Aalborg Universitet

Floor Map Reconstruction Through Radio Sensing and Learning By a Large Intelligent Surface

Vaca-Rubio, Cristian J.; Pereira, Roberto; Mestre, Xavier; Gregoratti, David; Tan, Zheng-Hua; Carvalho, Elisabeth de; Popovski, Petar

FLOOR MAP RECONSTRUCTION THROUGH RADIO SENSING AND LEARNING BY A LARGE INTELLIGENT SURFACE

Cristian J. Vaca-Rubio[★,§], *Roberto Pereira*[†,§], *Xavier Mestre*[†], *David Gregoratti*[‡],
Zheng-Hua Tan[★], *Elisabeth de Carvalho*[★] and *Petar Popovski*[★]

[★] Department of Electronic Systems, Aalborg University, Aalborg, Denmark.
Email: {cjvr, zt, edc, petarp@es.aau.dk}
[†] ISPIC, Centre Tecnològic Telecomunicacions Catalunya, Barcelona, Spain.
Email: {rpereira, xmestre}@cttc.es
[‡] Software Radio Systems, Barcelona, Spain. Email: david.gregoratti@ieee.org

ABSTRACT

Environmental scene reconstruction is of great interest for autonomous robotic applications, since an accurate representation of the environment is necessary to ensure safe interaction with robots. Equally important, it is also vital to ensure reliable communication between the robot and its controller. Large Intelligent Surface (LIS) is a technology that has been extensively studied due to its communication capabilities. Moreover, due to the number of antenna elements, these surfaces arise as a powerful solution to radio sensing. This paper presents a novel method to translate radio environmental maps obtained at the LIS to floor plans of the indoor environment built of scatterers spread along its area. The usage of a Least Squares (LS) based method, U-Net (UN) and conditional Generative Adversarial Networks (cGANs) were leveraged to perform this task. We show that the floor plan can be correctly reconstructed using both local and global measurements.

Index Terms— Sensing, Computational Imaging, LIS, Machine Learning for Communication.

1. INTRODUCTION

Mobile robots with mapping devices have been used to estimate indoor floor plans. The state-of-the-art methods are usually based on optical sensors to obtain the indoor maps in detail. For instance, Light Detection and Ranging (LIDAR) [1], depth cameras [2] and RGB cameras [3]. Although these methods achieve acceptable accuracy, they have some limitations. LIDAR might not be able to capture all types of materials meanwhile approaches based on cameras are dependent on the lighting conditions. To circumvent this, acoustic-based approaches such as microphones [4] and ultrasonic sensors [5] are robust to lighting conditions. However, they have a limited sensing range and might malfunction in noisy environments.

On a related note, radar devices such as millimeter Wave (mmWave) radars have become popular for indoor sensing applications. They actively transmit Radio Frequency (RF) signals to monitor the reflections to sense nearby scatters' parameters such as range, speed or angle. Hence, they can be used indoors in poor lighting conditions. These radars have been used in applications such as human sensing [6] and floor-plan reconstruction [7]. However, mmWave radars work in high frequency bands, leading to short wavelengths. This leads to high energy attenuation over distance and weak penetrability through walls.

In the context of 6G, sensing has become a fundamental feature. For instance, the authors in [8] explore this sensing-style capabilities of a mMIMO BS to jointly learn an antenna selection and a range-azimuth map of a beamforming gain. With the increasing number of receivers, Large Intelligent Surface (LIS) becomes a natural extension of the massive MIMO technology which designates a continuous electromagnetic surface able to transmit and receive radio waves. In practice, they are planar arrays conformed by a huge amount of closely spaced tiny antenna elements. There is a vast range of studies analyzing its application in communications [9]. However, very little has been studied regarding its sensing capability [10]. Consequently, in [11] we presented a method that enables reconstructing a radio map of the propagation environment using an indoor LIS deployment in the ceiling. This allows for tracking both active and passive users.

