(\widehat{\mathbf{g}}_0 - \widehat{\mathbf{g}}), \tag{6}$$

where we have replaced $\mathbf{g}_0$ by $\widehat{\mathbf{g}}_0$. In (6), $\xrightarrow{p}$ stands for convergence in probability and $\mathbf{J} \in \mathbb{R}^{M\times M}$ is the Fisher information matrix of $\mathbf{w}_k$ with respect to $\mathbf{g}_0$ [41]. In our case, $\mathbf{J}$ is a diagonal matrix whose i-th element is given by

$$J_i(g_i) = \frac{e^{-g_i/\sigma^2}}{\sigma^6 g_i}\int_0^\infty w_i e^{-w_i/\sigma^2}\frac{I_1^2\left(\frac{2}{\sigma^2}g_i\sqrt{w_i}\right)}{I_0\left(\frac{2}{\sigma^2}g_i\sqrt{w_i}\right)}dw_i - \frac{1}{\sigma^4}. \tag{7}$$

Eq. (6) can be rewritten as

$$-2\log\Lambda|_{H_0} \xrightarrow{p} (\boldsymbol{\epsilon}_0 - \boldsymbol{\epsilon})^T N_v \mathbf{J}(\boldsymbol{\epsilon}_0 - \boldsymbol{\epsilon}), \tag{8}$$

where $\boldsymbol{\epsilon}_0 = (\epsilon_{0,1}, \ldots, \epsilon_{0,M})^T$ and $\boldsymbol{\epsilon} = (\epsilon_1, \ldots, \epsilon_M)^T$ are the error vectors of estimators in (5). Note that both error vectors are Gaussian distributed, but they vary at very different time scales. The true channel $\mathbf{g}_0$ is estimated during the training phase, and thus the error $\boldsymbol{\epsilon}_0$, albeit random, remains constant during the whole evaluation phase until the system is retrained. In turn, each time the system evaluates a point, $\boldsymbol{\epsilon}$ takes a different (random) realization. With that in mind, we propose choosing η based on a worst case design, i.e., we consider an estimation error $\boldsymbol{\epsilon}_0$ that overestimates the true error at $1 - \alpha_0$ percent of the time. That is,

$$\epsilon'_{0,i} = F^{-1}_{\epsilon_{0,i}}(1 - \alpha_0/2), \tag{9}$$

where $F_{\epsilon_{0,i}}$ stands for the cumulative distribution function (CDF) of a Gaussian variable with zero mean and variance $\frac{\sigma^2}{N_t}(\sigma^2 + 2\widehat{g}_{0,i})$. Note that we have replaced the true channel value by its estimation in the calculation of the aforementioned percentile.

Finally, conditioned on $\epsilon'_{0,i}$, the distribution of the test for large number of samples is given by

$$-2\log\Lambda|_{H_0,\epsilon'_0} \xrightarrow{p} D = \sum_{i=1}^{M} N_v J_i(g_i)\left(\epsilon_i - \epsilon'_{0,i}\right)^2, \tag{10}$$

1. Note that, due to the circular symmetry of the noise, the distribution of $\mathbf{w}$ does not depend on the complex channel h_i but on its squared modulus $g_i = \|h_i\|^2$.

Algorithm 1: Statistical Test for Sensing

Training phase:
for *each* $\mathbf{p} \in \mathbf{P}_c$ **do**
  I. Estimate $\widehat{\mathbf{g}}_0$ using (5)
  II. Compute $\epsilon'_{0,i}$ for $i = 1, \ldots, M$ from (9) for a confidence value α_0
  III. Compute $J(\widehat{g}_{0,i})$ for $i = 1, \ldots, M$ from (7)
  IV. Compute -2η using (11) for a confidence value α
end
Evaluation phase: Received $\mathbf{w}_j$ for $j = 1, \ldots, N_v$, do
for *each* $\mathbf{p} \in \mathbf{P}_c$ **do**
  I. Estimate $\widehat{\mathbf{g}}$ using (5)
  II. Compute $-2\log\Lambda$ using (4)
  III. Reject H_0 if $-2\log\Lambda > -2\eta$
end
The point does not belong to $\mathbf{P}_c$ if H_0 is rejected $\forall$ $\mathbf{p}$

which corresponds to a non-central Gaussian quadratic form. Therefore, given a predefined false alarm probability α, the test finally reads as

$$-2\log\Lambda \underset{H_0}{\overset{H_1}{\gtrless}} -2\eta = F_D^{-1}(1-\alpha|\widehat{\mathbf{g}}_0), \tag{11}$$

where F_D is the CDF of D, which can be obtained by Monte-Carlo simulations or by using some of the approximations given in the literature for Gaussian quadratic forms (see, e.g., [42]–[44]). Note that, in (11), we have again replaced the true channel values by their estimations. In our proofs, this seems to have a negligible impact on the threshold distribution unless the number of samples is very low (in which case the asymptotic analysis here presented does no longer hold). A summary of the proposed statistical test is provided in Algorithm 1, where the here presented pointwise comparison is performed along the whole route $\mathbf{P}_c$.

V. MACHINE LEARNING FOR RADIO IMAGE-BASED LIS SENSING

In the previous section, we have presented a statistical method to sense the environment based on the received power signal at the different antenna elements of the LIS, and hence detecting route deviations from a predefined correct trajectory. This approach exploits the large number of antennas in the LIS in the same way as in massive MIMO systems. However, the high spatial density of antennas and the large array aperture of LIS can be exploited in an alternative way. The basis of this novel technique is using the power of the received signal across the surface as a radio image of the environment, as stated in Section II.

A. MODEL DESCRIPTION

We introduce a ML model to perform the anomalous route classification task based on the radio-based images obtained at the LIS. The main advantage of this proposal, as we will see, is that it is independent on the data distribution, and no assumptions are needed to its implementation. This is in contrast with Section IV where we considered the noise is Gaussian distributed with noise variance known for the sake of simplicity in the analytical derivations. In reality, these assumptions may not hold.

In our considered problem, the training data is obtained by sampling the received power at certain temporal instants while the target device is moving along the correct route. In order to reduce both training time and scanning periods, which may be heavy tasks for large trajectories, we resort on transfer learning [45]. Because of this matter, the risk of overfitting due to our constraint of short scanning periods is quite significant, being transfer learning also a proper way to tackle it. This allows using a small dataset and therefore improving the flexibility of the system in real world deployments. In our case, we choose the VGG19 architecture [46]. Due to our specific use case, and the training data constraints, we propose the use of an unsupervised ML algorithm named as LOF which identifies the outliers presents in a dataset (i.e., the anomalous positions of the target robot) [47].

The proposed model is detailed in Fig. 2, where, in order to perform the feature extraction, we remove the last fully connected layer (FC) that performs the classification for the purpose of VGG19 and modify it for our specific classification task (anomaly/not anomaly in robot's route).

B. LOCAL OUTLIER FACTOR

LOF is an unsupervised ML algorithm that relies on the density of data points in the distribution as a key factor to detect outliers (i.e., anomalous events). In the context of anomaly detection, LOF was proposed in [47] as a solution to find anomalous data points by measuring the local deviation of a given point with respect to its neighbors.

LOF is based on the concept of local density, where the region that compounds a *locality* is determined by its K nearest neighbors. By comparing the local density of a point to the local densities of its neighbors, one can identify regions of similar density, and points that have a substantially lower density than their neighbors (the latter are considered to be outliers). This approach can be naturally applied to the anomalous trajectory deviation detection as deviated points that are really close to the correct trajectory could be really close in distance, but they would have a lower density compared with the points that actually belong to the correct one, being accurately detected as deviations. Hence, the points belonging to the correct route are used to learn the correct clusters. The strength of the LOF algorithm is that it takes both local and global properties of datasets into consideration, i.e., it can perform well even in datasets where anomalous samples have different underlying densities because the comparison is carried out with respect to the surrounding neighborhood. For the reader's convenience, a brief description of the LOF theory is provided in the following.[2]

2. For a more detailed description, the reader is gently referred to [47].

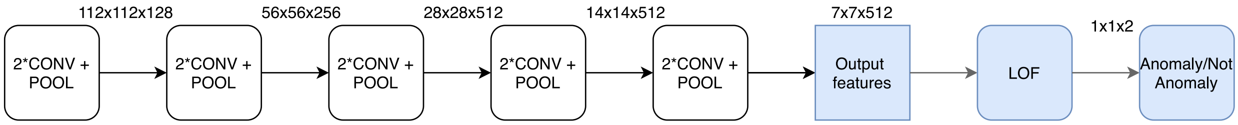


FIGURE 2. Proposed model. White and blue blocks refer to VGG19 re-used original architecture and to the additional blocks for our task, respectively.

