

PAPER • OPEN ACCESS

# A deep learning enhanced inverse scattering framework for microwave imaging of piece-wise homogeneous targets

To cite this article: Álvaro Yago Ruiz *et al* 2024 *Inverse Problems* **40** 045001

View the [article online](#) for updates and enhancements.

## You may also like

- [MRI-based electric properties tomography with a quasi-Newton approach](#)  
Anar Rahimov, Amélie Litman and Guillaume Ferrand
- [Increasing stability of a linearized inverse boundary value problem for a nonlinear Schrödinger equation on transversally anisotropic manifolds](#)  
Shuai Lu and Jian Zhai
- [Fluctuations in heat engines](#)  
Viktor Holubec and Artem Ryabov

# A deep learning enhanced inverse scattering framework for microwave imaging of piece-wise homogeneous targets

Álvaro Yago Ruiz<sup>1</sup> , Maria Nikolic Stevanovic<sup>2</sup> ,  
Marta Cavagnaro<sup>3</sup>  and Lorenzo Crocco<sup>1,\*</sup> 

<sup>1</sup> CNR—IREA, National Research Council of Italy, Institute for the Electromagnetic Sensing of the Environment, Napoli, Italy

<sup>2</sup> University of Belgrade, Belgrade, Serbia

<sup>3</sup> DIET, Sapienza University of Rome, Roma, Italy

E-mail: [crocco.l@irea.cnr.it](mailto:crocco.l@irea.cnr.it)

Received 17 July 2023; revised 15 January 2024

Accepted for publication 1 February 2024

Published 21 February 2024



CrossMark

## Abstract

In this paper, we present a framework for the solution of inverse scattering problems that integrates traditional imaging methods and deep learning. The goal is to image piece-wise homogeneous targets and it is pursued in three steps. First, raw-data are processed via orthogonality sampling method to obtain a qualitative image of the targets. Then, such an image is fed into a U-Net. In order to take advantage of the implicitly sparse nature of the information to be retrieved, the network is trained to retrieve a map of the spatial gradient of the unknown contrast. Finally, such an *augmented shape* is turned into a map of the unknown permittivity by means of a simple post-processing. The framework is computationally effective, since all processing steps are performed in real-time. To provide an example of the achievable performance, Fresnel experimental data have been used as a validation.

Keywords: inverse scattering, microwave imaging, deep learning, qualitative methods, orthogonality sampling method

\* Author to whom any correspondence should be addressed.



Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

## 1. Introduction

Microwave imaging (MWI) exploits the capability of electromagnetic (EM) waves to penetrate material bodies to enable the non-invasive inspection of unknown scenarios that are otherwise not directly accessible. As such, MWI is relevant to several applications fields as different as biomedical imaging [1, 2], subsurface sensing [3], food security monitoring [4], or through-wall imaging [5].

By measuring the scattered field arising from the interaction of a known EM incident field with a target, it is possible to obtain an image depicting the EM properties (i.e. dielectric permittivity and conductivity) of such target, as well as its morphology. From a mathematical point of view, MWI corresponds to the solution of an inverse scattering problem (ISP), which is a well-known non-linear and ill-posed inverse problem [6].

To cope with the difficulties of the ISP, many solution methods have been developed [7–9]. However, still no ‘universal’ method exists. For instance, *quantitative* methods [7, 8], which aim at the complete solution of the ISP, are computationally demanding, prone to the occurrence of *false solutions*<sup>4</sup>, and often rely on available a-priori information on the targets to perform a successful reconstruction. Conversely, *qualitative* methods [9] cast the ISP in terms of an auxiliary linear ill-posed problem, thus overcoming non-linearity and requiring an almost negligible computational burden. However, besides still having to face an ill-posed problem, they can only provide explicit information on the target’s morphology and not on its EM properties.

Recently, a huge interest in the literature has been devoted to the possibility of addressing the ISP non-linearity and ill-posedness resorting to computational methods based on the deep learning (DL) paradigm [11, 12]. Different from traditional approaches, DL is data-driven: common DL architectures run an optimization procedure (the training) from which a model is built by analyzing a collection of examples.

Among the possible ways to exploit such data-driven approach in ISP solution [11], *physics-assisted* techniques are worth to be considered. In these approaches, domain knowledge in the specific problem at hand is incorporated in the internal structure of the DL architecture or provided into its inputs by pre-processing the raw data. For MWI, this represents a particularly convenient strategy, since MWI data are not ‘homogeneous’, as they can be collected in different conditions (e.g. number and position of the probes or operating frequency). As such, MWI data are usually not abundant enough to enable a *direct learning* approach which solely relies on the scattering measurements. In fact, embedding domain knowledge allows the training with less examples than a *direct learning* counterpart, as the model does not have to ‘learn’ the all the physics involved in the problem [11].

The most common domain knowledge incorporation is carried out by pre-processing the MWI scattering measurements with traditional imaging algorithms. In doing so, convolutional neural network (CNN) models, which are known to be excellent image processing frameworks [13], could be employed. To this end, a crucial aspect that must be considered is the choice of the MWI algorithm. First of all, since DL models work in real-time (once trained), computationally intensive quantitative methods are not suitable if speed is a requirement, since they would act as a bottleneck in the processing workflow. Also, it is worth recalling that the

<sup>4</sup> A false solution is an estimate of the unknown that fits the data but is different from the ground truth. False solutions are a consequence of both the non-linearity and ill-posedness of the ISP and arise when the ISP is solved via local iterative optimization. Theoretically, global optimization methods could circumvent false solutions occurrence and converge to a global optimum. In practice they cannot, due to the *curse of dimensionality* [10] arising from the exponential growth of the computational cost with the number of unknowns.

possible occurrence of false solutions is an issue, as the output of quantitative methods is to some extent not predictable. Finally, in both qualitative and quantitative methods, the need of tuning regularization parameters poses an issue on the possibility of full-automated operation, which is another attractive feature of the DL paradigm.

In [14], such difficulties have been addressed training a physics-assisted CNN to image piece-wise homogeneous targets from input images obtained using two techniques or schemes, back-propagation scheme (BPS) and dominant current scheme (DCS). The chosen CNN architecture is the U-Net [15], which is very popular as a tool to face computer vision tasks, as it can be configured to handle an image in input and provide an output which is still an image [16]. Notably, the authors show that the U-Net trained either with BPS or DCS outperforms a *direct learning* implementation with the raw scattering measurements. Although both BPS and DCS do not slow down the processing workflow, and may be used in real-time, they still present some drawbacks. More precisely, BPS is based on the linearized back-propagation algorithm and therefore may lead to significantly inaccurate images when the assumptions underneath the linearization are violated. Whereas, DCS do not require approximations, but are dependent on the measurement configuration, so that the network has to be retrained every time the set-up is changed. Also, both methods provide discretized images in which the discretization step is dictated by the working wavelength, thus posing a constraint on the size of the images fed into the network.