Motivated by these results and on the increasing interest in both sensing and LIS, in this work, we further explore the capabilities of LIS and provide a floor plan reconstruction based on the LIS received signal. We leverage the usage

of Deep Generative Model (DGM) to learn a map from the complex-valued received signals at the LIS to the RGB image of the corresponding floor plan. Differently from previous works [1, 4, 7], we are mainly concerned with enabling sensing capabilities in a system primarily designed for communication. Consequently, such applications further exploit the usability of already deployed hardware. We assess the reconstruction performance with the original (ground-truth) floor plan composed of multiple elements representing, for example, robots, furniture and appliances

2. SYSTEM AND PROBLEM FORMULATION

Consider an indoor factory scenario in which K_a active devices are deployed. These devices can be anything using a transmitter such as robots, smartphones or IoT devices. More importantly, since they are active devices, we can assume these devices to be communicating with a receiver through a wireless channel. In a real-world scenario, this communication often assists with the task at hand, e.g., sending commands to the robots. In our scenario, this receiver is a LIS placed on the factory's roof and, apart from assisting with the communication task, the LIS also tries to map the current environment of this factory. This environment sensing can be divided into two steps:

1. Mapping from the *environmental signals* into a radio map of the environment from the LIS viewpoint, similar to [11]. This allows converting the complex-valued data into a RGB real-valued data which is more natural to neural networks.

2. Translation of the radio map into its current floor plan pattern, i.e., the disposition of the elements in this environment. In summary, based on the raw complex-valued signals received at the LIS, we aim to reconstruct the current arrangement of the passive (non-active) elements present in the environment.

For the second step, we rely on conditional Generative Adversarial Networks (cGANs) which has been widely used in the literature for the task of image-to-image reconstruction [12].

Moreover, we assume this LIS to be equipped with $N = N_x \times N_y$ antenna elements and its physical aperture comprises its whole area. Concretely, we consider a square LIS composed of isotropic antennas and physical effects such as mutual coupling are ignored. A bit more formally, the sensing problem consists in obtaining a radio map that describes the environment from the superposition of the received signals from each of the $1 < k \leq K_a$ users at every element of the LIS. This map contains information on the K_a active devices involved in the scenario as well as the K_p passive objects/scatters.

The superposed complex baseband signal received at the LIS is given by

$$\mathbf{y} = \sum_{k=1}^{K_a} \mathbf{h}_k x_k + \mathbf{n}, \tag{1}$$

with x_k the transmitted (sensing) symbol from user k (we consider $x_k = 1$ without loss of generality), $\mathbf{h}_k \in \mathbb{C}^{N \times 1}$ the channel vector from a specific position of user k to each antenna-element, and $\mathbf{n} \sim \mathcal{CN}_N(\mathbf{0}, \sigma^2 \mathbf{I}_N)$ the noise vector. To avoid frequency selectivity, we consider a narrowband transmission.

As a consequence of the large physical dimensions of the LIS in comparison with the distance from the transmitters to the roof, we need to account for spherical wave propagation during the modeling. The spherical-wave channel coefficient $h_{sp,n}$ at the LIS n-th element from an arbitrary active device transmission is proportional to [13]

$$h_{sp,n} \propto \frac{1}{d_n} e^{-j\frac{2\pi}{\lambda} d_n}, \tag{2}$$

where $d_n = \sqrt{(x_n - x_k)^2 + (y_n - y_k)^2 + (z_n - z_k)^2}$ denotes the distance between the active device k and the n-th antenna. To obtain the radio map, expressing the LIS in a vectorized version for ease of notation, we can derive a Matched Filter (MF) procedure such that [11]