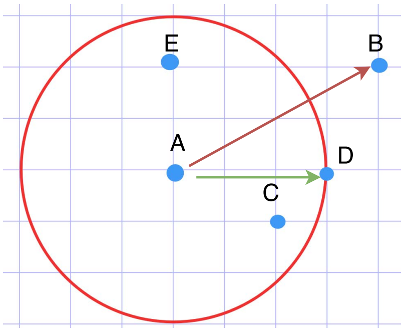


FIGURE 3. Illustration of Reachability Distance with $K = 3$. Manhattan distance used for illustration purposes.

The algorithm is based on two metrics, namely the K-distance of a point A, denoted by $D_K(A)$, and its K-neighbors, which is the set $N_K(A)$ composed by those points that are in or on the circle of radius D_K with respect to the point A. Note that K is a hyperparameter to be chosen and fixed for computing the clusters. Also note that this implies $|N_K(A)| \geq K$, where $|N_K(A)|$ is the number of points in $N_K(A)$. With these two quantities, the reachability distance between two arbitrary points A and B is defined as

$$RD_K(A,B) = \max\{D_K(A), d(A,B)\}, \tag{12}$$

where $d(A,B)$ is the distance between points A and B. Figure 3 illustrates the RD_K concept. This means that if a point B lies within the K-neighbors of A, the reachability distance will be $D_K(A) = 3$ (the radius of the circle containing points C, D and E), else the reachability distance will be the distance between A and B. In the example, $RD_K(A,B) = 6$.

Note that the distance measure is problem-specific, being in our case the Euclidean distance between the different features extracted by the VGG19 network. From (12), the local reachability density (LRD) of a point A is defined as the inverse of the average reachability distance of A from its neighbors, i.e.,

$$LRD_K(A) = \left(\sum_{B \in N_K(A)} \frac{RD_K(A,B)}{|N_K(A)|} \right)^{-1}. \tag{13}$$

According to the LRD, if neighbors are far from the point (i.e., more the average reachability distance), less density of points are presented around a particular point. Note this would be the distance at which the point A can be reached from its neighbors, meaning this measures how far a point is from the nearest cluster of points, acquiring low values of LRD when the closest cluster is far from the point. This finally give rise to the concept of LOF, which is given by

$$LOF_K(A) = \frac{\sum_{B \in N_K(A)} LRD_K(B)}{LRD_K(A)|N_K(A)|}. \tag{14}$$

Observe that, if a point is an inlier, the ratio is approximately equal to one because the density of a point and its neighbors are practically equal. In turn, if the point is an outlier, its LRD would be less than the average LRD of neighbors, and hence the LOF would take large values. In our specific problem, we propose using the LOF values to determine whether a point belong to the correct trajectory or from any other point due to a robot deviation.

C. DATASET FORMAT

With the algorithm and the model introduced, the remaining component to fully characterize the proposed ML solution is the dataset. In our considered problem, the dataset is obtained by sampling the received signal power at each element of the LIS while the robot moves along the trajectories. Formally, we can define the possible trajectories (including those composed by both correct and anomalous points) as the set of points in the space $\mathbf{P}_t \in \mathbb{R}^{N_p \times 3}$ being N_p the total number of points in the route. Let us assume the system is able to obtain N_s samples at each channel coherence interval $\forall\, \mathbf{p}_j \in \mathbf{P}_t$, being $\mathbf{p}_j$ for $j = 1, \dots, N_p$ an arbitrary point of the route. Hence, the dataset consists of $T = N_p \times N_s$ samples (monocromatic radio image snapshots of received power). Each sample is a gray-scale image which is obtained by mapping the received power into the range of [0, 255]. To that end, we apply min-max feature scaling, in which the value of each pixel $m_{i,j}$ for $i = 1, \dots, M$ and $j = 1, \dots, N_p$ is obtained as

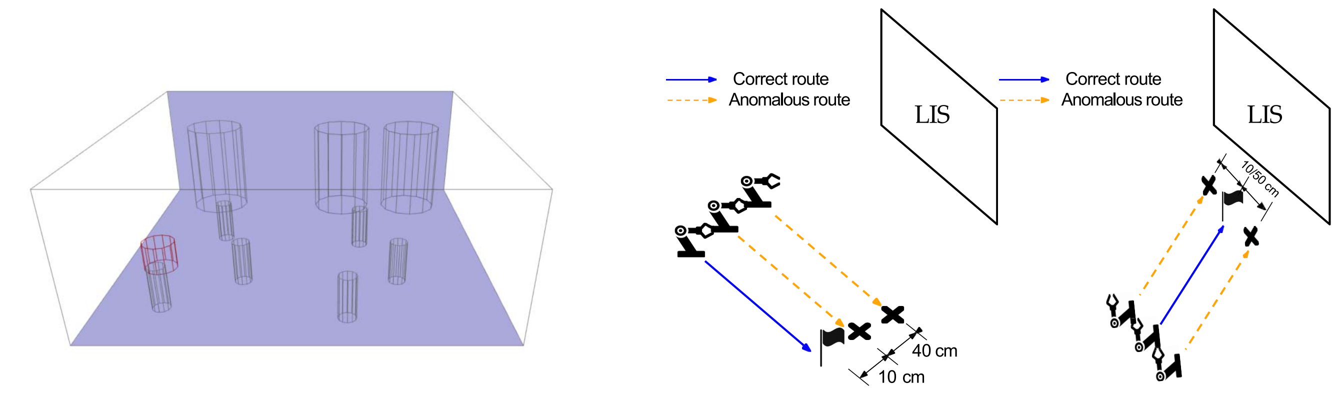
$$m_{i,j} = \left\lceil m_{\text{MIN}} + \frac{(w_{i,j} - w_{\text{MIN},j})(m_{\text{MAX}} - m_{\text{MIN}})}{w_{\text{MAX},j} - w_{\text{MIN},j}} \right\rfloor, \tag{15}$$

where $w_{i,j}$ are the elements of $\mathbf{w}_j$, i.e., $w_{i,j} = \|h_{i,j} + n_{i,j}\|^2$, $m_{\text{MAX}} = 255$ and $m_{\text{MIN}} = 0$, and

$$w_{\text{MAX},j} = \max_{\{i=1,\dots,M\}} \mathbf{w}_{i,j}, \quad w_{\text{MIN},j} = \min_{\{i=1,\dots,M\}} \mathbf{w}_{i,j} \tag{16}$$

are the maximum and minimum received power value from a point $\mathbf{p}_j$ along the surface.

The input structure supported by VGG19 is a RGB image of $n_c = 3$ channels. Due to our monocromatic measurements, our original gray-scale input structure is a one-channel image. To solve this problem, we expand the values by copying them into a $n_c = 3$ channels input structure.



(a) Use case scenario. (b) Parallel deviation. (c) Normal deviation.

FIGURE 4. Simulated scenario.

Once the feature extraction is performed, the output is $n_c = 512$ channels of size $n_w = 7$ and $n_h = 7$ pixels. Since LOF works with vectors, the data is reshaped into an input feature vector formed by $7 \times 7 \times 512 = 25088$ dimensions, meaning our dataset is $\{x^{(i)}\}_{i=1}^{T}$, where $x^{(i)}$ is the i-th n-dimensional training input features vector (being $n = 25088$) and $x_j^{(i)}$ is the value of the j-th feature.

VI. MODEL VALIDATION

In order to validate both sensing solutions, namely the statistical hypothesis testing and the radio-based image sensing algorithm, we carried out an extensive set of simulations to analyze the performance of the systems in a simple, yet interesting, industrial scenario. To properly obtain the received power values, we use a ray tracing software, therefore capturing the effects of the multipath propagation in a reliable way. Specifically, we consider ALTAIR FEKO WINPROP [48].

