# Landmine Detection Using Autoencoders on Multi-polarization GPR Volumetric Data

Paolo Bestagini, *Member, IEEE,* Federico Lombardi, *Student Member, IEEE,* Maurizio Lualdi, Francesco Picetti, *Student Member, IEEE,* and Stefano Tubaro, *Senior Member, IEEE*

*Abstract*—Buried landmines and unexploded remnants of war are a constant threat for the population of many countries that have been hit by wars in the past years. The huge amount of casualties has been a strong motivation for the research community toward the development of safe and robust techniques designed for landmine clearance. Nonetheless, being able to detect and localize buried landmines with high precision in an automatic fashion is still considered a challenging task due to the many different boundary conditions that characterize this problem (e.g., several kinds of objects to detect, different soils and meteorological conditions, etc.). In this paper, we propose a novel technique for buried object detection tailored to unexploded landmine discovery. The proposed solution exploits a specific kind of convolutional neural network (CNN) known as autoencoder to analyze volumetric data acquired with ground penetrating radar (GPR) using different polarizations. This method works in an anomaly detection framework, indeed we only train the autoencoder on GPR data acquired on landmine-free areas. The system then recognizes landmines as objects that are dissimilar to the soil used during the training step. Experiments conducted on real data show that the proposed technique requires little training and no ad-hoc data pre-processing to achieve accuracy higher than 93% on challenging datasets.

*Index Terms*—Demining, GPR, Machine Learning, Deep Learning, CNN.

## I. INTRODUCTION

In the last seventy years, landmines have been massively deployed in a huge amount of countries all over the world. It is not possible to provide a global estimate of the total area contaminated by landmines, due to a lack of data. However, global deaths and injuries from landmines have hit a ten-year high, with the latest available figures from 2016 showing a sharp increase on the previous year [1].

The United Nations Department of Humanitarian Affairs (UNDHA) has very strict regulations in terms of civil area demining. As a matter of fact, the 99.6% of mines and unexploded ordnance must be safely removed from an area in order to consider it landmine-free [2]. This is strongly different from military demining, in which case only a few paths need to be freed to enable vehicles passage, and little attention is paid to the surrounding areas. For this reason, humanitarian demining operations are typically performed manually, thus making demining an extremely risky and long operation.

In order to ease demining operations and make them safer, different systems have been proposed in the literature. In particular, landmine clearance systems generally follow a common pipeline, and different solutions have been proposed for each step of it. These steps can be summarized as:

- *Detection*: given an area of investigation, detecting whether possibly dangerous objects are buried. During this step, it is important to have a very low false negative rate (i.e., percentage of landmines not detected), whereas it is possible to tolerate high false positive rates (i.e., non-explosive objects detected as dangerous), which will be reduced during the next step.
- *Localization*: given that a target has been detected within an area, providing more accurate localization information about the position of the buried object.
- *Recognition*: given that a buried object has been detected and localized, being able to understand whether it actually is a landmine, or it is a misclassification of the detection step (e.g., a stone, tree branches, etc.).

In this paper, we focus on the detection problem, by proposing a method based on convolutional neural networks (CNNs) applied to ground penetrating radar (GPR) data to detect the presence of buried objects not coherent with the ground under analysis. The choice of using GPR, methodology capable of detecting even small dielectric changes in the subsurface, is determined by the fact that the vast majority of modern landmines are mainly composed of plastic materials, thus significantly reducing the efficacy of traditional metal detector surveys.

In the last few years, many methods exploiting GPR technique to detect buried objects and landmines have been proposed [3], [4]. Some of these methods are based upon model-based solutions that exploit the knowledge of wave propagation laws. As an example, in [5] an automatic data interpretation system is proposed to estimate buried pipes from GPR B-scan images. In [6] GPR B-scans are packed together and then processed in a 3D domain in order to efficiently detect buried pipes. In [7] the authors devise a 3D-to-2D data conversion filter in order to apply the 2D Reverse-Time Migration to 3D GPR data.

Due to the impressive results obtained by machine learning in the last years, other detection methods exploit supervised learning strategies. As an example, in [8], reflection hyperbolas are localized using pattern recognition techniques. Recently,

the problem of properly training machine learning algorithm for buried threat detection has been also studied [9].