Motivated by the above considerations, the authors of this work have considered the use of the orthogonality sampling method (OSM) [17] as the domain knowledge-embedding imaging algorithm [18, 19]. The OSM is a qualitative method introduced by Roland Potthast, in which an indicator function is computed to estimate the shape of the unknown targets. The OSM has the remarkable feature of being based on an implicit regularization, thus not requiring any regularization parameter tuning and not being limited by underlying approximations. Moreover, similar to other *sampling* methods [9], the spatial discretization of the resulting image is arbitrary and thus not influenced in any way by the measurement configuration. Last but not least, it has been shown in [20] that OSM images encode information on the spatial behavior of the EM properties of the targets, owing to the relation between the OSM indicator and the radiating component of the induced currents. Based on these considerations, it was shown that a U-Net fed with OSM images could be trained to achieve an objective reconstruction of the targets' shape [18] or an estimate of the targets shape and EM properties, provided they belonged to a fixed and known set of values [19].

In this paper, we show how an OSM-informed U-Net can be trained to solve the more general problem of imaging piece-wise homogeneous targets, i.e. retrieve their shape and EM properties, without limiting the possible contrast to a finite set of values as in [19]. To face the increased complexity of such a problem, the following strategies are put into action:

- Different from our previous works where U-Net task consisted in classification problems (binary segmentation [18] or categorical segmentation [19]), the U-Net is herein trained within a pixel-wise regression framework, to allow retrieving a continuous set of values;
- The *a priori* information on the piece-wise nature of the targets is encoded by representing the spatial map of the EM properties distribution to be predicted by the network in terms of the corresponding spatial gradient, which allows to explicitly enforce into the training process the implicitly *sparse* nature of the information to be retrieved. We refer to this map as the *augmented shape*, to recall that it conveys information on both the target's internal and external boundaries and the relative contrast variation with respect to the (known) background medium;

- Finally, a simple post-processing procedure is developed to turn the network's output (i.e. the pixel-wise regression of the gradient's values) into the map of the targets showing both their shape and permittivity.

In the following, the proposed framework is employed to solve the canonical 2D scalar problem (TM polarized fields) in free space. After training the U-Net on simulated data, the resulting model is tested on the Fresnel experimental scattering measurements [21], to provide a performance assessment against this broadly adopted benchmark.

The remainder of the paper is organized as follows. In section 2, the problem is formulated. The physics-assisted DL framework proposed to retrieve the contrast is presented in section 3, wherein each processing step is described. In section 4, implementation details of the U-Net and its training/validation on simulated data are given. Section 5 presents the validation of the overall framework against Fresnel experimental data [21], conclusions follow. Throughout the paper a time-harmonic behavior was supposed and the corresponding time factor  $e^{j\omega t}$  was assumed and dropped.

Note that preliminary results concerned with this work were presented in [22].

## 2. Formulation of the problem

Let  $\Omega$  denote the imaging domain embedded in a homogeneous and lossless medium of relative permittivity  $\varepsilon_b$ , which hosts the cross-section  $\Sigma$  of a collection of possibly overlapping targets invariant along one direction (say the  $z$ -axis). The targets are piece-wise homogeneous. Hence each of them is characterized by a relative dielectric permittivity  $\varepsilon(\mathbf{r})$  and an electric conductivity  $\sigma(\mathbf{r})$ , with  $\mathbf{r} = (x, y)$ . All materials are supposed to be non-magnetic, i.e. the magnetic permeability is everywhere that of vacuum,  $\mu_0$ .

The unknown targets are probed with TM-polarized incident fields  $E_{inc}$ , transmitted by a set of antennas located in  $\mathbf{r}_t \in \Gamma$ , with  $\Gamma$  being a closed curve located in the far-zone of  $\Omega$ . For each transmitter, the interaction between the incident field and the targets gives raise to the scattered field  $E_s$ . The superposition of these two fields becomes the total field  $E = E_{inc} + E_s$  which is measured by a set of receivers that, without any loss of generality, is assumed to be located on  $\Gamma$  as well, with the receiver position being  $\mathbf{r}_s$ .

For each frequency  $f$  belonging to the set of frequencies adopted for the imaging experiment, the overall phenomenon is cast through a Fredholm type integral equation as:

$$E_s(\mathbf{r}_s, \mathbf{r}_t) = \int_{\Omega} G(\mathbf{r}_s, \mathbf{r}') \tau(\mathbf{r}') E(\mathbf{r}', \mathbf{r}_t) d\mathbf{r}', \quad (1)$$

where  $G(\mathbf{r}_s, \mathbf{r}')$  is the Green's function of the assumed homogeneous background medium and  $\tau(\mathbf{r}) = \varepsilon_{eq}(\mathbf{r})/\varepsilon_b - 1$  is the contrast function encoding the properties of the targets.  $\varepsilon_{eq}(\mathbf{r}) = \varepsilon(\mathbf{r}) - j\sigma(\mathbf{r})/\omega\varepsilon_0$  denotes the relative complex permittivity of the targets, with  $j$  being the imaginary unit,  $\omega = 2\pi f$  the pulsation and  $\varepsilon_0$  the dielectric permittivity of vacuum.

The total field  $E$  is defined through another Fredholm integral equation of the first kind as:

$$E(\mathbf{r}, \mathbf{r}_t) = \tau(\mathbf{r}) E_{inc}(\mathbf{r}, \mathbf{r}_t) + \int_{\Omega} G(\mathbf{r}, \mathbf{r}') W(\mathbf{r}', \mathbf{r}_t) d\mathbf{r}' \quad (2)$$

where  $W(\mathbf{r}', \mathbf{r}_t) = \tau(\mathbf{r}') E(\mathbf{r}', \mathbf{r}_t)$  is the contrast source.

The retrieval of the contrast function  $\tau$  from measurements of the fields they scatter is the objective of the ISP. However, due to the smoothing kernel of (1) and the dependence of the total field on  $\tau$ , the problem turns out to be non-linear and ill-posed [6, 7].