$$\mathbf{y}_{mf} = \mathbf{h}_{sp} * \mathbf{y}, \tag{3}$$

denoting $*$ the spatial convolution operator, $\mathbf{h}_{sp} \in \mathbb{C}^{N_f \times 1}$ the expected spherical pattern (steering vector) for N_f antennas LIS deployment on (2), $\mathbf{y}$ the received signal from (1) and $\mathbf{y}_{mf} \in \mathbb{C}^{N \times 1}$ the filtered output signal. As the convolution operator would reduce the output dimension (Due to the 2D convolution along the received signal at the LIS), we zero-pad $\mathbf{y}$ such that we guarantee $\mathbf{y}_{mf} \in \mathbb{C}^{N \times 1}$ dimension. To obtain a radio map, we just need to compute the energy at the output of the MF procedure $|\mathbf{y}_{mf}| \in \mathbb{R}^{N \times 1}$. We then map the values to the RGB scale such that $F : \mathbb{R}^{N \times 1} \rightarrow \{[0, 255] \cap \mathbb{N}\}^{N \times 3}$

$$\mathbf{y}_m = F(|\mathbf{y}_{mf}|) \tag{4}$$

that represents the radio map. Please note, we have to know the frequency f and assume a distance z from the transmitter to the LIS to design the filter[1]. In our work, we design a filter for $f = 3.5$ GHz, $z = 8$ m and $N_f = 100 \times 100$ antenna elements $\frac{\lambda}{2}$ spaced[2]. Let us assume we can obtain S samples at each channel coherence interval. In this way, we can use these extra samples to perform an S-averaging of the received signal measurements at the LIS viewpoint, reducing the noise variance contribution and enhancing the quality

---

[1]For a more detailed explanation of the radio maps, we gently refer the readers to [11]. The distance z is a parameter for the filter design. This does not imply that in the evaluation, all the transmitters or scatters are fixed at this distance.

[2]Note we use a fixed kernel with $z = 8$ m which corresponds to the height of the building. We do not need to calibrate the MF to specific distances.

(a) Floor Plan   (b) $S = 1$

(c) $S = 100$ (d) $S = 1000$
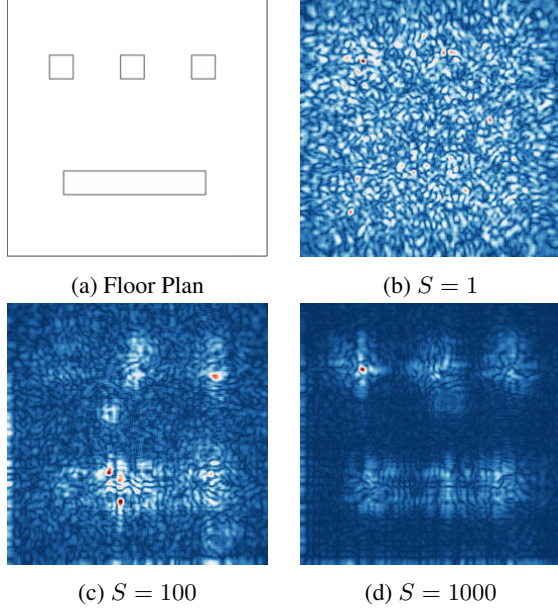
Fig. 1: Radio maps with its corresponding floor plan obtained over S-averaging channel samples acquired by LIS in a $\gamma = -10$ dB Signal-to-Noise Ratio (SNR) condition.

of the obtained radio map. Figure 1(a) shows a ground-truth floor plan. Looking at Figure 1(b)-(d) we can see three exemplary radio maps representing the ground-truth floor plan. The radio map captures the reflections of the scatters (rectangular and square shapes representing walls/objects) that act as virtual sources. The S-averaging effect in reducing the noise contribution leads to an enhancement in the radio map quality. The target of our model will be translating these radio maps to the ground-truth floor plan.