Due to the development of deep learning solutions that often outperform classical methods in many fields (e.g., computer vision, pattern recognition, etc.), these kinds of techniques are being applied to landmine detection. As an example, in [10], the authors devise a real-time method for hyperbola recognition that employs a CNN for discriminating hyperbola from non-hyperbola traces in B-scans. In [11] the authors compare a neural network against logistic regression algorithms in discriminating between potential targets and clutter. Additionally, in [12], [13], the authors propose different CNN-based strategies for landmine detection in B-scans, while in [14] the authors compare implementations of shallow and deep learning approaches, showing that there are few conditions that prefer the former to the latter.

In this work we exploit a particular CNN architecture known as autoencoder to analyze GPR volumetric data acquired with multiple polarizations, building upon our recent findings reported in [13]. The proposed method is able to detect whether an investigated volume contains buried objects that are not coherent with the kind of soil being analyzed. This is done exploiting an anomaly detection scheme, i.e., the autoencoder is trained only using mine-free data in order to learn how to model the soil of interest. At deployment time, the autoencoder recognizes the presence of anomalous objects such as landmines, as they are not coherent with the training data. The proposed solution is tested on two different datasets consisting of real GPR acquisitions. Specifically, in two different sand pits we have buried inert landmines, along with plastics and metal objects that produce GPR scans similar to those of landmines. Results show promising when compared with state-of-the-art methods developed for the same kind of data.

In terms of novelty, the following aspects can be pointed out:

- This is the first solution jointly exploiting horizontal and vertical polarizations for landmine detection using CNNs, to the best of our knowledge.
- We investigate the use of a CNN that works on 3D volumetric data, rather than simple 2D B-scans.
- We propose a comprehensive exploration of autoencoders for GPR buried target detection.
- The proposed architecture is able to work in a cross-dataset scenario, i.e., we can train the CNN on a dataset, and use it on another one with limited accuracy loss, thus making the solution robust to different environments.
- Datasets can be relatively small for training purposes, thus the proposed architecture can be daily refined for the specific area of investigation; indeed the need for generalization is not strict for landmine detection.

The code representing a running example of the proposed solution on a part of the dataset is also available[1].

The rest of the paper is structured as it follows. Section II reports the background concepts useful to understand the rest of the paper. Section III formally introduces the detection problem faced in this work. Section IV contains all the algorithmic details of the proposed solution. Section V reports information about the used experimental setup. Section VI provides the achieved numerical results compared with state-of-the-art deep-learning solutions. Finally, Section VII concludes the paper.

## II. BACKGROUND

In this section, we report some background information useful to understand the rest of the manuscript. We start introducing some literature on the use of machine learning for geophysical data processing. We then introduce the main notions behind autoencoders. We finally report relevant information about GPR polarization.

### A. Machine Learning for geophysics

Deep learning, and in particular convolutional neural networks (CNNs), have shown very good performance in several computer vision applications such as image classification, face recognition, pedestrian detection and handwriting recognition [15].

Recently, learning-based methodologies have been increasingly explored for different applications also by the geophysical community. As an example, an application that has been analyzed by different authors is the analysis and understanding of Earth's subsurface structures. To this purpose, in [16] the authors propose to use machine-learning based tools. Additionally, different techniques have been compared in [17] and [18] for the problem of facies recognition and data filtering/analysis [19]. [20] employs a deep CNN classifier to detect faults in seismic images. [21] spots parallels between natural and seismic images, and employs autoencoders for feature extraction. [22] clusters seismic signals; [23] exploits the reduced dimensionality of the hidden space of under-complete autoencoders for developing a denoising machine to be applied on seismic shot gathers; [24], [25] apply deep autoencoders to SAR imaging. Moreover, a wide variety of more classical image processing applications have been studied in a learning-based framework. These involve image restoration, tomography, and inverse problems [26], [27], [28], [29]. Another interesting application is that of anomaly detection, a well-established task for computer systems [30]. In this framework, in [31], a method based on artificial neural networks to extract relevant features is proposed. Moreover, inspired by the brilliant performance that autoencoders have achieved on any kind of signals and tasks (e.g., audio and speech [32], image and video [33], [34], [35], robotics [36] and network traffic analysis [37]), also the imaging inverse problems started to be investigated through the CNN anomaly detection schemes [38], [39].