A. SIMULATED SCENARIO

The baseline set-up is described in Fig. 4(a), a small size industrial scenario of size 484 m^2. We address the detection of the deviation of the target robot (highlighted in red color) in 2 cases: *i)* it follows a fixed route parallel [Fig. 4(b)], and *ii)* the correct route is normal to the bottom wall, in which the LIS is deployed [Fig. 4(c)]. To evaluate the performance in the detection of anomalies, we consider that the robot may deviate from the correct route at any point, and we test the ability of both systems to detect potential deviations at a distance of, at least, $\Delta d = 50$ cm and $\Delta d = 10$ cm. These two distances correspond to the cases $\Delta d \gg \lambda$ and $\Delta d \approx \lambda$, respectively, denoting λ the wavelength.

For the aforementioned cases, we simulate in the ray tracing software N_p points, which correspond to different positions of the robot in both the correct and anomalous routes. Then, N_s radio image snapshots of the measurements are taken at every $\mathbf{p}_j$, $j = 1, \ldots, N_p$. The most relevant parameters used for simulation are summarized in Table 1.

In our simulations, we set $N_p = 367$ and $N_s = 10$, thus the dataset is composed of $T = N_p \times N_s = 3670$ radio propagation snapshots containing images of both anomalous and non-anomalous situations, as described in Section V-C. Out of $N_p = 367$, $N_c = 185$ are the snapshots corresponding to the correct route, meaning we have $T_c = N_c \times N_s = 1850$ correct data samples, while the remaining are anomalous points. To train the algorithm with the correct points, we split the correct dataset into a 80% training set 10% validation set and the remaining 10% for the test set. During the training phase, the optimum value of $K = 3$ (the LOF parameter) is obtained by maximizing the accuracy score in the correct validation set. The training procedure was performed in an Intel Xeon machine with 32 CPUs taking around 15 seconds in the offline scanning period.

TABLE 1. Parameters.

Frequency (GHz)	Tx Power (dBm)	Nray paths	Antenna type	Antenna Spacing (cm)	Propagation model
3.5	20	10	Omni	$\frac{\lambda}{2}$	Free Space

B. RECEIVED POWER AND NOISE MODELING

The complex electric field arriving at the i-th antenna element at sample time t, $\widetilde{E}_i(t)$, can be regarded as the superposition of each path, i.e.,[3]

$$\widetilde{E}_i(t) = \sum_{n=1}^{N_r} \widetilde{E}_{i,n}(t) = \sum_{n=1}^{N_r} E_{i,n}(t) e^{j\phi_{i,n}(t)}, \tag{17}$$

where N_r is the number of paths and $\widetilde{E}_{i,n}(t)$ is the complex electric field at i-th antenna from n-th path, with amplitude $E_{i,n}(t)$ and phase $\phi_{i,n}(t)$. From (17), and assuming isotropic antennas, the complex signal at the output of the i-th element is therefore given by

$$y_i(t) = \sqrt{\frac{\lambda^2 Z_i}{4\pi Z_0}} \widetilde{E}_i(t) + n_i(t), \tag{18}$$

3. Note that the electric field also depends on the point $\mathbf{p}_j$. However, for the sake of clarity, we drop the subindex j throughout the following subsections.

with λ the wavelength, $Z_0 = 120\pi$ the free space impedance, Z_i the antenna impedance, and $n_i(t)$ is complex Gaussian noise with zero mean and variance σ^2. Note that (18) is exactly the same model than (1); the only difference is that we are explicitly denoting the dependence on the sampling instant t. For simplicity, we consider $Z_i = 1\,\forall i$. Thus, the power $w_i(t) = \|y_i(t)\|^2$ is used at each temporal instant t both to perform the hypothesis testing in (4) and to generate the radio image, as pointed out before. Finally, in order to test the system performance under distinct noise conditions, the average signal-to-noise ratio (SNR) over the whole route, $\overline{\gamma}$, is defined as[4]

$$\overline{\gamma} \triangleq \frac{\lambda^2}{4\pi Z_0 M T \sigma^2} \sum_{t=1}^{T} \sum_{i=1}^{M} \left|\widetilde{E}_i(t)\right|^2, \tag{19}$$

where M denotes the number of antenna elements in the LIS.

C. NOISE AVERAGING STRATEGY

The statistical solution presented in Section IV has been derived taking into account the presence of noise, and consequently it has implicit mechanisms to reduce its impact in the performance. However, the presence of noise may be more critical in the radio image sensing, since it impacts considerably the image classification performance [49].

Referring to (2) and (18), since we are considering only received powers, the signal at the output of the i-th antenna detector is given by

$$w_i = \left\| \sqrt{\frac{\lambda^2 Z_i}{4\pi Z_0}} \widetilde{E}_i + n_i \right\|^2, \tag{20}$$

where we have dropped the dependence on t. Also, let us assume the system is able to obtain S extra samples at each channel coherence interval $\forall\, \mathbf{p}_j \in \mathbf{P}_t$. That is, at each point $\mathbf{p}_j$, the system is able to get $N_s' = N_s \times S$ samples. Since the algorithm only expects N_s samples from each point, we can use the extra samples to reduce the noise variance at each pixel. To that end, the value of each pixel $m_{i,j}$ is not computed using directly $w_{i,j}$ as in (15) but instead

$$w_{i,j}' = \frac{1}{S} \sum_{s=1}^{S} w_{i,j,s}, \tag{21}$$

where $w_{i,j,s}$ denote the received signal power at each extra sample $s = 1, \ldots, S$. Note that, if $S \to \infty$, then

$$w_{i,j}'\Big|_{S\to\infty} = \mathbb{E}\left[w_{i,j} | h_{i,j}\right] = \left\|h_{i,j}\right\|^2 + \sigma^2, \tag{22}$$

meaning that the noise variance at the resulting image has vanished, i.e., the received power at each antenna (conditioned on the channel) is no longer a random variable. Observe that the image preserves the pattern with the only addition of an additive constant factor σ^2. This effect is only possible if the system would be able to obtain a very large number S of samples within each channel coherence interval.

4. This is equivalent to averaging over all the points $\mathbf{p}_j$ of the trajectory $\mathbf{P}$.

D. STACKED DENOISING AUTOENCODER FOR IMAGE SUPER-RESOLUTION

An autoencoder is a type of neural network which tries to learn a representation (denoted as encoding) from the data given as input, usually for dimensionality reduction purposes. Along with the encoding side, a reconstructing side is learnt, where the autoencoder tries to reconstruct from the reduced encoded data a representation as close as possible to its original input. There are several variations of the basic model in order to enforce the learned representation to fulfill some properties [50]. Among all of them, we are interested in the DAEs [51].

The goal of the denoising autoencoder is to reconstruct a "clean" target version from a corrupted input. In our context, let us assume that we can obtain a target image $\mathbf{t} \in \mathbb{N} \cap [0, 255]^N$ and a corrupted input $\mathbf{c} \in \mathbb{N} \cap [0, 255]^M$ result of the received power mapping explained in (15). Also, consider $N \gg M$ and that $\mathbf{t}$ was obtained from a less noisy environment, i.e., the average SNR $\overline{\gamma}_t$ is greater than that of $\mathbf{c}$, denoted by $\overline{\gamma}_t \geq \overline{\gamma}_c$. Then, we can perform a resizing of both images such as $R : [0, 255]^d \to \mathbb{R} \cap [0, 1]^R$, meaning we resize both images towards the same dimension R and we normalize the values dividing by 255, being $\mathbf{t_r} = R(\mathbf{t})$ and $\mathbf{c_r} = R(\mathbf{c})$ the target and corrupted input used to train our DAE. Note that, although the two images ($\mathbf{t_r}$ and $\mathbf{c_r}$) are identical in dimension (R pixels) after the resizing procedure, the resolution of the target one is higher because it is obtained in a more favorable scenario (larger SNR) and from an initially higher number of pixels N.