2.1. Received signal and noise modeling

In order to simulate the propagation environment in the most reliable way, we resort to ray tracing [14]. From the ray-tracing simulation, the received signal in (1) is obtained as the complex electric field arriving at the n-th antenna element, $\widetilde{E}_n$, which can be regarded as the superposition of each ray path $r \in N_r$ from every $k \in K_a$ user, i.e.,

$$\widetilde{E}_n = \sum_{k=1}^{K_a} \sum_{r=1}^{N_r} \widetilde{E}_{n,r,k} = \sum_{k=1}^{K_a} \sum_{r=1}^{N_r} E_{n,r,k} e^{j\phi_{n,r,k}}. \quad (5)$$

Then, the complex signal at the output of the n-th element is therefore given by

$$y_n = \sqrt{\frac{\lambda^2 Z_n}{4\pi Z_0}} \widetilde{E}_n + n_n, \quad (6)$$

with λ the wavelength, $Z_0 = 120\pi$ the free space impedance and Z_n the antenna impedance. For simplicity, we consider

$Z_n = 1 \,\forall\, n$. Finally, we define the SNR, γ, as

$$\gamma \triangleq \frac{\lambda^2}{4\pi Z_0 N \sigma^2} \sum_{n=1}^{N} |\widetilde{E}_n|^2, \quad (7)$$

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

3. RECONSTRUCTION LEARNING

3.1. Least Squares

This method tries to naively find the best linear mapping $\mathbf{W}^* \in \mathbb{R}^{N \times N}$ from the i-th radio map $\mathbf{y}_m^{(i)} \in \mathbb{R}^N$ into the ith floor plan $\mathbf{x}^{(i)} \in \mathbb{R}^N$ by minimizing the average least square error

$$\mathbf{W}^* = \arg\min_{\mathbf{W}} \frac{1}{\mathrm{T}} \sum_{i=1}^{\mathrm{T}} \left(\mathbf{x}^{(i)} - \mathbf{W}\mathbf{y}_m^{(i)} \right)^2 \quad (8)$$
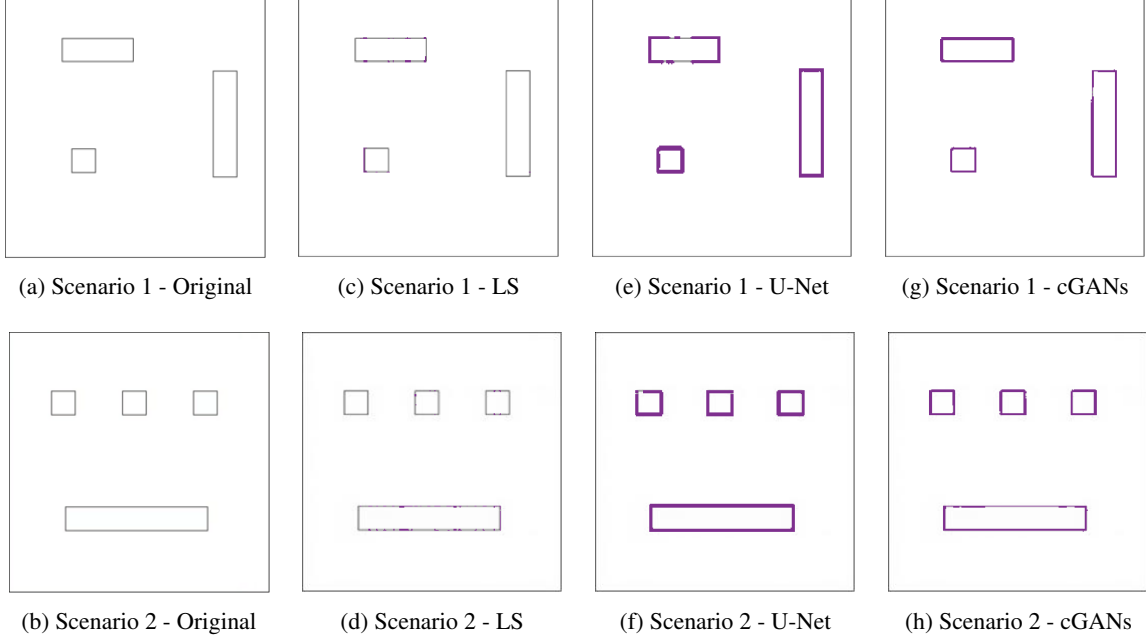
over the T training samples. Prediction is then performed on a new sample by $\hat{\mathbf{x}}^{(j)} = \mathbf{W}\mathbf{y}_m^{(j)}$. Despite its simplicity, similar solutions have been widely applied in many signal-to-image reconstruction, such as medical applications [15] and synthetic aperture radars sensing [16]. In our scenario, LS method often returns a noisy version of the environment reconstruction. Hence, to simulate the best possible set of post-processing operations that improve the LS estimator, we redefine $\hat{\mathbf{x}}^{(j)} = \min(\mathbf{x}^{(j)}, \hat{\mathbf{x}}^{(j)})$ with the entry-wise minimum operator, i.e., we remove the noise area that is present outside of the area of interest.

