Deep Learning-Enhanced Qualitative Microwave Imaging: Rationale and Initial Assessment

Alvaro Yago*§, Marta Cavagnaro*§, Lorenzo Crocco*,
*CNR-IREA National Research Council of Italy, Institute for Electromagnetic Sensing of Environment,
Napoli, Italy, crocco.l@irea.cnr.it yago.a@irea.cnr.it*
§DIET, Sapienza University of Rome, Rome, Italy, marta.cavagnaro@uniroma1.it

Abstract—In this paper, an innovative approach to microwave imaging that combines qualitative imaging and deep learning is presented. The goal is to set a framework for a reliable and user-independent retrieval of the shapes of unknown targets. To this end, the proposed approach exploits an inversion technique known as orthogonality sampling method, which is capable of providing a qualitative estimation of the shape of targets in real-time. The output of the qualitive inversion is processed by a deep learning fully convolutional network called U-Net. U-Net automatically generates binary masks depicting the geometrical properties of the targets, i.e., separates the scattering objects (foreground) from the background. A quantitative assessment of the performance of the processing framework is provided with simulated data to demonstrate the capabilities of the proposed approach.

I. INTRODUCTION

Microwave imaging (MWI) technology enables accurate inspection and exploration of unknown scenarios which are not directly accessible. As such, it is relevant to several applications, like biomedical imaging [1], [2], subsurface sensing [3], food security monitoring [4] or through-wall imaging [5].

The mathematical problem underlying MWI is a non-linear and ill-posed inverse scattering problem. Many techniques have been proposed in the literature to achieve reliable and accurate solutions [6]. Among them, qualitative imaging approaches [7], which aim at recovering just presence, position and shape of the unknown targets from the knowledge of the field they scatter, are worth to be mentioned. As a matter of fact, in many practical cases, retrieving the morphological information of a target is a sufficient piece of information, and qualitative methods pursue this goal smartly, without introducing approximations, and in real-time.

The typical outcome of qualitative imaging methods is the continuous map of an indicator function, which attains large values where the target is located and low values elsewhere. While this can be sufficient to provide a visual information about the targets, it does not provide a clear indication of the actual morphological properties. This is particularly important when an objective characterization of the targets is needed or when the morphological information is exploited as prior knowledge for a quantitative processing aimed at achieving the estimation of the electromagnetic features of the target. Such an issue is typically tackled by resorting to hard-thresholding the continuous map at some ad hoc defined values. This

however, is a case-dependent approach, often biased by the user's expertise.

In this paper, we present an innovative approach to overcome the above problem, which consists in enhancing qualitative imaging with an automated classification based on deep learning (DL), to enable a user-independent procedure.

In recent years, machine learning, and especially its sub-field of deep learning, proved to be extremely successful in solving complex problems and attracted the attention of many researchers who have explored their potential in the solution of inverse scattering problems [8]–[10]. As such, DL appears as a suitable candidate to address the problem of achieving objective results via qualitative MWI.

More in detail, the proposed approach exploits convolutional neural networks (CNNs) [11] as a backbone of the DL model. CNNs have been extensively used for classification purposes, in different fields involving imagery [8]–[10]. As a matter of fact, the problem at hand can be seen as a pixel-wise classification (segmentation), where the end goal is the location of structures and their boundaries, instead of a regular image classification problem.

As far as the qualitative imaging approach is concerned, the proposed processing framework exploits the orthogonality sampling method (OSM) [12]. In OSM, the evaluation of the indicator function does not require the explicit solution of an inverse problem, so that there is no need for determining a regularization parameter. Such a circumstance is a particularly attractive in view of the implementation of an automated processing chain. Moreover, OSM can be applied to a huge variety of measurement configurations (single-view, multifrequency, multiview, and a combination of them) and provides imaging results in real-time.

In the following, the proposed automated inversion framework is detailed and its assessment in a canonical 2D scalar problem (TM polarized fields) in free space is presented.