Recently, machine learning techniques have gained the trend also in GPR-specific applications. [40] uses deep learning classifiers for improving efficiency of cavity detection on GPR data. In [8], buried objects are localized through a Viola-Jones learning algorithm and a Hough Transform specifically tailored to hyperbolas. The authors of [41] proposed to classify GPR

---

[1]https://tinyurl.com/lndmn

B-scans patches through a random forest algorithm that processed Histograms of Oriented Gradients (HOG) information; then they proposed in [42] a deep CNN responsible for both the feature extraction and the classification. In [11] the authors compare a neural network against logistic regression algorithms in discriminating between potential targets and clutter in B-scans obtained from multi-frequency GPRs. Anomaly detection autoencoders are proposed in [43]. Specifically, the CNN learns how to reconstruct normalized auto-correlations of background A-scans. The same authors lately proposed a deep classifier on B-scan [44]. In [10], the authors devise a real-time method for hyperbola recognition that employs a CNN for discriminating hyperbola from non-hyperbola traces in B-scans, where areas of interest have been previously determined by a clustering algorithm. In [12] a CNN classifier is devised for detecting hyperbolas in GPR B-scans. The system is trained on a large dataset of synthetic along with a few background-only real data and it reaches good performance in terms of detection accuracy. In [45] a deep classifier is pre-trained with natural images from CIFAR-10 dataset and then refined on GPR hyperbolas; similarly, in [46] the authors deploy a deep CNN architecture (the well-known Inception-v2 [47]) pre-trained on COCO dataset and deployed for discriminating anti-tank (AT) from anti-personnel (AP) landmines. However, the relatively small availability of data limits the performance of deep learning approaches. To overcome the problem of acquiring GPR B-scans over real landmines, an alternative solution based on a one-class approach is proposed in [13] in which he authors train deep autoencoders for reconstructing background-only B-scans, thus failing in reconstructing hyperbolas. Therefore an anomaly detection scheme is proposed.

### B. Convolutional Autoencoders

In this section, we provide the concepts behind autoencoders to clarify the rest of the paper. For an in-depth autoencoder review, refer to [15]; for a review of autoencoder applications to geophysical problems, refer to [48], [49] .

A CNN is a complex computational model partially inspired by the human neural system that consists of a high number of interconnected nodes, called *neurons*. Such nodes are organized in multiple stacked layers, each one performing a simple operation on its input. In the literature, typical neuron layers are:

- *Fully connected*: layers performing dot product between the input and a parametric matrix.
- *Convolutional*: layers applying multi-dimensional convolution to the input.
- *Transposed Convolutional*: layers applying upsampling followed by convolution to the input. The output has the same spatial resolution a hypothetical deconvolutional layer would produce.
- *Activation*: layers applying a non-linear operation (e.g., hyperbolic tangent, rectification, etc.) to the input.
- *Pooling*: layers processing the input with a moving window, and outputting a selection of the samples within each window (e.g., the maximum, minimum, average, etc.).
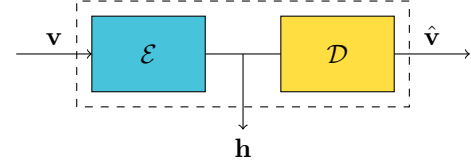


Fig. 1. Scheme of an undercomplete autoencoder. The encoder $\mathcal{E}$ turns the input $\mathbf{v}$ into its hidden representation $\mathbf{h}$, which is turned into $\hat{\mathbf{v}}$ by the decoder $\mathcal{D}$.

Many of these operations depends on numeric parameters that can be tuned based on experience, so that the model is able to learn complex functions. This is done by minimizing a cost function depending on the output of the last layer of the network. This minimization is achieved using back-propagation coupled with an optimization method such as gradient descent and the use of large annotated training datasets, thus enabling the network to capture patterns in the input data and automatically extract distinctive features. In [50] the authors discuss the development and improvements in gradient descend methods for geophysical error minimization problems.