To illustrate the approach, a one hidden layer explanation is made for simplicity. The denoising autoencoder can be split into two components: the encoder and the decoder, which can be defined as transformations Φ and Ψ such that $\Phi : \mathbb{R} \cap [0, 1]^R \to \mathcal{F}$, $\Psi : \mathcal{F} \to \mathbb{R} \cap [0, 1]^R$. Then, the encoder stage of the DAE takes the input $\mathbf{c_r}$ and maps it into $\mathbf{e} \in \mathbb{R}^l = \mathcal{F}$ as

$$\mathbf{e} = \rho(\mathbf{W}\mathbf{c_r} + \mathbf{b}), \tag{23}$$

being l the dimension of the compressed representation of the data, known as the latent space, ρ the element-wise activation function, $\mathbf{W}$ the weight matrix and $\mathbf{b}$ the bias vector. These weights and biases are randomly initialized and updated iteratively through backpropagation. After this process, the decoder stage maps $\mathbf{e}$ to the reconstruction $\widetilde{\mathbf{c_r}}$ of the same shape as $\mathbf{c_r}$

$$\widetilde{\mathbf{c_r}} = \rho\left(\mathbf{W}'\mathbf{e} + \mathbf{b}'\right), \tag{24}$$

being ρ the activation function, and $\mathbf{W}'$ and $\mathbf{b}'$ the parameters used in the decoder part of the network. In our specific case, the reconstruction error, also known as loss, is given by the mean squared error (MSE) of the pixel values of the target image and the reconstructed image (R pixels), that is[5]

$$\mathcal{L}(\mathbf{t_r}, \widetilde{\mathbf{c_r}}) = \frac{\sum \|\mathbf{t_r} - \widetilde{\mathbf{c_r}}\|^2}{R}$$

5. Note that the summation is made along all the pixel values. However, for the sake of clarity, we drop the subindex in this expression.

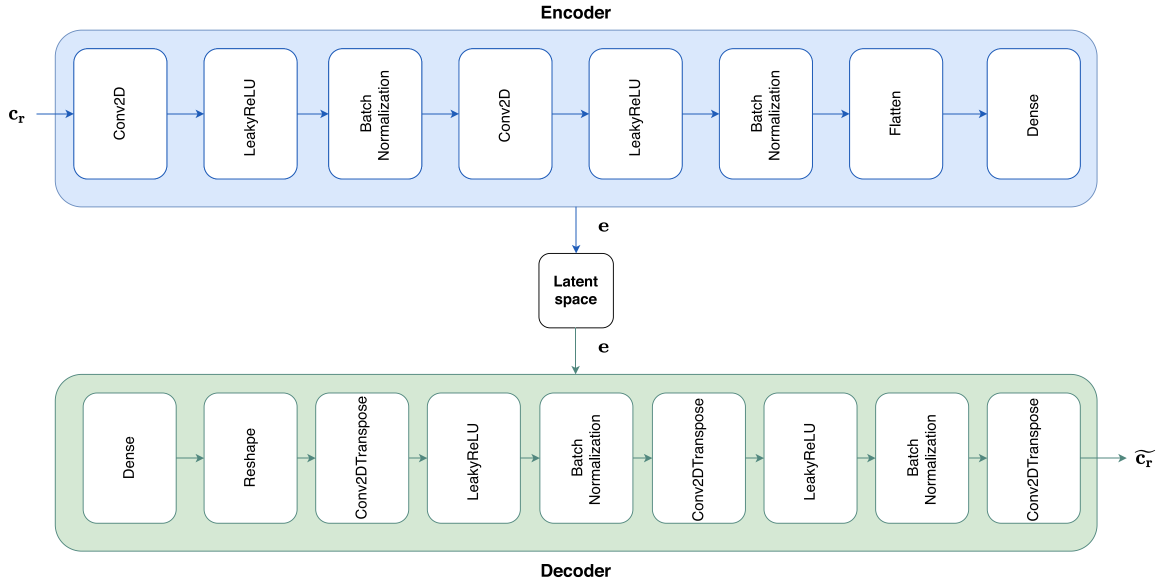


FIGURE 5. Denoising autoencoder architecture.

$$= \frac{\sum \left\| \mathbf{t_r} - \sigma\left(\mathbf{W}'\sigma\left(\mathbf{W}\mathbf{c_r} + \mathbf{b}\right) + \mathbf{b}'\right)\right\|^2}{R}. \quad (25)$$

For the sake of reproducibility, a detailed summary of the proposed architecture is provided in Figure 5. We have used the Keras [52] library, so the description of the layers corresponds to its notation. The ADAM optimizer with a learning rate $\alpha = 0.001$, exponential decay for the 1st and 2nd moment estimates $\beta_1 = 0.9$ and $\beta_2 = 0.999$ and $\epsilon = 10^{-7}$ have been used for updating the gradient, minimizing the MSE loss function. For the encoder part, Conv2D layers with filter size 64 and 32 have been used, kernel sizes of 3x3 and stride $= 2$. The activation function LeakyReLU has been used with a slope coefficient $\beta = 0.2$. Then, a Batch Normalization layer has been used to maintain the mean output close to 0 and the output standard deviation close to 1. The Flatten layer is used to reshape the output into a vector to feed the Dense layer with a number of neurons $l = 16$ which corresponds to the dimension of the latent space. The dimension l was determined by analyzing the learning curves in the training procedure.

In the decoder part, a Dense layer is used again to recover the previous size of the feature vector while the reshaping recovers the initial 2D input structure. Then, Conv2DTranspose layers have been used to perform the reconstruction of the input structure, having an identical configuration than in the encoder side but changing the order of the filters (32 and 64). The LeakyReLU activations and the Batch Normalization are identical. The last layer is composed by 1 filter, kernel size of 3x3 and stride = 1. This last layer is for recovering completely the size as the input structure. Furthermore, the DAE network trains itself to augment the resolution of the input image, because it will remove artifacts resulting from a lower resolution, by learning from a high resolution target. Then, this reconstructed image will be used for our anomaly detection algorithm. This is advantageous for our problem, leading to a strategy for improving the performance of the system.

E. PERFORMANCE METRICS

To evaluate the prediction effectiveness of our proposed method, we resort on common performance metrics that are widely used in the related literature. Concretely, we are focusing on the F1-Score which is a metric based on the Precision and Recall metrics. First, we need to describe what we consider as a positive or negative event. In our problem, TP and FP stand for True and False Positive (anomalous event) while TN and FN for True and False Negative (correct event). In this way, the applied metrics are defined as follows:

- Precision positive (PP) and negative (PN) as the proportion of correct predictions of a given class

$$\text{PP} = \frac{\text{TP}}{\text{TP+FP}}, \quad \text{PN} = \frac{\text{TN}}{\text{TN+FN}}. \quad (26)$$

- Recall positive (RP) and negative (RN) as the proportion of actual occurrences of a given class which has been correctly predicted.

$$\text{RP} = \frac{\text{TP}}{\text{TP+FN}}, \quad \text{RN} = \frac{\text{TN}}{\text{TN+FP}}. \quad (27)$$

- Positive F1-Score (PF_1) and Negative F1-Score (NF_1) as the harmonic mean of precision and recall:

$$PF_1 = 2 \cdot \frac{\text{PP} \cdot \text{RP}}{\text{PP + RP}}, \quad NF_1 = 2 \cdot \frac{\text{PN} \cdot \text{RN}}{\text{PN + RN}}. \quad (28)$$

Note that although the training procedure is fully unsupervised, for our specific evaluation we know the labels of the data samples, meaning we can calculate these metrics, well-known in the supervised learning literature.