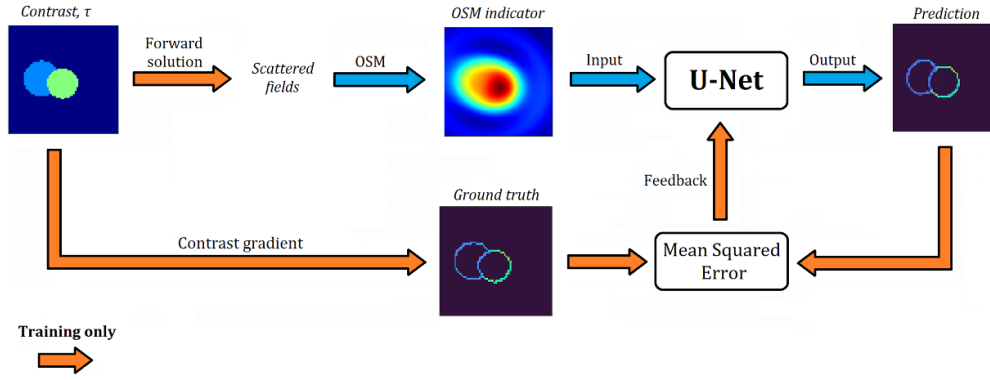


Figure 1. The proposed physics-assisted MWI-DL framework.

### 3. The proposed physics-assisted DL framework

Figure 1 shows the processing flow of the proposed MWI-DL framework, whose steps are detailed in the following.

#### 3.1. The OSM and the domain knowledge it supplies

In the first step of the proposed approach, the measured scattered fields (raw data) are processed with the OSM to obtain a set of images (one for each working frequency).

As most *qualitative* methods [9], OSM provides an estimate of the targets shape through an indicator function, which attains its higher values when evaluated in points belonging to the targets and lower values elsewhere [17]. However, the OSM indicator function is not achieved through the solution of an auxiliary linear ill-posed problem. This entails that unlike other qualitative methods, in OSM there is no need of determining any regularization parameter. This is not a negligible advantage since the estimation of the proper regularizer is a tedious optimization problem [23].

This remarkable OSM property descends from the fact the indicator is built exploiting the reduced scattered field  $E_{\text{red}}$ , which, for each frequency and for each scattered field, is computed as:

$$E_{\text{red}}(r_p, r_t) = \langle E_s(r_s, r_t), G_{(r_s, r_p)} \rangle_{\Gamma} \quad (3)$$

where  $\langle, \rangle$  denotes the scalar product on  $\Gamma$  and  $r_p$  a point of an arbitrary grid sampling the imaging domain  $\Omega$ . As discussed in [20, 24], the reduced field is related to the adjoint solution of an inverse source problem, as such, it is implicitly regularized.

The OSM indicator function is calculated as:

$$\mathcal{I}(r_p) = \|E_{\text{red}}(r_p, r_t)\|_{\Gamma}^2, \quad (4)$$

with  $\| \cdot \|$  denoting the  $L^2$  norm computed on  $\Gamma$ . It is worth noting that the computational burden required to evaluate  $\mathcal{I}$  is negligible, as it is only a scalar product in each sampling point (which is in addition an intrinsically parallelizable process) has to be computed. As such, OSM image formation can be performed in real-time.

In addition to this, as shown in [20], the reduced field is related to the radiating component of the contrast source. Accordingly, the indicator  $\mathcal{I}$  will not only provide an estimation

of the targets support, but it will also bear information on the behavior of their EM properties. In particular, higher permittivity values will correspond to higher intensity values of  $\mathcal{I}$ . However, the relationship between the  $\mathcal{I}$  values and the corresponding ones of the contrast is not straightforward.

### 3.2. The DL architecture

In the second step, the OSM images are fed into the network, which is in charge of estimating the augmented shape.

From the perspective of DL, the process of solving the ISP is driven by data [11]. In particular, assuming a supervised learning procedure, the adopted DL architecture, say  $F_\theta$ , is specialized for the problem at hand through a process called *training*. This is an iterative optimization procedure in which a set of  $N$  training pairs  $(x_n, y_n)$  is exploited to optimize the parameters  $\theta$  that characterize the network against some loss function  $\mathcal{M}$ , i.e.:

$$\hat{\theta} = \arg \min_{\theta} \sum_{n=1}^N \mathcal{M} \{F_\theta(y_n), x_n\} \quad (5)$$

where  $F_\theta(y_n)$  is the prediction made by the architecture corresponding to the ground truth value  $x_n$ .

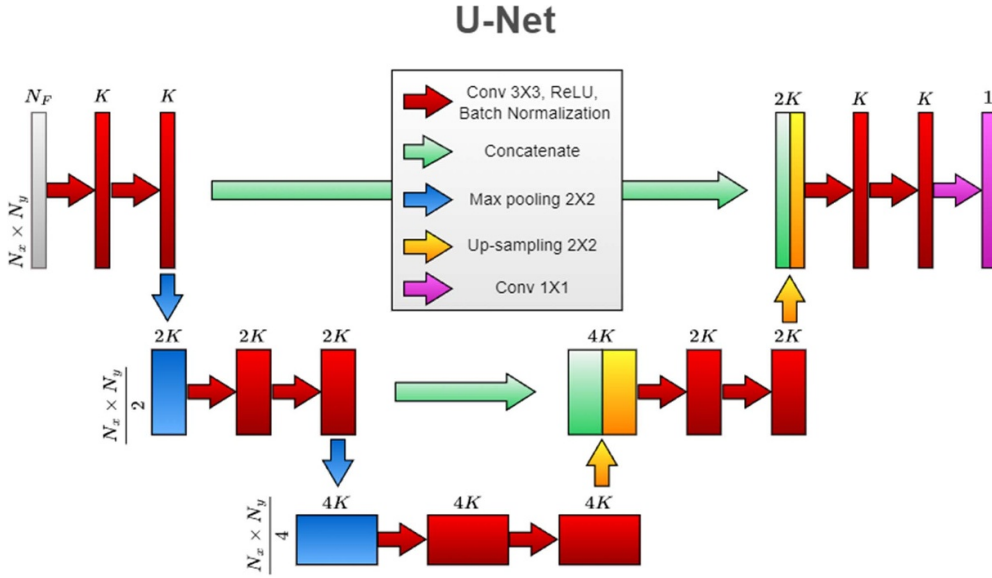
To exploit the above general scheme for the specific problem at hand, the architecture  $F_\theta$ , the loss function  $\mathcal{M}$ , and the training pairs  $(x_n, y_n)$ , have to be defined:

- The network's input  $y_n$  is a stack of one or more OSM images, depending on the number of frequencies
- As far as the ground truth  $x_n$  is concerned, the most straightforward choice would be to define it as an image in which each pixel is associated with the local value of the contrast. However, for piece-wise homogeneous targets, a more efficient way to encode  $\tau$  is to express it through its spatial gradient. Different from the original image, in which all pixels belonging to the target will be different from zero, the gradient only assumes nonzero values at boundaries. As well known, this naturally provides a *sparse* representation of the unknown, which encodes all the required information with a minimal number of non-zero coefficients. Accordingly, the network's output  $\hat{x}_{n,i}$  is the predicted *augmented* shape, i.e. a map of the spatial gradient of the EM properties of the targets. More in detail, the image gradient is computed using the intermediate difference gradient method as:

$$\frac{\partial \tau}{\partial x} = \tau(x+1) - \tau(x) \quad (6)$$

$$\frac{\partial \tau}{\partial y} = \tau(y+1) - \tau(y) \quad (7)$$

and the *augmented shape* fed into the network (as ground truth during training) is obtained as  $\|\nabla \tau\| = \sqrt{(\frac{\partial \tau}{\partial x})^2 + (\frac{\partial \tau}{\partial y})^2}$ . Note the gradient can be computed using different gradient operators, like Sobel or Prewitt [25], but they involve the convolution of the image with a  $3 \times 3$  filter, thus resulting in a less sparse version of the gradient, and therefore less effective for our purposes.



**Figure 2.** U-Net architecture diagram. The network consists of several layers in which the operations depicted with arrows are performed. The size of the matrices is detailed for each layer, the specific values being reported in the text. In our implementation of U-Net,  $K = 32$ .

- The task to be performed by the network is to transform the input OSM image into an image depicting the estimated augmented shape. Such a task can be cast in terms of a pixel value regression, i.e. transform each pixel of the input image into an estimated contrast gradient  $\nabla\tau$  value. To this end, a U-Net architecture [15, 26] is considered, whose specific structure and different processing steps are detailed in figure 2. The U-Net training for the (non-linear) regression task at hand is assessed taking as loss function  $\mathcal{M}$  the mean squared error MSE defined as [13]:

$$MSE = \frac{1}{N_B \cdot N_P} \sum_{n=1}^{N_B} \sum_{i=1}^{N_P} (x_{n,i} - \hat{x}_{n,i})^2 \quad (8)$$

where  $N_B$  is the batch size [13] and  $N_P$  is the total number of pixels per image. Additionally, it is worth noting that U-Net is not necessarily limited to single input images. Hence, when multi-frequency data are available, OSM images for each single frequency data can be supplied stacked together using the U-Net channel dimension [13].

### 3.3. Contrast estimate

The last step of the processing flow is to determine the contrast map from the augmented shape predicted by the network.

For a homogeneous target in free space, such a task is straightforward, as  $\|\nabla\tau\| = \tau$ . Hence, the contrast map in this case could be readily retrieved by assigning to each pixel belonging to



the identified contour the contrast value obtained by averaging the values of  $\|\nabla\tau\|$  estimated by the network.

In the more general case of targets embedded in a homogeneous medium of known or estimate permittivity, the above also applies<sup>5</sup>. Hence, it is possible to extend the above straightforward approach also to nested targets through the following post-processing procedure:

- (i) for each contour, create a separate image having the same size as the original image;
- (ii) for each image, assign a contrast value to the pixel internal to each contour, by averaging the estimated gradient on the contour;
- (iii) sum all the partial images to obtain the final result.

Note that, by means of the above procedure, the superposition of the targets having overlapping supports allows to restore the contrast with respect to the host medium which embeds the targets.

#### 4. Network implementation and optimization

To cope with the 2D canonical ISP in free space at hand, the implementation of the U-Net is carried out optimizing its parameters  $\theta$  with a training set of simulated data similar to the one used in [26]. In particular, the training set consisted of cylinders placed in groups of two with variable size, location and permittivity. However, as opposed to [26], single targets were not considered in the simulations. Also, no profile was allowed to be partially outside of the imaging domain, while target overlapping was permitted. Details of the measurement conditions are listed in table 1.

For the training and assessment, a total set of  $N = 7000$  scattering experiments was simulated. Among them, 85% were used as the training set and 15% as validation set. In particular, for each simulated target,  $N_F = 8$  OSM indicator functions were built using equation (4). The data have been numerically computed using a proprietary forward solver based on Richmond's implementation of the method of moments [27]. The code has been validated against the reference paper for consistency [27]. For reproducibility, the training dataset is publicly available [28].

Accordingly, the input of our U-Net is a stack of  $N_F$  matrices each encoding the  $64 \times 64$  image of the OSM indicator  $\mathcal{I}$  at each frequency. A normalization to  $[0, 1]$  was carried out for each indicator [29].

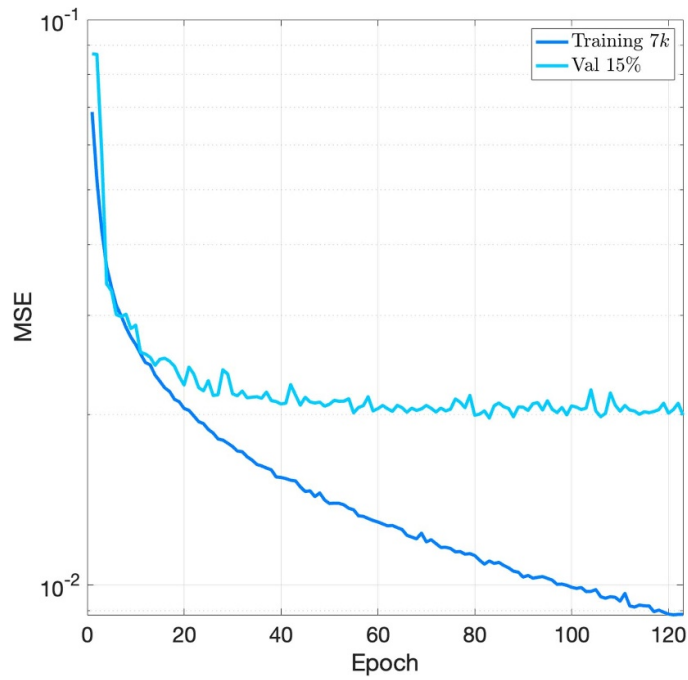
The optimization of the loss function in equation (8) was carried out using Adam optimizer [30], with a learning rate of  $10^{-4}$  and a batch size of 16. An optimal solution is found after several passes through all samples in the training set. A complete pass of the whole training set is known as epoch. A 200 epoch-long training was performed.

The result of the training process is depicted in figure 3, which reports the behavior of the MSE for both the training and validation set along the epochs. As can be seen no overfitting occurs.

<sup>5</sup> Apart for an ambiguity on the signum of the contrast, which can be solved exploiting the different qualitative behavior of the OSM indicator for negative and positive contrasts [20].