3.2. U-Net

U-nets are essentially autoencoder networks with skipped connections [17]. At the encoder side, we learn the feature mapping of an image while converting it to a vector. U-Net extends this vector to a segmented picture, utilizing the same feature maps as were used for compression at the encoder side (i.e. skipped connections). This would keep the image's structural integrity and provide information to the decoder to perform the segmentation. Image segmentation can be applied to our specific problem. We can see the groundtruth floor plan $\mathbf{x}$ as the segmentation of the radio map $\mathbf{y}_m$ composed of two classes, background and shapes. To reduce training time and prevent overfitting, we will use a pre-trained model for the encoder, the mobileNet-v2 [18], while for the decoder we will use the up-sampling block of Pix2Pix [12]. Finally, the U-Net minimizes the Binary Cross-Entropy (BCE) loss.

3.3. Conditional Generative Adversarial Network

We design our problem as a cGANs learning procedure by designing a generator for the distribution p_g over the floor plan data of the environment $\mathbf{x}$, given as conditional information the radio map $\mathbf{y}_m$ from (4). The generator models a mapping function from a prior noise distribution $p_{\mathbf{z}}(\mathbf{z})$ to the floor

(a) Scenario 1 - Original	(c) Scenario 1 - LS	(e) Scenario 1 - U-Net	(g) Scenario 1 - cGANs
(b) Scenario 2 - Original	(d) Scenario 2 - LS	(f) Scenario 2 - U-Net	(h) Scenario 2 - cGANs

Fig. 2: Visual representation of the two original (a)-(b) floor plans and their and reconstructions (c)-(h) using the three different methods. Reconstructions are highlighted in purple and the original floor plans in black.

plan space $G(\mathbf{z}|\mathbf{y}_m; \theta_g)$. Similarly, we design a discriminator $D(\mathbf{x}|\mathbf{y}_m; \theta_d)$ that outputs a scalar value representing the probability that $\mathbf{x}$ came form training data instead of p_g. The target min-max cost function is given by

$$\mathcal{L}_{cGANs} = \min_G \max_D \mathbb{E}_{\mathbf{x} \sim p_{\text{data}}(\mathbf{x})}[\log D(\mathbf{x}|\mathbf{y}_m)] +$$
$$\mathbb{E}_{\mathbf{z} \sim p_{\mathbf{z}}(\mathbf{z})}[1 - \log D(G(\mathbf{z}|\mathbf{y}_m))]. \quad (9)$$

More specifically, we re-implement a well-known model used in the field of image-to-image translation [19]. The main difference with respect to the general cGANs framework resides in the loss function. The authors of [19] improve the adversarial loss by incorporating a feature matching loss. It works by extracting features from several layers of the discriminator, trying to match these in-between representations among the real and the fake data. The new loss function can be expressed as

$$\mathcal{L}^*_{cGANs} = \min_G \max_D \mathcal{L}_{cGANs} \quad (10)$$
$$+ \lambda \mathbb{E}_{\mathbf{x}, \mathbf{y}_m, \mathbf{z}} \sum_{l=1}^{L} \frac{1}{N_l}[\|D^{(l)}(\mathbf{x}|\mathbf{y}_m) - D^{(l)}(\mathbf{z}|\mathbf{y}_m)\|_1],$$

where L denotes the number of layers, N_l denotes the number of elements in each layer and λ is an hyper-parameter that controls the weight of the terms. The model will learn a mapping between the wireless environmental signals and its corresponding floor plan, i.e., it will perform a translation from the radio map $\mathbf{y}_m$ to the floor plan $\mathbf{x}$.