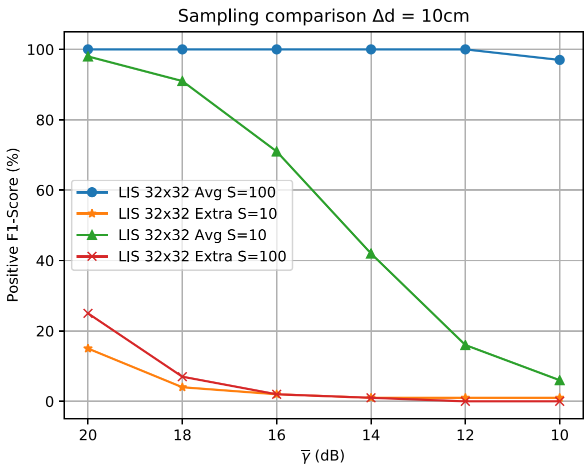


FIGURE 6. ***PF*$_1$ score for radio image sensing with $M = 32 \times 32$ antennas, inter-element distance of $\Delta s = \lambda/2$, correct route parallel to the LIS, anomalous points placed at $\Delta d = 10$ cm, and different numbers of samples.**

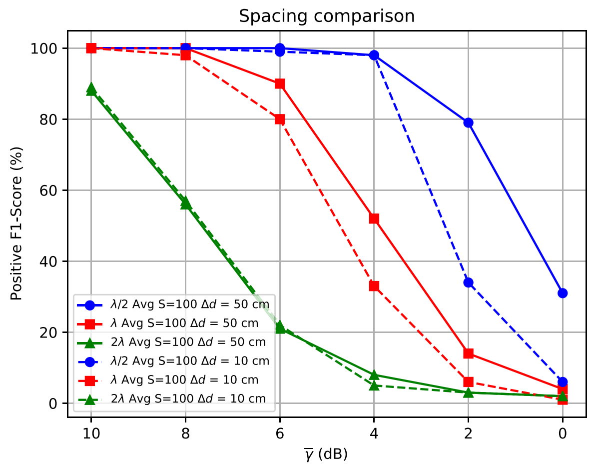


FIGURE 7. ***PF*$_1$ score for radio image sensing with fixed LIS aperture of 5.44 × 5.44 m, correct route parallel to the LIS, anomalous points placed at $\Delta d = 10$ and $\Delta d = 50$ cm, $S = 100$ samples, and different number of antennas and spacings.**

VII. NUMERICAL RESULTS AND DISCUSSION

We here present some numerical results in order to analyze the performance of both proposals (statistical test and radio image sensing) in our evaluation setup described in Section VI. Generally, in the considered industrial setup, it would be more desirable to avoid undetected anomalies (which may indicate some error in the robot or some external issue in the predefined trajectory) than obtaining a false positive. Hence, all the figures in this section shows the algorithm performance in terms of the PF_1 metric.

Also, we mainly focus our results on the radio image sensing algorithm since it is the proposal with a larger number of tunable parameters, whilst the statistical hypothesis testing is used as a benchmark of the ML based solution.

A. IMPACT OF SAMPLING AND NOISE AVERAGING

First, we evaluate the impact of both available number of samples and the noise averaging technique in the radio image sensing algorithm. To that end, we consider a LIS compounded by $M = 32 \times 32$ antennas and a spacing $\Delta s = \lambda/2$ for a $\Delta d = 10$ cm parallel deviation.

We evaluate two approaches: *i)* using the S extra samples directly as input to the algorithm, being $Tc = N_c \times N_s \times S$, and *ii)* using the S extra samples for averaging. For this particular case, $N_s' \in \{1000, 100\}$. Then $\forall$ $\mathbf{p}_j$ we use $S = \frac{N_s'}{N_s}$ samples for obtaining N_s S-averaged samples for training the algorithm, being still $Tc = N_c \times N_s = 1850$. Note that the number of samples N_s' would depend on the sampling frequency and the second order characterization of the channel, i.e., the channel coherence time and its autocorrelation function.

Figure 6 shows the performance of the system when using S extra samples and S averaged ones respectively. As highlighted previously, noise contribution is critical in image classification performance, leading to not achieving a valuable improvement when augmenting the number of samples presented to the algorithm. However, when performing the averaging, results are significantly improved due to the noise variance reduction. As expected, when noise level is higher, more samples are needed to preserve the pattern by averaging, being $N_s' = 1000$ the one which yields a better performance. For the following discussions, this sampling strategy will be used, meaning we are using $S = 100$ extra samples.

B. IMPACT OF ANTENNA SPACING

The next step is evaluating the impact of inter-antenna distance in the ML sensing solution. We fix the aperture to 5.44×5.44 m and $S = 100$ averaged samples. Then, we assess the performance in both $\Delta d = 50/10$ cm for the parallel deviation, and we analyze different spacings with respect to the wavelength ($\lambda/2$, λ and 2λ).

The performance results for the distinct configurations are depicted in Fig. 7. As observed, the spacing of 2λ — which is far from the concept of LIS — is presenting really inaccurate results showing that the spatial resolution is not enough. We can conclude that the quick variations along the surface provide important information to the classifier performance. Besides, this information becomes more important the lower the distance between the routes is. Specially in the range of $\overline{\gamma} \in [10, 4]$ for $\lambda/2$ where the detection is almost identical regardless of the extra precision needed to detect the deviation when the routes are closer. Furthermore, the effect of antenna densification for a given aperture is highlighted and it can be seen that the lowest spacing leads to the best results.

C. LIS APERTURE COMPARISONS

In this case, LIS with different apertures have been evaluated. The spacing is fixed to $\lambda/2$, $S = 100$ averaged samples are used while the deviation is $\Delta d = 10$ cm parallel.

Looking at Fig. 8, the aperture plays a vital role in the sensing performance. Increasing the number of antennas leads to a higher resolution image, being able to capture the

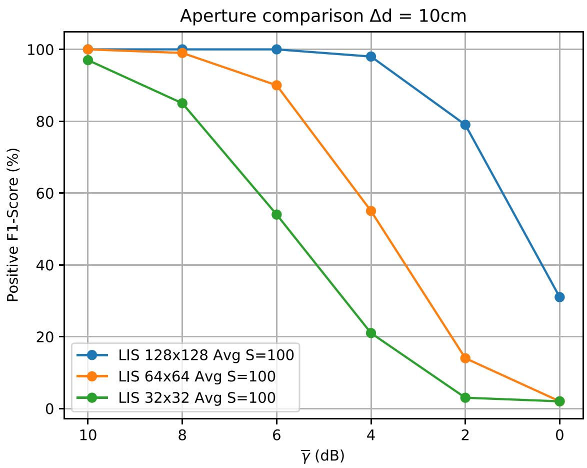


FIGURE 8. ***PF*$_1$ score for radio image sensing with variable aperture, inter-antenna distance of $\lambda/2$, correct route parallel to the LIS, anomalous points placed at $\Delta d = 10$ and $S = 100$ samples.**

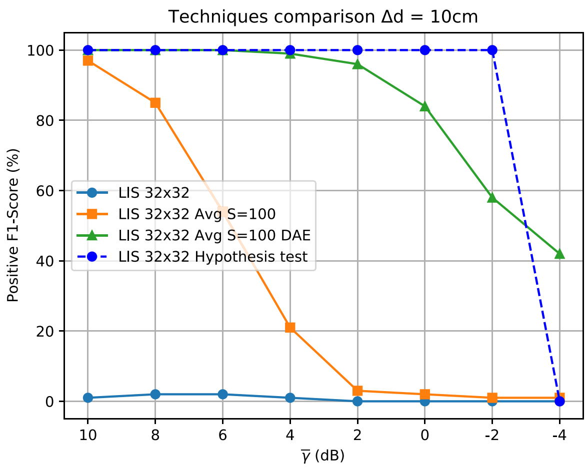


FIGURE 9. **Comparison between radio image sensing and the statistical solution in (4) for $M = 32 \times 32$ antennas, correct route parallel to the LIS, spacing of $\lambda/2$ and different processing techniques.**

large-scale events occurring in the environment more accurately. Note the usage of incoherent detectors is yielding to a good performance when the aperture is large enough. The key feature for this phenomena is the LIS pattern spatial consistency, i.e., the ability of representing the environment as a continuous measurement image.

D. DAE FOR IMAGE SUPER-RESOLUTION EVALUATION

In this case, the impact on performance by using DAE is evaluated and compared to the hypothesis test in (4). We fix the aperture to $M = 32 \times 32$, for a parallel deviation of $\Delta d = 10$ cm and an antenna spacing of $\lambda/2$.

For this evaluation, the performance is analyzed in 4 cases: *i)* no pre-processing of images performed, *ii)* $S = 100$ averaging strategy, *iii)* image resolution augmentation using DAE, and *iv)* The hypothesis test proposed in Algorithm 1. For the DAE, we assume we have access to a target reference image $\mathbf{t} \in \mathbb{N} \cap [0, 255]^N|_{N=128\times128}$ with $\overline{\gamma} = 10$ dB and our corrupted input is $\mathbf{c} \in \mathbb{N} \cap [0, 255]^M|_{M=32\times32}$ with $\overline{\gamma} \in [10, -10]$ dB. Then, we finally resize both images $R : [0, 255]^d \to \mathbb{R} \cap [0, 1]^R|_{R=224\times224}$, obtaining images of $\mathbf{t_r} \in \mathbb{N} \cap [0, 1]^{224\times224}$ and $\mathbf{c_r} \in \mathbb{N} \cap [0, 1]^{224\times224}$ pixels.