**Table 1.** Simulations for training data generation.

Dataset size $N$	7000
Training/Validation Split	85/15
Target count	2
Size of imaging domain	$25 \times 25 \text{ cm}^2$
Image size in pixel, $N_x \times N_y$	$64 \times 64$
Pixel size $x$	0.39 cm
Pixel size $y$	0.39 cm
Background medium	Air
Number of illuminating antennas $N_T$	8
Angular spacing between emitters	$45^\circ$
Number of receiving antennas $N_R$	241
Angular spacing between receivers	$3/2^\circ$
Distance of the source from the center of $\Omega$	167 cm
Distance of the receiver from the center of $\Omega$	167 cm
Number of frequencies $N_F$	8
Frequency range	[2, 9] GHz
Frequency step	1 GHz
Target radius range	[1.2, 5] cm
Target relative permittivity range	[1.3, 3.5]

**Figure 3.** Training procedure of the proposed physics-assisted DL. MSE of the training and validation split with MSE axis in logarithmic scale.

#### 4.1. Performance evaluation metrics

To quantitatively assess the performances of the optimized models, two metrics were used. The first considered metric is the *mean absolute percentage error (MAPE)* [31], which is defined as:

$$MAPE = \frac{100\%}{N_P} \sum_{i=1}^{N_P} \left| \frac{x_i - \hat{x}_i}{x_i} \right|. \quad (9)$$

Although *MAPE* provides an estimation of the performance, it can suffer from weighting down the reported performance as a consequence of the high number of pixels with zero values. For this reason, a modified version where the *MAPE* is only computed over the pixels with positive values in the ground truth was calculated as well:

$$MAPE_{>0} = \frac{100\%}{N_{\bar{P}}} \sum_{i=1}^{N_{\bar{P}}} \left| \frac{x_i - \hat{x}_i}{x_i} \right| \quad \forall i, x_i > 0. \quad (10)$$

While *MAPE* can be interpreted as a performance metric of the qualitative error, i.e. how well the framework retrieves the shapes of the targets, *MAPE*<sub>>0</sub> reports the performance concerning the retrieval of the actual values of  $\|\nabla\tau\|$ .

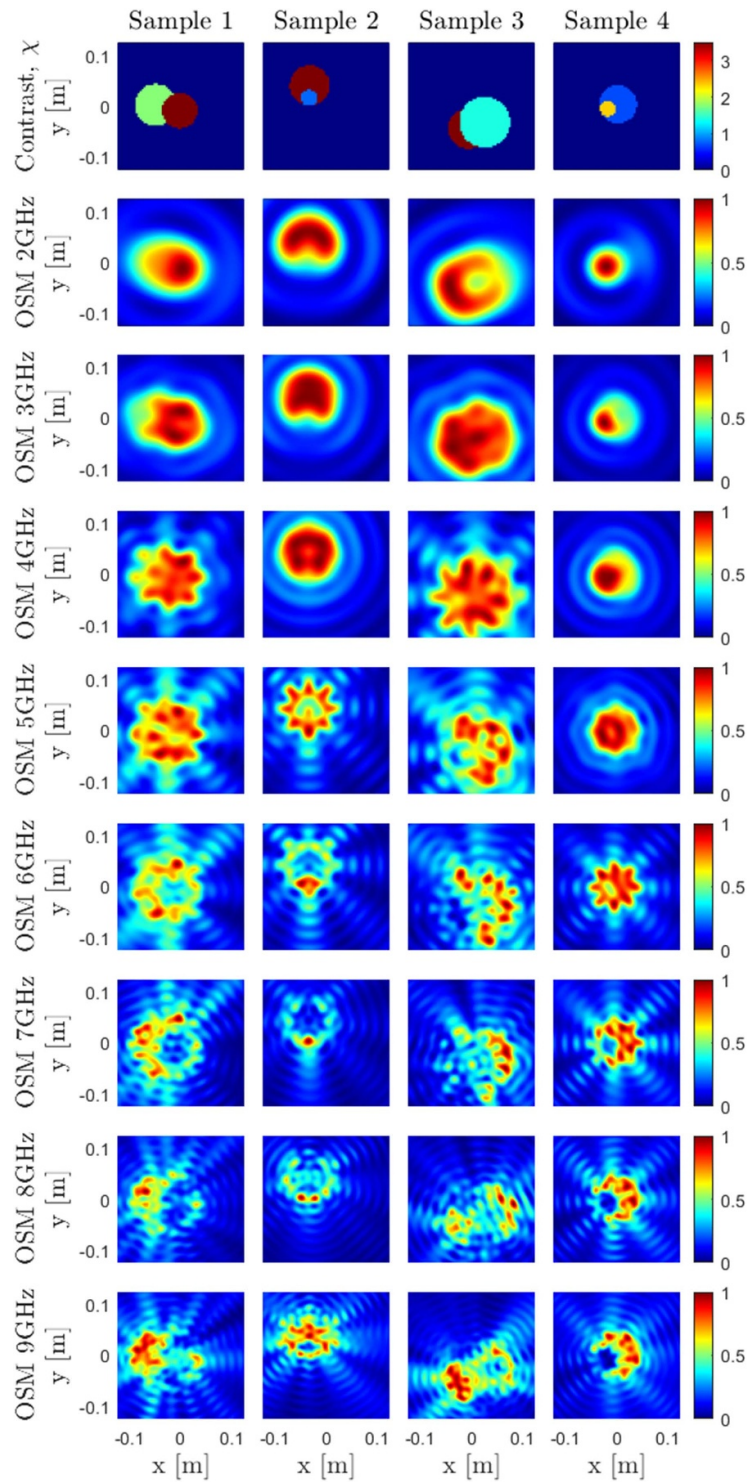
For each sample of the validation set (1050 samples), the metrics were calculated and averaged over the whole validation set. The resulting *MAPE* is 0.94%, which confirms that the trained network is capable of performing satisfactory estimations. On the other hand, when restricting the error on the non-zero pixels the error grows, being *MAPE*<sub>>0</sub> of 13.64%. This is related to the fact the MSE appraises the image as a whole so that the loss value is biased by the background pixels whose number largely exceeds the non-zero pixels.

Four randomly selected samples of the validation set are shown in figure 4, along with the OSM indicator images at the considered frequencies. As can be seen the OSM images visually suggest several properties of the target, but their contour and permittivity are not at all evident. The U-Net predictions are shown in figure 5. These results are consistent with the aforementioned performance metrics: U-Net not only does successfully find the boundaries between the targets and the background but also the ones between the two targets. When it comes to the quantitative gradient values, the accuracy is lower.