To train a CNN model for a specific image classification task we need:

- To define the meta-parameters of the CNN, i.e., the sequence of operations to be performed, the number of layers, and the shape of the filters.
- To define a proper cost function to be minimized during the training process.
- To prepare a dataset of training and test images, annotated with labels according to the specific tasks.

In this paper we exploit a specific class of CNN, the *autoencoder*, that takes its name from the ability of being logically split into two separate components, as shown in Fig. 1:

- The *encoder*, which is the operator $\mathcal{E}$ mapping the input $\mathbf{v}$ into the so called hidden representation $\mathbf{h} = \mathcal{E}(\mathbf{v})$.
- The *decoder*, which is the operator $\mathcal{D}$ that decodes the hidden representation into an estimate of the original input $\hat{\mathbf{v}} = \mathcal{D}(\mathbf{h})$.

In general, we can state that autoencoders learn how to reconstruct the input data by going to/coming from the hidden space. This is done by using as training cost function some distance metric between the input $\mathbf{v}$ and the output $\hat{\mathbf{v}} = \mathcal{D}(\mathcal{E}(\mathbf{v}))$.

Specifically, we make use of an *undercomplete convolutional autoencoder*, i.e., a specific autoencoder characterized by a hidden representation $\mathbf{h}$ of reduced dimensionality with respect to the input $\mathbf{v}$ [15]. By using this kind of autoencoder it is possible to estimate an almost-invertible dimensionality reduction function $\mathcal{E}$ directly from a representative set of training data (i.e., observations of $\mathbf{v}$). In the light of this, we can interpret the hidden representation $\mathbf{h} = \mathcal{E}(\mathbf{v})$ as a compact feature vector capturing salient information from $\mathbf{v}$.

### C. GPR Polarization

In this section we briefly describe the electromagnetic polarization of a GPR antenna, with a special focus on the

benefits provided by combining orthogonally polarized dipole antennas in soil imaging applications.

The polarization information contained in the waves backscattered from a given target is highly related to its geometrical structure and orientation as well as to its physical properties and it affects how a radar system sees the object in the scenario. This implies that the visibility of a subsurface scatterer can be enhanced or reduced by taking advantage of its expected polarimetric response [51], [52], [53], [54], [55], [56].

In our work, we consider a landmine as a regular and homogeneous object, and hence a polarization-independent scattering behavior can be assumed (i.e., the power of the reflected wave is independent of the transmitted polarization). This ensures that coherently combining data acquired with different antenna polarizations would enhance the signal strength, due to the fact that no differences are expected in the target response [57], [58], [59]. On the other hand, the GPR response of the surrounding soil is not expected to exhibit this property, and hence a coherent combination of polarimetric data might lead to a reduction of its amplitude level [60], [61], [62], [63].

The use of multi polarization GPR data did not always prove beneficial in the literature [64], [65]. Indeed, polarization effects are also strongly linked to the considered geometrical setup. In this work, we consider investigating the polarization effect an important step in order to better optimize the proposed buried object detection system. Therefore, in order to better capture the presence of buried objects in a radar image, we explore the possibility of leveraging polarization diversity by stacking the GPR images obtained from two orthogonal polarizations.

## III. PROBLEM FORMULATION

An adequate solution to the problem posed by landmines implies that the percentage of detected mines in the area under analysis should approach $100\%$ with the lowest possible number of false negatives (i.e., mistaking a mine for a non-dangerous object). To give a reference, the United Nations have set the detection goal at $99.6\%$; in order to achieve this rate, in principle, a system could accept an high false positive rate (i.e. the percentage of inhert buried objects that are wrongly classified as landmine). However the U.S. Army allows one false alarm every $1.25$ square meters [66]. Unfortunately, no existing landmine detection system meets these criteria. This could be conferred to the properties of the mines themselves and the variety of environments in which they are buried, as well as to the limits and flaws in the current technology [66].

The process of landmine identification is threefold: first, the soil is analyzed in order to find the presence of buried objects; then, a more accurate analysis is performed in order to localize the targets in 3D; finally, the volume of interest is investigated with classification techniques in order to discriminate landmines from rocks and other buried objects.