Regarding Fig. 9, one can see the raw-data (blue line) is yielding to a really poor performance. This is expected taking into account noise can interfere significantly in the local density of the clusters, leading to wrong results. Also, the noise averaging strategy is good enough when noise contribution is negligible, meaning that for improving the results in lower SNR scenarios we would need to obtain a higher S which would be unpractical. Finally, the usage of DAE for image super-resolution outperforms both methods, allowing to improve the system performance and even work in quite unfavourable SNR scenarios. In turn, the hypothesis test derived in Section IV provides in general the best performance.

However, we must take into account that the statistical test is built based on some key a priori knowledge, namely Gaussian noise with known variance. In the context of estimation, the radio image sensing solution can be seen as a non-parametric approach, which is valid for any baseline distribution and no further assumptions are required. Nevertheless, the performance of the ML solution (when DAE is employed), presents almost no difference with respect to the ad-hoc test up to 2 dB of average SNR. This is a promising result, since the application of more refined image processing techniques may lead to an increase in performance. Also, note that here we are considering a rather simple scenario, where the scatterers do not move. In a more realistic environment, with the rapid changes in the channel and the temporal dependencies due to the relative positions between users and scatterers in movement, ML-based sensing seems a promising solution to learn temporal dependencies in those scenarios where classical solutions become impractical.

E. ROUTE DEVIATIONS EVALUATION

We evaluate now the impact of the separation of deviations and different types of routes in both radio image sensing and the statistical test. To that end, we fix the aperture to $M = 32\times32$, and a antenna spacing of $\lambda/2$. We will be using all the improvements in the preprocessing of the images to leverage the performance of the ML system.

In Fig. 10, we can see the performance of the system under different deviations and SNR conditions. We can see the system works better the closer the deviation of the routes are. This is an advantage of our proposed approach, the closer the routes are, the more accurate the reconstruction of the DAE is, taking into account the corrupted image $\mathbf{c_r}$ will be more similar to the target image $\mathbf{t_r}$, allowing a better augmentation of the image resolution, so the correct clusters can be learned more accurately. In this way, the ML proposed algorithm works better in the cases a standard wireless sensing system would be more prone to failure. Also, the parallel deviations are easier to detect than the

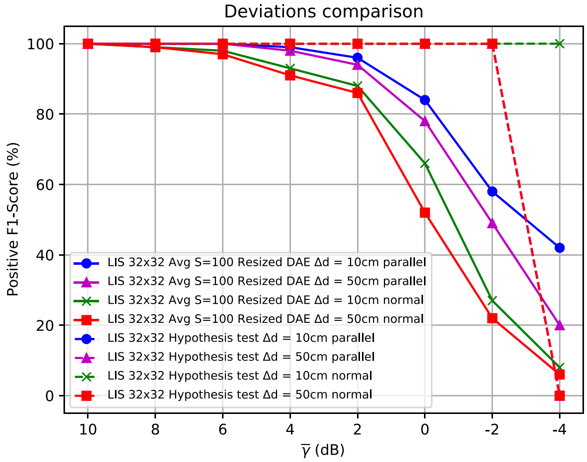


FIGURE 10. Comparison between radio image sensing and the statistical solution in (4) for $M = 32 \times 32$ antennas, different correct routes and spacing of $\lambda/2$.

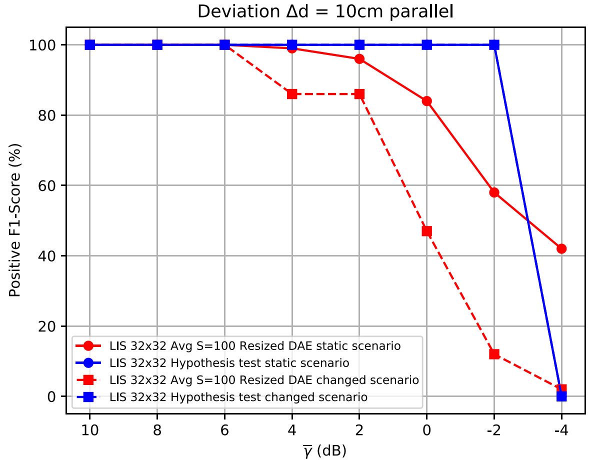


FIGURE 11. Performance of radio image sensing and the statistical solution in (4) for $M = 32 \times 32$ antennas, correct route parallel to the LIS, anomalous points placed at $\Delta d = 10$ and spacing of $\lambda/2$ in a changing environment.

normal deviations. The path loss of the points in the parallel routes remains almost identical regardless of the specific point, making it easier to detect. It is important to highlight the SNR definition presented in (19) can influence in the pattern acquisition in the normal deviation cases when points are far from the LIS, which will have a significantly lower instantaneous SNR leading to a more difficult detection.

Note that the abrupt decrease on performance for the hypothesis test is due to the fact that we are using a pointwise test to perform a detection over a whole route (collection of points), as shown in Algorithm 1. Whilst the ML algorithm performs an anomaly detection over the whole correct route, the proposed statistical test is a pointwise comparison, i.e., it checks the validity of the null-hypothesis for each point in the correct route separately. This implies that, in order to detect a point as anomalous, the test has to reject H_0 on all the correct points. Consequently, failing in a single point is equivalent to fail in all the points.

F. PERFORMANCE EVALUATION UNDER CHANGING ENVIRONMENT

Finally, we here evaluate the performance of our system when a major change in the scenario occurred, i.e., the relative positions between the scatterers and the transmitter has changed considerably and thus the pattern capture in the radio image no longer matches the original one used in the training phase. Note that, although the considered scenario for testing was assumed to be fixed, we may be interested in extrapolate the performance of the proposed solutions when dealing with environmental changes. To that end, we evaluate the anomaly detection accuracy of both the hypothesis test and the ML solution. We fix the aperture to $M = 32 \times 32$, for a parallel deviation of $\Delta d = 10$ cm and an antenna spacing of $\lambda/2$. Again, we will use all the improvements in the pre-processing of the images to leverage the performance of the ML system. Fig. 11 shows the performance of the system under a changing environment. The hypothesis test is robust to an environmental change, as its performance remains similar as the static case. With respect to the ML solution, in the range of $\overline{\gamma} \in [0, -4]$ drops significantly. However, in the proposed scenario, we assume a change in the environment is a really unlikely event, leading to a worsening in the performance in some SNR cases.

VIII. CONCLUSION

We have made the first step towards the use of LIS for sensing the propagation environment, exploring and proposing two different solutions: *i)* an statistical hypothesis test based on a generalization of the likelihood ratio, and *ii)* a ML based algorithm, which exploits the high density of antennas in the LIS to obtain radio-images of the scenario. We provide a complete characterization of the statistical solution, and also pave the way to the use of ML technique to improve the performance in the second case. As an example, we have shown that the use of denoising autoencoders considerably boosts the performance of the ML algorithm. Both proposals are tested in an exemplary industrial scenario, showing that, up to relatively low values of SNR, the performance of the two presented techniques is rather similar. The ML solution implies a larger computational effort than the statistical test, but in turn does not require any a priori knowledge, as is the case of the test in which the variance of the noise is assumed in order to reduce analytical complexity. Finally, the results obtained in this system motivate a further study with more complex detectors of I/Q components to quantify the potential performance gain obtained from using I/Q receivers, i.e., analyzing the trade-off between the system complexity and its performance.

REFERENCES

[1] C. J. Vaca-Rubio *et al.*, “A primer on large intelligent surface (LIS) for wireless sensing in an industrial setting,” 2020. [Online]. Available: arXiv:2006.06563.

[2] J. G. Andrews *et al.*, “What will 5G be?” *IEEE J. Sel. Areas Commun.*, vol. 32, no. 6, pp. 1065–1082, Jun. 2014.

[3] E. G. Larsson, O. Edfors, F. Tufvesson, and T. L. Marzetta, "Massive MIMO for next generation wireless systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 186–195, Feb. 2014.

[4] N. Shlezinger, G. C. Alexandropoulos, M. F. Imani, Y. C. Eldar, and D. R. Smith, "Dynamic metasurface antennas for 6G extreme massive MIMO communications," *IEEE Wireless Commun.*, early access, Jan. 14, 2021, doi: 10.1109/MWC.001.2000267.