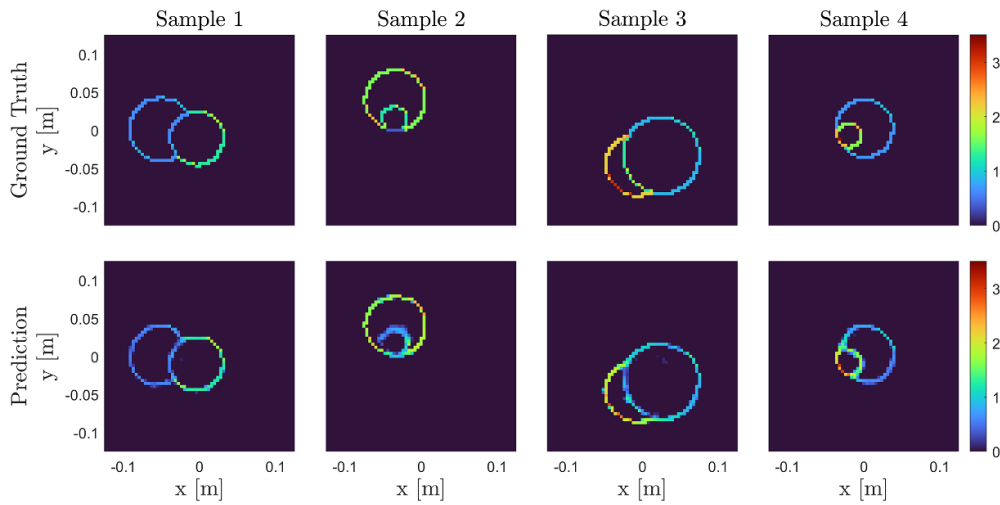
## 5. Assessment of the proposed physics-assisted framework against Fresnel experimental data

To show the capability of the proposed framework to retrieve the contrast map of piece-wise homogeneous targets, the widely adopted benchmark data provided by the Institut Fresnel [21] have been considered.

The Fresnel targets and the OSM indicator maps for each frequency are depicted in the first column of figure 6. While the results of the analysis are reported in figure 7. In particular, the first row reports the expected augmented shape, the second row the augmented shape retrieved by the U-Net, the third row the ground truth permittivity assuming the average values given in the database, while the last row shows the permittivity map estimated using the post-processing procedure.



**Figure 4.** Four randomly selected samples from the validation dataset. The first row depicts the target's contrast map while the other rows report the OSM indicator at each frequency and for each target.



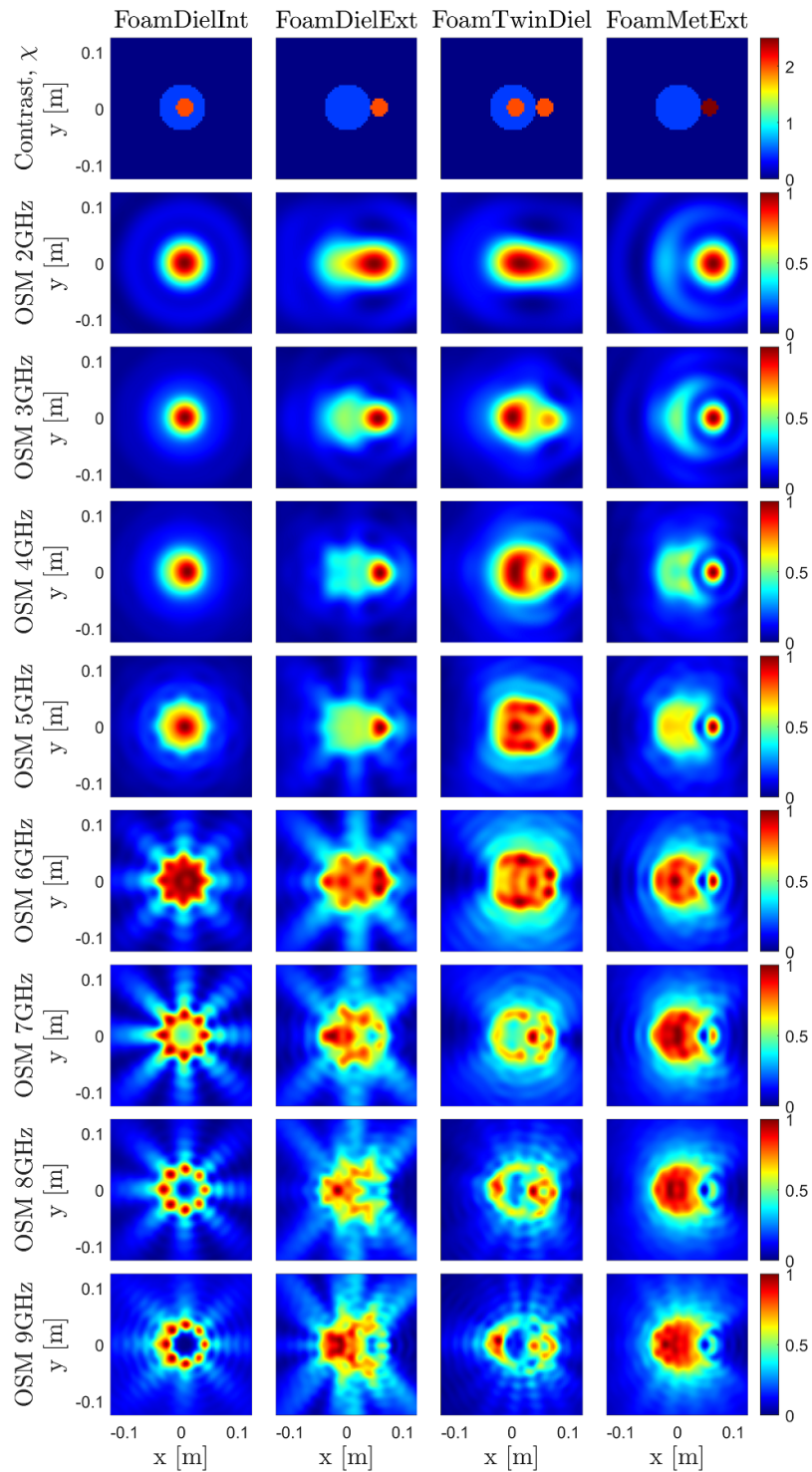
**Figure 5.** Augmented shapes predicted by the network for the four validation samples. One sample per row, with the first column representing the ground truth and the second presenting the prediction made by U-Net.