II. FORMULATION OF THE PROBLEM

The problem which is faced is the retrieval of the shape of a collection of targets which extend up to infinity along one direction (say the z-axis), and whose cross-section Σ is enclosed in the imaging domain Ω, embedded in a homogeneous and lossless medium of permittivity ε_b. Each target is characterized by a dielectric permittivity $\varepsilon(\underline{r})$ and an electric conductivity

$\sigma(\underline{r})$, with $\underline{r} = (x, y)$. All materials are supposed to be non-magnetic, i.e., the magnetic permeability is everywhere that of vacuum, μ_0.

The unknown targets are probed with a set of incident fields transmitted by a set of antennas located in $\underline{r}_t \in \Gamma$, with Γ being a closed curve located in the far-zone of Ω and $\underline{r}_t$ being the position of the generic antenna. Without any loss of generality, let us assume that the scattered fields are measured by receiving antennas located in $\underline{r}_s \in \Gamma$.

III. OUTLINE OF THE PROPOSED APPROACH

A. Qualitative Inversion: Orthogonality Sampling Method

The OSM provides an approximation of the geometrical properties of the unknown targets by plotting an indicator function $I(\underline{r}_p)$ on an arbitrary grid of points $\underline{r}_p = (x_p, y_p)$ sampling the imaging domain Ω [12].

To build the indicator function, the *reduced field* is first computed for each scattered field due to an incident field, as:

$$E_{red}(\underline{r}_t, \underline{r}_p) = <E_s(\underline{r}_s, \underline{r}_t), G_b(\underline{r}_s, \underline{r}_p)>_\Gamma \quad (1)$$

where $E_s(\underline{r}_s, \underline{r}_t)$ is the scattered field measured in $\underline{r}_s \in \Gamma$ due to the incident field radiated from $\underline{r}_t \in \Gamma$ and $G_b(\underline{r}_s, \underline{r}_p)$ is the Green's function of the assumed homogeneous background medium for the sampling point $\underline{r}_p \in \Omega$ and computed in $\underline{r}_s \in \Gamma$. $<, >$ denotes the scalar product.

The indicator function is then obtained as:

$$I(\underline{r}_p) = ||E_{red}(\underline{r}_p)||_\Gamma^2, \quad (2)$$

with $|| \; ||$ denoting the L^2 norm computed on Γ.

By evaluating the orthogonality between the scattered field pattern and the Green's function, the indicator retrieves the geometry by assuming large values where the target is present. It is important to notice that the evaluation of I does not require an explicit inversion process, so that there is no need of computing any regularization parameter. Moreover, the image formation is carried out in real-time.

It is worth noting that the above formulation of the OSM is valid for single frequency data. When data at multiple frequencies are available, a collection of images can be generated, one for each frequency. As an alternative, not exploited in the following, a multifrequency indicator can be built by adding the amplitudes of the indicators computed at each frequency.

B. Deep Learning

In segmentation tasks, a special type of CNN, called Fully Convolutional Network (FCN) [13], is normally employed. This is due to its inner structure, which makes a FCN capable of retrieving modified versions of an input image. In particular, FCN allows to retrieve a direct segmentation map by means of an image-to-image mapping. Accordingly, a peculiar implementation of FCN, called U-Net [14], [15] is explored for our problem. The selected network architecture was originally designed for image segmentation in cases where the number of images for training was scarce, like in medical applications.

One important concept in Deep Learning is indeed the learning process, called training, where the parameters of the network are adjusted iteratively as part of an optimization of an objective function. As detailed in the next Section, a training set made of circular cylinders similar to [15] was exploited in this paper. It is worth noting that such training data were generated via numerical simulations. As these simulations allow us to generate virtually whatever number of examples needed, the data availability limitation [16] is not an issue in the present problem.

Different from the implementation suggested in [15], where U-Net was trained to retrieve the original permittivities from the information provided by an inversion technique, in the case at hand, U-Net was trained for the generation of binary masks, so that the user can objectively discriminate the foreground (the profiles) from the background in the output image.