4. DATASET DESCRIPTION

In the ray-tracing simulation, we consider two scenarios of size $10.34 \times 10.34 \times 8$ m with different scatters arrangement. We deploy an LIS with 259×259 elements separated $\lambda/2$ apart. Every K_a active device transmits a narrowband signal of 20 dBm at 3.5 GHz. The distance from which the MF is calibrated is $z = 8$ m (building height). The scatters are modeled as metallic (with conductivity $s = 19444$ S/m, relative permittivity $\epsilon = 1$ and relative permeability $\mu = 20$) and the walls as brick (with $s = 0.078$ S/m, $\epsilon = 4$ and $\mu = 1$)[3]. Figure 2(a)-(b) show the distribution of the metallic elements in the two floor plans considered in this work. The dataset is obtained by sampling the received signal at each element of the LIS from all the different active K_a in the scenario. Then the resulting signal is processed following the process described in Section II to generate the radio map.

To guarantee generalization, we conform a dataset composed by a selection of $S \in [1, 100, 1000]$ and $K_a \in [5, 20]$ for the two scenarios presented. Finally, we are primarily interested in poor signaling conditions, hence, we set $\gamma = -10$ dB which means that there is much noise in the communication between the active transmitters and the LIS. Then, formally, the dataset is denoted as $\{\mathbf{y}_m^{(i)}, \mathbf{x}^{(i)}\}_{i=1}^T$, where $\mathbf{y}_m^{(i)}$ is the i-th N-dimensional training input features vector (radio map) and $\mathbf{x}^{(i)}$ is the i-th N-dimensional target vector (floor plan). Our dataset is composed of 2400 samples which are split into 70/10/20 for training, validation and test set, respectively.

---

[3]These values are provided by the software manual [14].

	Method	PSNR	SSIM	Error distance (cm)		
				$K_a = 5$	$K_a = 20$	Average
Scenario 1	LS	12.71 ± 1.29	0.19 ± 0.03	-	-	-
	U-Net	21.51 ± 0.24	0.93	5.68 ± 5.67	3.51 ± 3.50	4.59 ± 4.59
	cGANs	$\mathbf{31.77 \pm 0.3}$	$\mathbf{0.95}$	$\mathbf{4.58 \pm 4.50}$	$\mathbf{1.96 \pm 1.95}$	$\mathbf{3.27 \pm 3.27}$
Scenario 2	LS	11.896 ± 1.09	0.232 ± 0.04	-	-	-
	U-Net	21.51 ± 0.17	0.94	$\mathbf{5.27}$	$\mathbf{3.64 \pm 8.02}$	$\mathbf{4.45 \pm 4.01}$
	cGANs	$\mathbf{32.66 \pm 0.24}$	$\mathbf{0.95}$	$7.99 + \pm 8.9$	7.89	7.94 ± 4.45

Table 1: Average qualitative comparison of LS, U-Net and cGANs in terms of PSNR, SSIM and scattering centroids error distance.
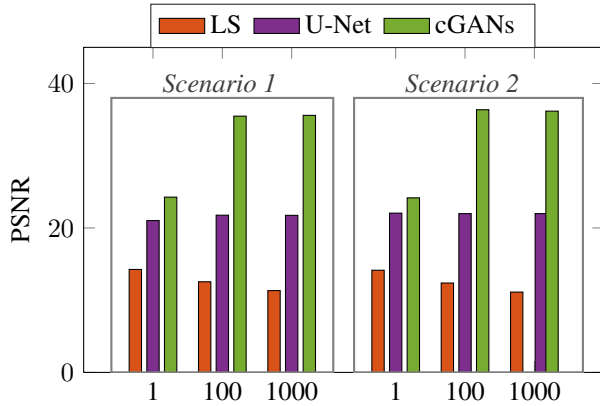
Fig. 3: Comparison of floor plan reconstruction quality based on the Average Peak Signal-to-Noise Ratio (PSNR) metric along the test set for varying number of S.