[5] C. Huang *et al.*, "Holographic MIMO surfaces for 6G wireless networks: Opportunities, challenges, and trends," *IEEE Wireless Commun.*, vol. 27, no. 5, pp. 118–125, Oct. 2020.

[6] S. Hu, F. Rusek, and O. Edfors, "Beyond massive MIMO: The potential of data transmission with large intelligent surfaces," *IEEE Trans. Signal Process*, vol. 66, no. 10, pp. 2746–2758, May 2018.

[7] E. Basar, M. Di Renzo, J. De Rosny, M. Debbah, M.-S. Alouini, and R. Zhang, "Wireless communications through reconfigurable intelligent surfaces," *IEEE Access*, vol. 7, pp. 116753–116773, 2019.

[8] E. Björnson, L. Sanguinetti, H. Wymeersch, J. Hoydis, and T. L. Marzetta, "Massive MIMO is a reality—What is next?: Five promising research directions for antenna arrays," *Digit. Signal Process.*, vol. 94, pp. 3–20, Nov. 2019. [Online]. Available: https://doi.org/10.1016/j.dsp.2019.06.007

[9] M. Di Renzo *et al.*, "Smart radio environments empowered by reconfigurable intelligent surfaces: How it works, state of research, and the road ahead," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2450–2525, Nov. 2020.

[10] G. C. Alexandropoulos, G. Lerosey, M. Debbah, and M. Fink, "Reconfigurable intelligent surfaces and metamaterials: The potential of wave propagation control for 6G wireless communications," 2020. [Online]. Available: arXiv:2006.11136.

[11] M. Di Renzo *et al.*, "Smart radio environments empowered by reconfigurable AI meta-surfaces: An idea whose time has come," *EURASIP J. Wireless Commun. Netw.*, vol. 2019, no. 1, pp. 1–20, 2019.

[12] Q. Wu and R. Zhang, "Intelligent reflecting surface enhanced wireless network via joint active and passive beamforming," *IEEE Trans. Wireless Commun.*, vol. 18, no. 11, pp. 5394–5409, Nov. 2019.

[13] H. Huang, Y. Song, J. Yang, G. Gui, and F. Adachi, "Deep-learning-based millimeter-wave massive MIMO for hybrid precoding," *IEEE Trans. Veh. Technol.*, vol. 68, no. 3, pp. 3027–3032, Mar. 2019.

[14] M.-A. Badiu and J. P. Coon, "Communication through a large reflecting surface with phase errors," *IEEE Wireless Commun. Lett.*, vol. 9, no. 2, pp. 184–188, Feb. 2020.

[15] H. Wymeersch, J. He, B. Denis, A. Clemente, and M. Juntti, "Radio localization and mapping with reconfigurable intelligent surfaces," 2019. [Online]. Available: arXiv:1912.09401.

[16] R. Alghamdi *et al.*, "Intelligent surfaces for 6G wireless networks: A survey of optimization and performance analysis techniques," 2020. [Online]. Available: arXiv preprint:2006.06541.

[17] J. He, H. Wymeersch, L. Kong, O. Silvén, and M. Juntti, "Large intelligent surface for positioning in millimeter wave MIMO systems," in *Proc. IEEE 91st Veh. Technol. Conf. (VTC-Spring)*, 2020, pp. 1–5.

[18] H. Shen, W. Xu, S. Gong, Z. He, and C. Zhao, "Secrecy rate maximization for intelligent reflecting surface assisted multi-antenna communications," *IEEE Commun. Lett.*, vol. 23, no. 9, pp. 1488–1492, Sep. 2019.

[19] J. D. V. Sánchez, P. Ramírez-Espinosa, and F. J. López-Martínez, "Physical layer security of large reflecting surface aided communications with phase errors," *IEEE Wireless Commun. Lett.*, vol. 10, no. 2, pp. 325–329, Feb. 2021.

[20] C. Huang, G. C. Alexandropoulos, C. Yuen, and M. Debbah, "Indoor signal focusing with deep learning designed reconfigurable intelligent surfaces," in *Proc. IEEE 20th Int. Workshop Signal Process. Adv. Wireless Commun. (SPAWC)*, Cannes, France, 2019, pp. 1–5.

[21] D. Dardari, "Communicating with large intelligent surfaces: Fundamental limits and models," *IEEE J. Sel. Areas Commun.*, vol. 38, no. 11, pp. 2526–2537, Nov. 2020.

[22] S. Hu, F. Rusek, and O. Edfors, "Beyond massive MIMO: The potential of positioning with large intelligent surfaces," *IEEE Trans. Signal Process.*, vol. 66, no. 7, pp. 1761–1774, Apr. 2018.

[23] S. Hu, F. Rusek, and O. Edfors, "Cramér-rao lower bounds for positioning with large intelligent surfaces," in *Proc. IEEE 86th Veh. Technol. Conf. (VTC-Fall)*, Toronto, ON, Canada, 2017, pp. 1–6.

[24] H. Wang, D. Zhang, Y. Wang, J. Ma, Y. Wang, and S. Li, "RT-fall: A real-time and contactless fall detection system with commodity WiFi devices," *IEEE Trans. Mobile Comput.*, vol. 16, no. 2, pp. 511–526, Feb. 2017.

[25] Q. Pu, S. Gupta, S. Gollakota, and S. Patel, "Whole-home gesture recognition using wireless signals," in *Proc. 19th Annu. Int. Conf. Mobile Comput. Netw.*, 2013, pp. 27–38.

[26] Y. Zhao, N. Patwari, J. M. Phillips, and S. Venkatasubramanian, "Radio tomographic imaging and tracking of stationary and moving people via kernel distance," in *Proc. ACM/IEEE Int. Conf. Inf. Process. Sens. Netw. (IPSN)*, Philadelphia, PA, USA, 2013, pp. 229–240.

[27] F. Adib, Z. Kabelac, H. Mao, D. Katabi, and R. C. Miller, "Demo: Real-time breath monitoring using wireless signals," in *Proc. 20th Annu. Int. Conf. Mobile Comput. Netw.*, 2014, pp. 261–262.

[28] M. Zhao *et al.*, "Through-wall human pose estimation using radio signals," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Salt Lake City, UT, USA, 2018, pp. 7356–7365.

[29] J. Wilson and N. Patwari, "Radio tomographic imaging with wireless networks," *IEEE Trans. Mobile Comput.*, vol. 9, no. 5, pp. 621–632, May 2010.

[30] S. Bozkurt, G. Elibol, S. Gunal, and U. Yayan, "A comparative study on machine learning algorithms for indoor positioning," in *Proc. Int. Symp. Innovat. Intell. SysT. Appl. (INISTA)*, Madrid, Spain, 2015, pp. 1–8.

[31] J. Joung, "Machine learning-based antenna selection in wireless communications," *IEEE Commun. Lett.*, vol. 20, no. 11, pp. 2241–2244, Nov. 2016.

[32] O. T. Demir and E. Björnson, "Channel estimation in massive MIMO under hardware non-linearities: Bayesian methods versus deep learning," *IEEE Open. J. Commun. Soc.*, vol. 1, pp. 109–124, 2020.

[33] X. Ma and Z. Gao, "Data-driven deep learning to design pilot and channel estimator for massive MIMO," *IEEE Trans. Veh. Technol.*, vol. 69, no. 5, pp. 5677–5682, May 2020.

[34] H. Huang, J. Yang, H. Huang, Y. Song, and G. Gui, "Deep learning for super-resolution channel estimation and DOA estimation based massive MIMO system," *IEEE Trans. Veh. Technol.*, vol. 67, no. 9, pp. 8549–8560, Sep. 2018.

[35] C.-K. Wen, W.-T. Shih, and S. Jin, "Deep learning for massive MIMO CSI feedback," *IEEE Wireless Commun. Lett.*, vol. 7, no. 5, pp. 748–751, Oct. 2018.

[36] C. Luo, J. Ji, Q. Wang, X. Chen, and P. Li, "Channel state information prediction for 5G wireless communications: A deep learning approach," *IEEE Trans. Netw. Sci. Eng.*, vol. 7, no. 1, pp. 227–236, Jan.–Mar. 2020.