When training CNNs on regular images, the information is usually separated in three channels, one for each RGB color component. Such a feature is herein exploited taking advantage of the ability of OSM to provide different images of the target when processing data at different frequencies. In particular, by indicating with N_F the number of frequencies at which the scattering experiment is carried out, the input of U-Net is given by a stack of N_F images with the values of the OSM indicator function I for each selected frequency.

Fig. 1 shows the complete processing scheme. In particular, binary cross-entropy [17] is employed as objective function for the binary segmentation. By minimizing this objective function during the training, a maximization of the probability of correctly labeling the pixels of an input image is enforced.

One interesting strategy for boosting the performance of a network when the available data is limited is data augmentation. Data augmentation is a technique for artificially increasing the size of a small training set by generating new samples out of the ones present on the actual training set [18]. An additional option is dropout [19], also known as implicit data augmentation, where more data are created during the training by randomly deactivating connections in the inner structure of the network. From the perspective of solving an ill-posed problem, dropout acts as a regularizer.

IV. NUMERICAL ASSESSMENT

A. Training and Test Data Generation

Similar to what is done in [15], the training of the network was carried out by simulating a number of scattering experiments involving combinations of lossless homogeneous circular cylinders enclosed in the imaged domain. In particular, cylinders placed in groups of two or three with variable size, location and permittivity have been considered for the simulations. Additionally, no profile was allowed to be partially outside of the imaging domain and some profiles could overlap; more details of the setup conditions are listed in Tab. I. Note that the adopted measurement configuration corresponds to the aspect limited one [20], in the view of the assessment of the approach with the targets available in those experiments.
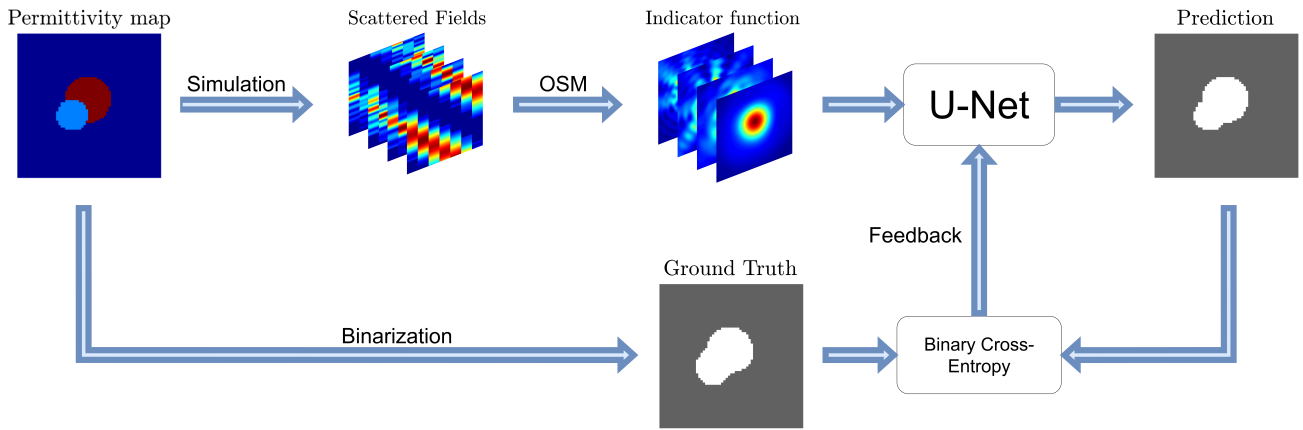
Fig. 1. Deep learning scheme for the automatic generation of binary masks. From the scattered fields generated through simulation, the OSM indicator map is built and feed into U-Net, which automatically retrieves the predicted binary mask for those inputs. When U-Net is on the training stage, the feedback loop is used to optimize the inner parameters of the network by minimizing the binary cross-entropy between the ground truth and the prediction of the samples.

In Tab. I, SNR indicates the level of the Gaussian noise that is added to the simulated scattering matrices. For each computed scattered field, the OSM indicator function is built according to Eq. 2.