5. NUMERICAL RESULTS AND DISCUSSION

In order to validate the proposed methodology, we rely on standard image quality metrics, to evaluate the performance of the reconstructed environment using the three methods considered in this work (LS, U-Net and cGANs), specifically, we evaluate these methods using PSNR and Structural Similarity Index Measure (SSIM) [20] as quality metrics. They provide a general evaluation of how similar the reconstruction is to the original floor plan. However, this evaluation is performed based on the whole image. Hence, we also calculate the distance (in centimeters) between the centres of the scatters in the original and predicted environment. This further provides us with a way to assess local spatial predictions.

Let us first analyse the general reconstructions capability of each of these methods. Figure 3 compares the average PSNR obtained on the test set by the three methods considering different S-averaging of the received signal. As expected, the original signal $S = 1$ provides little information and hence makes it unfeasible to obtain good quality reconstruction (perfect reconstruction would lead to a maximum PSNR of 48 dB). Moreover, we notice that specifically for the linear mapping (LS), the quality of the reconstruction dras-tically decays as S-averaging increases. This happens despite the post-processing mechanisms performed (see Sec. 3) and the quality of the data used for training - we also considered training different models for specific scenarios, i.e., S-averaging and number of active transmitters, but obtained similar results. Consequently, it is reasonable to assume that it is extremely hard to find a direct linear mapping from the input radio map to the desired output. This further motivates the application of machine learning to this problem. In fact, for U-Net and cGANs, we notice the opposite behaviour, i.e., the larger S, the larger the PSNR obtained. In all scenarios the best results are obtained using cGANs. The overall generalisation of the methods is compared by considering the average reconstruction results over all different signal configurations $S \in [1, 100, 1000]$, $K_a \in [5, 20]$ and for the two different scenarios. Table 1 contains the PSNR, SSIM and average distance error of the predicted scattered objects[4]. The best results are highlighted. Indeed, the average quantity of these metrics is consistent with what is explained above. More interesting, though, is the error distance (in centimetres) between the central position of the scattered objects and their reconstruction. For the original floor plan, these centres are the middle point of each rectangle. In the reconstructed images, we obtain these central positions by approximating polygons to the scatter. The cGANs exhibit the best performance in most of the comparative analyses. It obtains the highest PSNRs and SSIM in both scenarios and the smallest average distance for *Scenario 1*. Specifically, for *Scenario 2*, U-Net obtains smaller error distances among the true and the predicted position of the scatters. However, in real-world space, this difference is almost negligible. An exemplary visual reconstruction for $S = 100$ is provided in Figure 2(e)-(h) for both U-Net and cGANs. Finally, since LS provides little information (see Figure 2(c)-(d) for a visual comparison) on the reconstructed scenario, it is impossible to perform any analysis on the error distance between the true location of the scatters and their predicted locations.

---

[4]We do not denote the variance of the results when it is negligible.

6. CONCLUSION

We have demonstrated a proof-of-concept that it is possible to learn an environmental scene reconstruction using signals received at a LIS while it is performing communication. This sensing procedure can be done in parallel and without interfering with the communication task. We have shown that using both U-Net and cGANs it is possible to accurately (with less than 16 cm error) estimate the central position of the scatters present in the environment directly from these image reconstructions. In future work, we plan to explore real-world environments as well as explore the learning environment to enhance communication performance.

7. REFERENCES

[1] H. Surmann, A. Nüchter, and J. Hertzberg, "An autonomous mobile robot with a 3D laser range finder for 3D exploration and digitalization of indoor environments," *Robotics and Autonomous Systems*, vol. 45, no. 3-4, pp. 181–198, 2003.