As a first comment, it is worth to remark that this validation has been carried out using the optimized U-Net resulting from the training process described in the previous section, without retraining. This is a noticeable aspect, since the Fresnel experimental data and targets are to some extent different from those considered in training the U-Net. In particular:

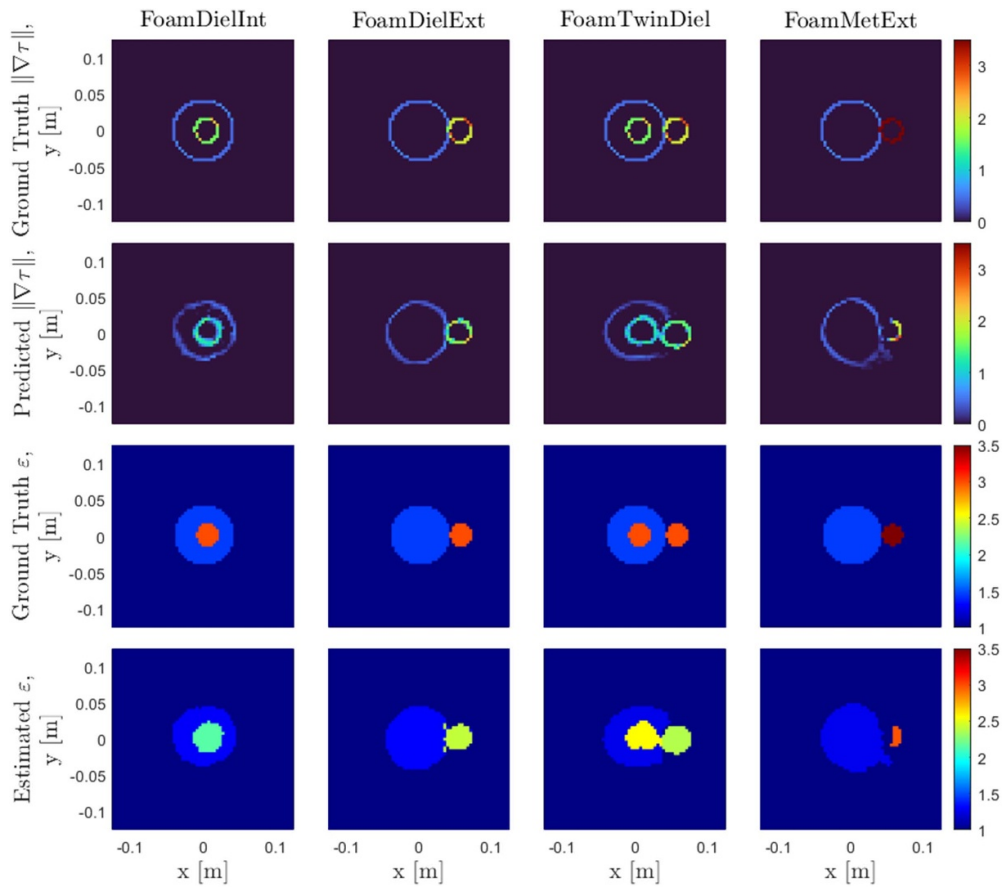
- the measurement configuration is different, since the Fresnel data are collected within an aspect-limited configuration, whereas the configuration used in section 4 is full aspect;
- in one of the Fresnel datasets, three targets are present, while only up to two targets were considered in the training;
- only dielectric targets were considered in the training, while one of the Fresnel targets includes a metallic object;
- the dielectric materials employed to build the Fresnel targets are not exactly lossless, opposite to the targets considered in the training.

As can be seen, the developed framework, while not optimized for the considered experimental data, successfully resolves the targets and provides quite accurate reconstructions of their augmented shapes.

As far as the estimation of the permittivity values is concerned, from figure 7, it appears that the retrieved values are quite close to the actual values for the low-permittivity foam targets, while they are quite different for the plastic targets. More in detail, as reported in table 2, where also the MAPE computed for each material is reported, the retrieved values are always close to the lower expected actual value for each material and in general appear to be underestimated. This is due to the fact the estimated augmented shape is a blurred version of the ground truth, so that the estimated gradient value is averaged on a larger number of pixels than the ground truth.



**Figure 6.** The Fresnel targets considered for the validation of the proposed framework. The first row depicts the target's contrast map while the other rows report the OSM indicator at each frequency and for each target.



**Figure 7.** The results of the experimental validation. For each target, the first row depicts the augmented shape ground truth. The second row represents the prediction made by U-Net. The third row represents the ground truth assuming the average values of the targets permittivity given in the database. The last row shows the permittivity map retrieved by the framework after the post-processing.

**Table 2.** Estimated permittivity values.

	Foam ( $1.45 \pm 0.15$ )	$MAPE_{\text{Foam}}$	Plastic ( $3 \pm 0.3$ )	$MAPE_{\text{Plast}}$
FoamDielInt	1.3	10%	2.14	29%
FoamDielExt	1.31	10%	2.41	20%
FoamTwinDiel	1.26	13%	2.56; 2.38	15%; 21%
FoamMetExt	1.27	12%	3	—

## 6. Conclusion

This work presents a MWI framework for real-time and user-independent imaging of piece-wise homogeneous targets. Besides the methodological interest, this class of targets is relevant in most applications, wherein EM properties of the targets of interest are indeed piece-wise

(like in non-destructive testing) or can be well approximated by average values (like in bio-medical imaging).

The core of the proposed framework is the efficient encoding of the *a priori* information on the piece-wise target's nature by means of their *augmented shape*, i.e., the amplitude of the spatial gradient of the contrast. Accordingly, the approach is implemented by training a U-Net CNN to retrieve this quantity, which embeds the information on both the shape and the EM properties of the targets. Then, the predicted augmented shape is processed by means of a simple deterministic procedure to turn it into a map of the targets' permittivity.

The network is trained by exploiting a physics-assisted approach in which domain knowledge is supplied in the form of OSM images at multiple frequencies. Such a qualitative imaging technique is a convenient way to pre-process the raw data, thanks to its capability to form the image in real-time and without any supervision. Moreover, these images result from the back-projection of the data from the measurement domain onto the imaging domain, using the adjoint operator. As such, they directly represent the information embedded into the data in the imaging domain, and they are not the outcome of an inversion process prone to the choice of a regularization parameter. For this reason, considering multiple frequencies allows us to include all the information embedded in the measured data in the learning process. In particular, including high-frequency images in the network's learning process is useful, even if they appear poor in terms of targets' reconstructions, since high-frequency data may contain pieces of information that are not present in the low-frequency data (e.g. in terms of details at a finer spatial resolution). Finally, the adopted physics-assisted approach allows the U-Net to manipulate images and transform them into the predicted augmented shape, taking advantage of its demonstrated capability to effectively deal with this kind of inputs.