For the training and the performance assessment, a total set of 2000 scattering experiments was simulated. Among them, 95% were used as the training set and 5% as the test set. For each experiment, an input data-ground truth pair for the network training was generated. The input is a stack of 4 images 64×64 with the obtained OSM indicator maps, wherein 4 refers to the 4 available frequencies. The ground truth is made by 64×64 resulting from the binarization of the simulated profile. In agreement with the use of the Softmax function to implement binary cross-entropy, the binary mask is splitted into a pair of images, each separately encoding foreground and background pixels.

Due to the simple nature of the chosen profiles, no further data augmentation was considered besides the dropout layers of the original implementation [14] with a dropout rate of 0.5.

The optimization of the objective function was carried out for 200 epochs, using Adam optimizer with a learning rate of 10^{-5} and a batch size of 16 [21].

B. Results and Performance Evaluation

When training is complete, U-Net can be used to make predictions on new data. Accordingly, the 100 samples of the test set were fed into the network and their corresponding predictions were retrieved. An example of the achieved imaging results is given in Fig. 2, for 4, randomly selected, samples.

To assess the performance of the U-Net, several metrics were calculated on each sample, using predictions and ground truth, and averaged over the whole test set.

The first metric which was considered is *accuracy* (ACC), which is defined as:

$$ACC = \frac{TP + TN}{TP + TN + FP + FN} \qquad (3)$$

TABLE I
SIMULATIONS FOR TRAINING DATA GENERATION

Size of imaging domain	25×25 cm^2
Image size in pixel	64×64
Background medium	Free Space
Number of illuminating antennas	8
Angular spacing between emitters	45°
Number of receiving antennas	241
Angular spacing between each emitter and the relevant first receiver	60°
Angular spacing between receivers	1°
Source - object center distance	167 cm
Receiver - object center distance	167 cm
Profile radius range	$[1.2, 5]$ cm
Profile permittivity range	$[1.1, 3.5]$ F·m^{-1}
Number of frequencies	4
Frequencies step	2 GHz
Frequency range	$[2, 8]$ GHz
SNR	70 dB

where TP, TN, FP, FN are true positives, true negatives, false positives and false negatives, respectively. In the problem at hand, TP was considered to be the count of all pixels correctly labeled as foreground, whereas TN was considered the count of all pixels correctly labeled as background.

The average ACC for the considered test set is 98.2%. However, such an excellent outcome is influenced by the fact that accuracy can be misleading if data are unbalanced, that is when there are considerably more pixels of one class than the others. In the test set, an average proportion of foreground pixels around 11.8% of the total, with a minimum of 3.1% and a maximum of 25.1% is found. Such an unbalance, results in TN and FN being much larger than TP and FP, hence making the formers dominate Eq. 3. Accordingly, U-Net showed a very optimistic average ACC.

As a second metric, the *dice similarity coefficient* (DCS), which is popular in medical image segmentation [22], was employed. DCS is defined as:

$$DSC = \frac{2 \cdot TP}{2 \cdot TP + FP + FN} \qquad (4)$$

An average DSC of 92.3% was obtained. While this is a lower score than ACC, it is more meaningful in this context. As a matter of fact, DSC does not rely on TN to give an estimation of the performance, therefore it is more robust against data unbalance. Furthermore, it doubles the contribution of TP. It is worth to mention that while *Precision* and *Recall* (*Sensitivity*) concepts are frequently used in the assessment of the performance, DSC is the harmonic mean of both metrics and hence gives a more general insight into the performance of the model.

Finally, motivated by some limitations of the DSC [23], the *Matthews correlation coefficient* (MCC)[24] was calculated as well:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{DM}} \qquad (5)$$

where $DM = (TP + FP)(TP + FN)(TN + FP)(TN + FN)$. The implemented U-Net achieved an average MCC of 91.1% i.e., close to the DSC value, thus showing reliability of the two considered metrics.

V. CONCLUSION

In this work, an innovative MWI framework for effective, objective and automatic retrieval of the shapes of unknown targets from measured scattered fields has been presented. To this end, the processing approach combines qualitative imaging and deep learning.