[2] Y. Zhang, C. Luo, and J. Liu, "Walk&Sketch: Create Floor Plans with an RGB-D Camera," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, 2012, pp. 461–470.

[3] R. Gao, M. Zhao, T. Ye, F. Ye, Y. Wang, K. Bian, T. Wang, and X. Li, "Indoor Floor Plan Reconstruction via Mobile Crowdsensing," in *Proceedings of the 20th annual international conference on Mobile computing and networking*, 2014, pp. 249–260.

[4] B. Zhou, M. Elbadry, R. Gao, and F. Ye, "BatMapper: Acoustic Sensing Based Indoor Floor Plan Construction Using Smartphones," in *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, 2017, pp. 42–55.

[5] K. S. Chong and L. Kleeman, "Feature-based Mapping in Real, Large Scale Environments using an Ultrasonic Array," *The International Journal of Robotics Research*, vol. 18, no. 1, pp. 3–19, 1999.

[6] A. D. Singh, S. S. Sandha, L. Garcia, and M. Srivastava, "RadHAR: Human Activity Recognition from Point Clouds Generated through a Millimeter-wave Radar," in *Proceedings of the 3rd ACM Workshop on Millimeter-wave Networks and Sensing Systems*, 2019, pp. 51–56.

[7] C. X. Lu, S. Rosa, P. Zhao, B. Wang, C. Chen, J. A. Stankovic, N. Trigoni, and A. Markham, "See Through Smoke: Robust Indoor Mapping with Low-cost mmWave Radar," in *Proceedings of the 18th International Conference on Mobile Systems, Applications, and Services*, 2020, pp. 14–27.

[8] T. Weiss, N. Peretz, S. Vedula, A. Feuer, and A. Bronstein, "Joint optimization of system design and reconstruction in mimo radar imaging," in *2021 IEEE 31st International Workshop on Machine Learning for Signal Processing (MLSP)*. IEEE, 2021, pp. 1–6.

[9] R. J. Williams, P. Ramírez-Espinosa, E. De Carvalho, and T. L. Marzetta, "Multiuser mimo with large intelligent surfaces: Communication model and transmit design," in *ICC 2021-IEEE International Conference on Communications*. IEEE, 2021, pp. 1–6.

[10] C. J. Vaca-Rubio, P. Ramirez-Espinosa, K. Kansanen, Z.-H. Tan, E. De Carvalho, and P. Popovski, "Assessing wireless sensing potential with large intelligent surfaces," *IEEE Open Journal of the Communications Society*, vol. 2, pp. 934–947, 2021.

[11] C. J. Vaca-Rubio, P. Ramirez-Espinosa, K. Kansanen, Z.-H. Tan, and E. de Carvalho, "Radio Sensing with Large Intelligent Surface for 6G," *arXiv preprint arXiv:2111.02783*, 2021.

[12] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-Image Translation with Conditional Adversarial Networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017.

[13] Z. Zhou, X. Gao, J. Fang, and Z. Chen, "Spherical wave channel and analysis for large linear array in LoS conditions," in *2015 IEEE Globecom Workshops (GC Wkshps)*. IEEE, 2015, pp. 1–6.

[14] Feko, altair engineering, inc. https://www.altairhyperworks.com/feko.

[15] N. Vaswani, "LS-CS-Residual (LS-CS): Compressive Sensing on Least Squares Residual," *IEEE Transactions on Signal Processing*, vol. 58, no. 8, 2010.

[16] M.-S. Kang and K.-T. Kim, "Compressive sensing based SAR imaging and autofocus using improved Tikhonov regularization," *IEEE Sensors Journal*, vol. 19, no. 14, pp. 5529–5540, 2019.

[17] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.

[18] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 4510–4520.

[19] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8798–8807.

[20] A. Hore and D. Ziou, "Image quality metrics: PSNR vs. SSIM," in *2010 20th international conference on pattern recognition*. IEEE, 2010, pp. 2366–2369.