The framework has been validated with the experimental data from the Fresnel Institut concerned with inhomogeneous targets [21]. The results achieved with this widely adopted benchmark showed the overall capability of the proposed framework to perform the task and operate in cases different from the specific conditions for which the U-Net was trained. In particular, while the network was trained with purely lossless dielectric targets, the overall framework also works successfully when dealing with slightly lossy dielectric targets and metallic targets. On the other hand, it can be expected that dealing with dielectric targets with larger losses would require including those cases in the training to preserve comparable performances. Similarly, while in this work only circular cylinders were considered as targets, the approach is fully general and can be applied to targets having other shapes, provided a suitable training set is implemented to consider those different profiles.

Finally, despite the positive results, there is still room for improvement, especially regarding the quantitative estimation of the permittivity values is concerned. Future research will address this issue as well as the application of the framework to more complex scenarios.

### Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://zenodo.org/records/8314615>.

### Acknowledgments

This research was partially supported by the EMERALD project funded from the European Union's Horizon 2020 under the Marie Skłodowska-Curie Grant Agreement No. 764479.



## ORCID iDs

Álvaro Yago Ruiz  <https://orcid.org/0000-0001-9390-9368>

Maria Nikolic Stevanovic  <https://orcid.org/0000-0002-9840-0322>

Marta Cavagnaro  <https://orcid.org/0000-0002-7624-1113>

Lorenzo Crocco  <https://orcid.org/0000-0003-2498-3219>

## References

- [1] Crocco L, Karanasiou I, James M L and Conceição R C 2018 *Emerging Electromagnetic Technologies for Brain Diseases Diagnostics, Monitoring and Therapy* (Springer)
- [2] Conceição R C et al 2016 *An Introduction to Microwave Imaging for Breast Cancer Detection* (Springer) (<https://doi.org/10.1007/978-3-319-27866-7>)
- [3] Turk A S, Hocaoglu K A and Vertiy A A 2011 *Subsurface Sensing* vol 197 (Wiley)
- [4] Vipiana F, Crocco L and LoVetri J 2020 Electromagnetic imaging and sensing for food quality and safety assessment [guest editorial] *IEEE Antennas Propag. Mag.* **62** 16–17
- [5] Amin M G 2017 *Through-the-Wall Radar Imaging* (CRC Press)
- [6] Colton D and Kress R 1992 *Inverse Acoustic and Electromagnetic Scattering Theory* (Springer)
- [7] Pastorino M 2010 *Microwave Imaging* vol 208 (Wiley)
- [8] Chen X 2018 *Computational Methods for Electromagnetic Inverse Scattering* (Wiley)
- [9] Cakoni F and Colton D L 2014 *A Qualitative Approach to Inverse Scattering Theory* vol 767 (Springer)
- [10] Bellman R 1966 Dynamic programming *Science* **153** 34–37
- [11] Chen X, Wei Z, Li M and Rocca P 2020 A review of deep learning approaches for inverse scattering problems (invited review) *Prog. Electromagn. Res.* **167** 67–81
- [12] Salucci M, Arrebola M, Shan T and Li M 2022 Artificial intelligence: new frontiers in real-time inverse scattering and electromagnetic imaging *IEEE Trans. Antennas Propag.* **70** 6349–64
- [13] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (MIT Press)
- [14] Xu S, Wu H and Bie R 2018 CXNet-m1: anomaly detection on chest x-rays with image-based deep learning *IEEE Access* **7** 4466–77
- [15] Ronneberger O, Fischer P and Brox T 2015 U-net: convolutional networks for biomedical image segmentation *Int. Conf. on Medical Image Computing and Computer-Assisted Intervention* (Springer) pp 234–41
- [16] Siddique N, Paheding S, Elkin C P and Devabhaktuni V 2021 U-net and its variants for medical image segmentation: a review of theory and applications *IEEE Access* **9** 82031–57
- [17] Potthast R 2010 A study on orthogonality sampling *Inverse Problems* **26** 074015
- [18] Ruiz A Y, Cavagnaro M and Crocco L 2022 A physics-assisted deep learning microwave imaging framework for real-time shape reconstruction of unknown targets *IEEE Trans. Antennas Propag.* **70** 1
- [19] Yago Ruiz A, Cavagnaro M and Crocco L 2023 An effective framework for deep-learning-enhanced quantitative microwave imaging and its potential for medical applications *Sensors* **23** 643
- [20] Bevacqua M T, Isernia T, Palmeri R, Akinci M N and Crocco L 2020 Physical insight unveils new imaging capabilities of orthogonality sampling method *IEEE Trans. Antennas Propag.* **68** 4014–21
- [21] Geffrin J M, Sabouroux P and Eyraud C 2005 Free space experimental scattering database continuation: experimental set-up and measurement precision *Inverse Problems* **21** S117
- [22] Yago Ruiz A, Stevanovic M N, Cavagnaro M and Crocco L 2022 A deep learning architecture for augmented shape reconstruction via microwave imaging 2022 *16th European Conf. on Antennas and Propagation (EuCAP)* (IEEE)
- [23] Benning M and Burger M 2018 Modern regularization methods for inverse problems *Acta Numer.* **27** 1–111
- [24] Bevacqua M T and Palmeri R 2021 Isernia T and Crocco L A simple procedure to design virtual experiments for microwave inverse scattering *IEEE Trans. Antennas Propag.* **69** 8652–63
- [25] Castleman K R 1996 *Digital Image Processing* (Prentice Hall Press)
- [26] Wei Z and Chen X 2018 Deep-learning schemes for full-wave nonlinear inverse scattering problems *IEEE Trans. Geosci. Remote Sens.* **57** 1849–60

- [27] Richmond J 1965 Scattering by a dielectric cylinder of arbitrary cross section shape *IEEE Trans. Antennas Propag.* **13** 334–41
- [28] Alvaro Y R, Nikolic M S, Cavagnaro M and Crocco L 2023 Microwave imaging dataset (available at: <https://zenodo.org/records/8314615>)
- [29] Ioffe S and Szegedy C 2015 Batch normalization: Accelerating deep network training by reducing internal covariate shift *Int. Conference on Machine Learning* (PMLR) pp 448–56
- [30] Shalev-Shwartz S and Ben-David S 2014 *Stochastic Gradient Descent* (Cambridge University Press) pp 150–66
- [31] de Myttenaere A, Golden B, Le Grand B and Rossi F 2016 Mean absolute percentage error for regression models *Neurocomputing* **192** 38–48