As well known, qualitative imaging methods are powerful inverse scattering tools, as they provide morphological information of unknown targets in real-time. However, their outcome is a continuous map which requires further processing to provide an objective output in which the boundaries of the target are unequivocally defined. Although hard-thresholding the continuous map may provide clear boundaries, the processing itself can easily biased in different ways. To overcome these limitations, the use of a deep learning architecture called U-Net has been proposed to enhance the qualitative imaging outcome of a technique called orthogonality sampling, by processing its output and automatically generating binary masks depicting the geometrical properties of the target.

In particular, the capability of OSM to provide images at multiple frequencies has been exploited to increase the input information supplied to U-Net. Though a reduced number of frequencies were considered in this paper, other models based on the same neural network could be designed to work with a higher number of OSM images from multiple frequencies.

The performance of the processing framework has been tested with simulated data and quantitatively assessed by means of several metrics showing, on average, excellent capabilities. The next step will be a performance assessment on experimental data from the Fresnel database, a popular benchmark for microwave imaging method [20], as a further proof of the achievable capabilities.

REFERENCES

[1] L. Crocco, I. Karanasiou, M. L. James, and R. C. Conceição, *Emerging Electromagnetic Technologies for Brain Diseases Diagnostics, Monitoring and Therapy*. Springer, 2018.

[2] R. C. Conceição, J. J. Mohr, M. O'Halloran, *et al.*, *An introduction to microwave imaging for breast cancer detection*, 2016.

[3] A. S. Turk, K. A. Hocaoglu, and A. A. Vertiy, *Subsurface sensing*. John Wiley & Sons, 2011, vol. 197.

[4] F. Vipiana, L. Crocco, and J. LoVetri, "Electromagnetic imaging and sensing for food quality and safety assessment [guest editorial]," *IEEE Antennas and Propagation Magazine*, vol. 62, no. 5, pp. 16–17, 2020.

[5] M. G. Amin, *Through-the-wall radar imaging*. CRC press, 2017.

[6] M. Pastorino, *Microwave imaging*. John Wiley & Sons, 2010, vol. 208.

[7] F. Cakoni and D. L. Colton, *A qualitative approach to inverse scattering theory*. Springer, 2014, vol. 767.

[8] S. Xu, H. Wu, and R. Bie, "Cxnet-m1: Anomaly detection on chest x-rays with image-based deep learning," *IEEE Access*, vol. 7, pp. 4466–4477, 2018.

[9] J. Liu, Y. Pan, M. Li, Z. Chen, L. Tang, C. Lu, and J. Wang, "Applications of deep learning to mri images: A survey," *Big Data Mining and Analytics*, vol. 1, no. 1, pp. 1–18, 2018.

[10] B. Gerazov and R. C. Conceicao, "Deep learning for tumour classification in homogeneous breast tissue in medical microwave imaging," in *IEEE EUROCON 2017-17th International Conference on Smart Technologies*, IEEE, 2017, pp. 564–569.

[11] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.

[12] M. T. Bevacqua, T. Isernia, R. Palmeri, M. N. Akinci, and L. Crocco, "Physical insight unveils new imaging capabilities of orthogonality sampling method," *IEEE Transactions on Antennas and Propagation*, vol. 68, no. 5, pp. 4014–4021, 2020.

[13] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[14] 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.

[15] Z. Wei and X. Chen, "Deep-learning schemes for full-wave nonlinear inverse scattering problems," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 4, pp. 1849–1860, 2018.

[16] S. Russell and P. Norvig, "Artificial intelligence: A modern approach (global 3rd edition)," *Essex: Pearson*, pp. 122–125, 2016.

[17] M. Rimer and T. Martinez, "Classification-based objective functions," *Machine Learning*, vol. 63, no. 2, pp. 183–205, 2006.

[18] L. Perez and J. Wang, "The effectiveness of data augmentation in image classification using deep learning," *arXiv preprint arXiv:1712.04621*, 2017.

[19] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *The journal of machine learning research*, vol. 15, no. 1, pp. 1929–1958, 2014.

[20] J.-M. Geffrin, P. Sabouroux, and C. Eyraud, "Free space experimental scattering database continuation: Experimental set-up and measurement precision," *inverse Problems*, vol. 21, no. 6, S117, 2005.