In this paper we focus on the first step. Making use of the GPR technology, the purpose of the proposed methodology is to detect whether a soil volume contains any trace of buried
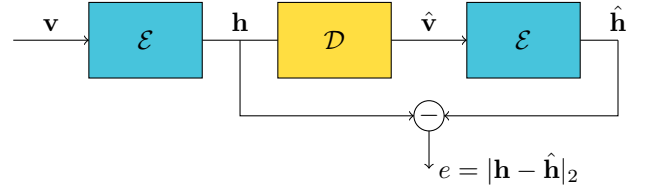


Fig. 2. The proposed anomaly detection scheme. A block under analysis $\mathbf{v}$ is autoencoded to $\hat{\mathbf{v}}$ and encoded again into $\hat{\mathbf{h}}$. High Euclidean distance in the hidden space indicates anomaly.

objects, whose dimensions and electromagnetic signature are compatible to those of landmines.

To formalize the goal of the proposed method, let us define a B-scan acquired with a GPR system as the 2D image $\mathbf{V}(t, x)$ whose coordinates are the reflection time $t$ and the *inline* axis $x$. By concatenating $Y$ consecutive B-scans, we sample the soil also in the third dimension, the *crossline* axis $y$, thus obtaining the volume $\mathbf{V}(t, x, y)$.

If the $y$-th B-scan has been acquired over a buried target, we associate to it the binary label $l(y) = 1$ indicating the presence of an object underground. If it has been acquired over a target-free area, we label it with $l(y) = 0$, indicating that no object traces are present.

Solving buried object detection problem under this hypothesis consists in taking the volume $\mathbf{V}(t, x, y)$ as input, and outputting a label $\hat{l}(y)$ (i.e., an estimate of $l(y)$) for each B-scan (i.e., $y \in [1, Y]$). Correct detection occurs every time $\hat{l}(y) = l(y)$. Misclassification occurs in case $\hat{l}(y) \neq l(y)$.

Notice that this way of proceeding is challenging the algorithm performance. As a matter of fact, in a practical scenario, detecting an object within a B-scan would be sufficient for its removal, even if the object is big enough to span multiple B-scans. However, we consider the more challenging scenario in which all B-scans showing traces of an object must be labeled correctly.

## IV. PROPOSED METHODOLOGY

As reported in Section II-B, the autoencoder learns how to encode and decode the training data. Once trained, if the input data are too different from that of the training set, the autoencoder fails in reconstructing it correctly, as the encoding / decoding procedure is highly sub-optimal. Therefore, the error introduced in encoded or decoded data can be used as an indicator of anomaly of the processed data with respect to the training data. For this reason, the autoencoder can be a powerful instrument for anomaly detection [67], [68].

In our application scenario, it is possible to train an autoencoder to learn a hidden representation of small GPR volumes not showing any object trace. After training, this autoencoder processes the whole data volume under analysis. Then, an anomaly mask is computed, and a label $\hat{l} = 0$ or $\hat{l} = 1$ is associated to each B-scan, depending on the value of the anomaly metrics.

In the following, we describe each step of the proposed method, from data pre-processing, to network training and testing procedures.

## A. Data Pre-processing

Let us define the data volume obtained with horizontal polarization as $\mathbf{V}_\mathrm{H}(t, x, y)$, and the relative volume obtained with vertical polarization as $\mathbf{V}_\mathrm{V}(t, x, y)$. The first step of our algorithm consists in merging these two volumes into a single one but taking advantage of what the different polarization can detect.