[37] G. D. Durgin, T. S. Rappaport, and D. A. de Wolf, "New analytical models and probability density functions for fading in wireless communications," *IEEE Trans. Commun.*, vol. 50, no. 6, pp. 1005–1015, Jun. 2002.

[38] J. Sijbers, A. J. den Dekker, P. Scheunders, and D. Van Dyck, "Maximum-likelihood estimation of Rician distribution parameters," *IEEE Trans. Med. Imag.*, vol. 17, no. 3, pp. 357–361, Jun. 1998.

[39] S. Kay, *Fundamentals of Statistical Signal Processing: Detection Theory* (Fundamentals of Statistical Si). Englewood Cliffs, NJ, USA: Prentice-Hall PTR, 1998.

[40] W. F. Scott, "On the asymptotic distribution of the likelihood ratio statistic," *Commun. Stat. Theory Methods*, vol. 36, no. 2, pp. 273–281, 2007.

[41] J. Idier and G. Collewet. (Oct. 2014). *Properties of Fisher Information for Rician Distributions and Consequences in MRI*. [Online]. Available: https://hal.archives-ouvertes.fr/hal-01072813

[42] S. Provost and A. Mathai, *Quadratic Forms in Random Variables: Theory and Applications* (Statistics : Textbooks and Monographs). New York, NY, USA: Marcel Dek ker, 1992.

[43] P. Ramírez-Espinosa, D. Morales-Jimenez, J. A. Cortés, J. F. París, and E. Martos-Naya, "New approximation to distribution of positive RVs applied to Gaussian quadratic forms," *IEEE Signal Process. Lett.*, vol. 26, no. 6, pp. 923–927, Jun. 2019.

[44] T. Y. Al-Naffouri, M. Moinuddin, N. Ajeeb, B. Hassibi, and A. L. Moustakas, "On the distribution of indefinite quadratic forms in Gaussian random variables," *IEEE Trans. Commun.*, vol. 64, no. 1, pp. 153–165, Jan. 2016.

[45] S. J. Pan and Q. Yang, "A survey on transfer learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 10, pp. 1345–1359, Oct. 2010.

[46] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014. [Online]. Available: arXiv:1409.1556.
[47] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "LOF: Identifying density-based local outliers," in *Proc. ACM SIGMOD Int. Conf. Manag. Data*, 2000, pp. 93–104.
[48] *WinProp*, Altair Eng., Inc., Troy, MI, USA, 2019. [Online]. Available: https//www.altairhyperworks.com/winprop
[49] P. Roy, S. Ghosh, S. Bhattacharya, and U. Pal, "Effects of degradations on deep neural network architectures," 2018. [Online]. Available: arXiv:1807.10108.
[50] I. Goodfellow, Y. Bengio, A. Courville, and Y. Bengio, *Deep Learning*, vol. 1. Cambridge, MA, USA: MIT Press, 2016.
[51] P. Vincent, H. Larochelle, I. Lajoie, Y. Bengio, P.-A. Manzagol, and L. Bottou, "Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion," *J. Mach. Learn. Res.*, vol. 11, no. 12, pp. 3371–3408, 2010.
[52] F. Chollet *et al.* (2015). *Keras*. [Online]. Available: https://keras.io

CRISTIAN J. VACA-RUBIO received the B.Sc. and M.Sc. degrees in telematics engineering and telematic and telecommunication networks from the University of Malaga, Spain, in 2018 and 2019, respectively. He is currently pursuing the Ph.D. degree with the Connectivity Section, Electronic Systems Department, Aalborg University, under a Marie Curie Fellowship as an Early Stage Researcher in the H2020 ITN WindMill Project, where he joined in September 2019. From 2018 to July 2019, he was an Assistant Researcher with the Communication Engineering Department, University of Malaga in a joint project with DEKRA Testing and Certification S.A.U. His main research activities are in wireless communications, machine learning, mobile networks, telematics, and applied statistics in video streaming.

PABLO RAMIREZ-ESPINOSA received the M.Sc. and Ph.D. degrees in telecommunication engineering from the University of Málaga, Spain, in 2017 and 2020, respectively. From 2017 to 2020, he was an Assistant Researcher with the Communication Engineering Department, University of Málaga. In March 2020, he joined Connectivity Section, Electronic Systems Department, Aalborg University, where he is currently a Postdoctoral Researcher. He has been a Visiting Researcher with Queen's University of Belfast in 2018. His main research activities are in wireless communications, particularly massive MIMO, dynamic metasurfaces antennas, ultra-reliable low-latency communications, channel modeling and physical layer security, as well as applied statistics. He received the IEEE TRANSACTIONS ON COMMUNICATIONS Exemplary Reviewer Certificate in 2019 and 2020.

KIMMO KANSANEN (Member, IEEE) received the M.Sc. (EE) and Dr.Tech. degrees from the University of Oulu, Finland, in 1998 and 2005, respectively, where he was a Research Scientist and a Project Manager with the Centre for Wireless Communications, University of Oulu. He is a Professor with the Norwegian University of Science and Technology, where he heads the Signal Processing Group. He has participated in research projects and Early Training Networks in FP6, FP7, H2020, TEKES, and Research Council of Norway. He is an Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS and *Frontiers in Communications and Networks*.

ZHENG-HUA TAN (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from Hunan University, Changsha, China, in 1990 and 1996, respectively, and the Ph.D. degree in electronic engineering from Shanghai Jiao Tong University (SJTU), Shanghai, China, in 1999. He is a Professor with the Department of Electronic Systems and a Co-Head of the Centre for Acoustic Signal Processing Research, Aalborg University, Aalborg, Denmark. He was a Visiting Scientist with the Computer Science and Artificial Intelligence Laboratory, MIT, Cambridge, USA, an Associate Professor with SJTU, and a Postdoctoral Fellow with KAIST, Daejeon, South Korea. He has authored/coauthored over 200 refereed publications. His research interests include machine learning, deep learning, pattern recognition, speech and speaker recognition, noise-robust speech processing, multimodal signal processing, and social robotics. He is the Chair of the IEEE Signal Processing Society Machine Learning for Signal Processing Technical Committee. He was the General Chair for IEEE MLSP 2018 and a TPC Co-Chair for IEEE SLT 2016. He is an Associate Editor for the IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH AND LANGUAGE PROCESSING. He has served as an Editorial Board Member for Computer Speech and Language and was a Guest Editor for the IEEE JOURNAL OF SELECTED TOPICS IN SIGNAL PROCESSING and *Neurocomputing*.

ELISABETH DE CARVALHO (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from Telecom ParisTech, France. After her Ph.D. degree, she was a Postdoctoral Fellow with Stanford University, and then worked in industry in the field of DSL and wireless LAN. Since 2005, she has been with Aalborg University, where she is a Professor and has led several research projects in wireless communications. She has coauthored the book *A Practical Guide to MIMO Radio Channel*. She is the Project Manager of the H2020 Innovative Training Network WindMill focusing on machine learning for wireless communications. Her main expertise is in the field of signal processing with emphasis on MIMO communications. She is the Co-Chair of the Globecom 2021, Selected Areas in Communication Symposium: Machine Learning for Communications. She is a member of IEEE Signal Processing Society SPCOM technical committee and the Vice Chair of the IEEE COMSOC Emerging Technology Initiative on Machine Learning for Communications.

PETAR POPOVSKI (Fellow, IEEE) received the Dipl.-Ing. and M.Sc. degrees in communication engineering from the University of Sts, and the Cyril and Methodius in Skopje and the Ph.D. degree from Aalborg University in 2005, where he is a Professor with Aalborg University, where he heads the Section on Connectivity. He has authored the book *Wireless Connectivity: An Intuitive and Fundamental Guide* (Wiley, 2020). His research interests are in the area of wireless communication and communication theory. He received an ERC Consolidator Grant in 2015, the Danish Elite Researcher Award in 2016, the IEEE Fred W. Ellersick Prize in 2016, the IEEE Stephen O. Rice Prize in 2018, the Technical Achievement Award from the IEEE Technical Committee on Smart Grid Communications in 2019, and the Danish Telecommunication Prize in 2020. He was the General Chair for IEEE SmartGridComm 2018 and IEEE Communication Theory Workshop 2019. He is also a Member at Large at the Board of Governors in IEEE Communication Society, the Vice-Chair of the IEEE Communication Theory Technical Committee, and Steering Committee Member of IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING. He is currently an Area Editor of the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.