[21] S. Shalev-Shwartz and S. Ben-David, "Stochastic gradient descent," in *Understanding Machine Learning: From Theory to Algorithms*. Cambridge University Press, 2014, 150–166. DOI: 10 . 1017 / CBO9781107298019.015.
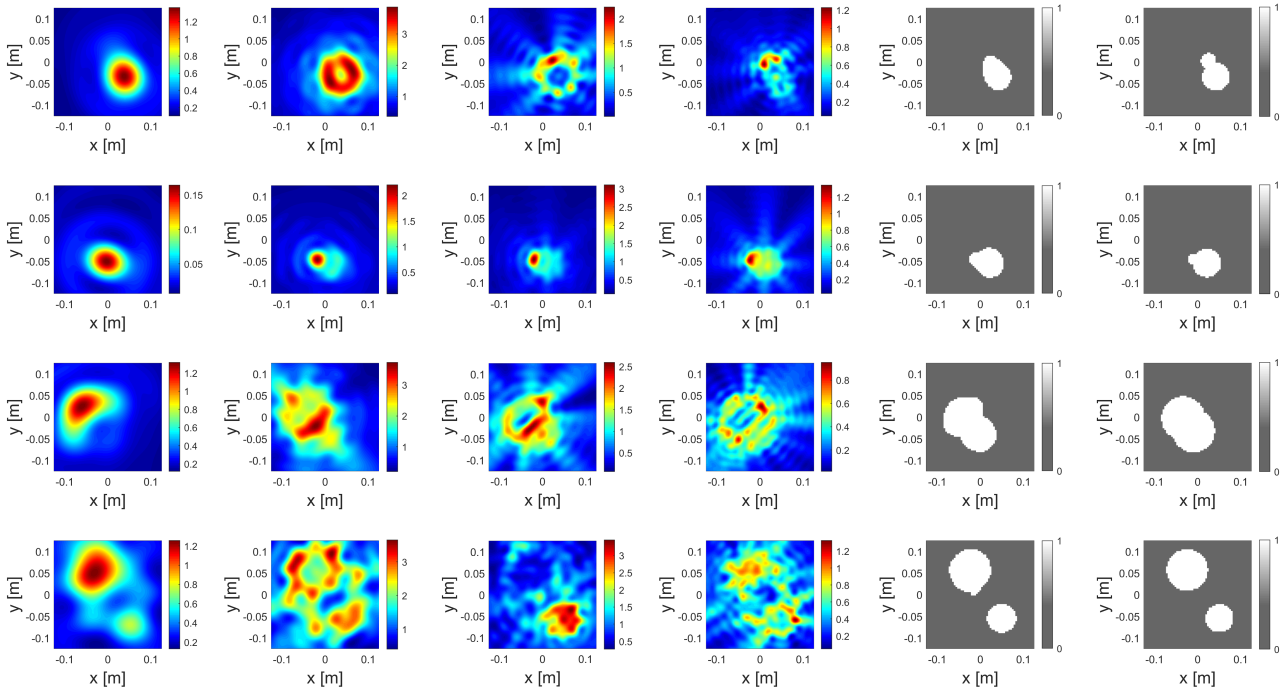
Fig. 2. Imaging results of 4 test samples. One sample per row, with the first 4 columns representing the Indicator function from measured scattered fields at 2, 4, 6 and 8 GHz. The fifth column gves the prediction made by U-Net, while the last column shows the ground truth binary mask used to evaluate the scattering data.

[22] A. A. Taha and A. Hanbury, "Metrics for evaluating 3d medical image segmentation: Analysis, selection, and tool," *BMC medical imaging*, vol. 15, no. 1, p. 29, 2015.

[23] D. Chicco and G. Jurman, "The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation," *BMC genomics*, vol. 21, no. 1, p. 6, 2020.

[24] B. W. Matthews, "Comparison of the predicted and observed secondary structure of t4 phage lysozyme," *Biochimica et Biophysica Acta (BBA)-Protein Structure*, vol. 405, no. 2, pp. 442–451, 1975.