To do so, we separately align each pair of A-scans acquired with horizontal and vertical polarization, i.e., $\mathbf{V}_\mathrm{H}(t)$ and $\mathbf{V}_\mathrm{V}(t)$. This is done by estimating the time shift $\tau$ between them as the amount of samples corresponding to the lag that maximizes the cross-correlation between the A-scans, i.e.,

$$\hat{\tau} = \arg\max_\tau \chi(\tau), \tag{1}$$

where $\chi(\tau)$ is the cross-correlation between $\mathbf{V}_\mathrm{H}(t)$ and $\mathbf{V}_\mathrm{V}(t)$. We then merge the two A-scans into a single one by re-alignment and averaging as

$$\mathbf{V}_\mathrm{A}(t) = \frac{\mathbf{V}_\mathrm{H}(t) + \mathbf{V}_\mathrm{V}(t - \hat{\tau})}{2}. \tag{2}$$

A new data volume $\mathbf{V}_\mathrm{A}(t, x, y)$ is then generated by simply concatenating all aligned A-scans $\mathbf{V}_\mathrm{A}(t)$ coherently.

## B. Convolutional Autoencoder

After the data volume has been processed, we can feed it to the proposed autoencoder. In the following, we report the strategy we propose to split the data volume into smaller blocks, train the autoencoder, and use it at test time.

*1) Block Extraction:* Rather than processing the available data volume $\mathbf{V}(t, x, y)$ as a whole[2], we split it into a subset of $I$ regular smaller blocks $\mathbf{v}_i(t, x, y)$, $i \in [1, I]$. These blocks can be either overlapped or not, depending on the desired trade-off between detection robustness and computational complexity. With reference to Figure 3, the $i$-th volume blocks can be defined as

$$\begin{aligned} \mathbf{v}_i(t, x, y) = &\mathbf{V}(t_i + t - 1, x_i + x - 1, y_i + y - 1), \\ &t \in [1, \Delta t], x \in [1, \Delta x], y \in [1, \Delta y], \end{aligned} \tag{3}$$

where $\Delta t$, $\Delta x$ and $\Delta y$ are parameters that define the block size, whereas the differences $\delta t = (t_{i+1} - t_i)$, $\delta x = (x_{i+1} - x_i)$ and $\delta y = (y_{i+1} - y_i)$ determine blocks overlap and are known as strides, i.e., the amount of samples of the original volume between one block and the next one. Notice that, in case of overlap (i.e., $\Delta t > \delta t$, or equivalently for $x$ and $y$), one sample of $\mathbf{V}$ will belong to many blocks $\mathbf{v}_i$. Moreover, in case of $\Delta y = 1$, volume blocks $\mathbf{v}_i$ collapse into 2D B-scan patches (a sub-optimal choice which will be cleared in the results presentation).

The motivation behind the choice of using small volumetric blocks is threefold:

- Splitting volumes into smaller blocks allows us to obtain a greater amount of data for training, considering that acquiring a B-scan is time consuming.
- Feeding the CNN with smaller data volumes enables working with lighter architectures.

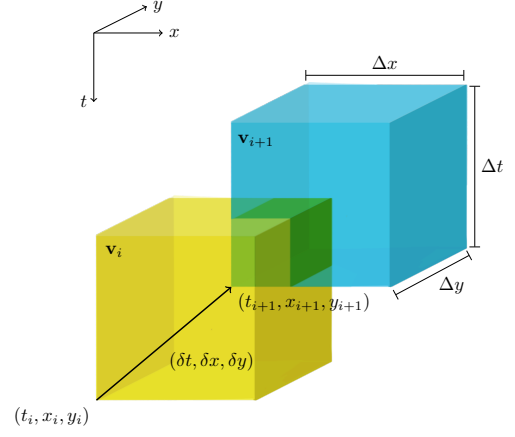[2]Let us drop the subscript A for the sake of notational compactness.



Fig. 3. Representation of the extraction of two adjacent blocks $\mathbf{v}_i$ (yellow) and $\mathbf{v}_{i+1}$ (blue) from volume $\mathbf{V}$. Being $\delta t < \Delta t, \delta x < \Delta x, \delta y < \Delta y$ the blocks are overlapped in the green sub-block.

- The system becomes independent from the whole volume size, i.e., any volume of arbitrary size can be split into equal-size blocks and can be analyzed.

*2) System Training:* As shown in Figure 1, the proposed architecture takes a volume $\mathbf{v}_i$ as input, it shrinks it to a smaller dimensionality representation $\mathbf{h}_i$, and it outputs a volume $\hat{\mathbf{v}}_i$ of the same size of the input. To train the autoencoder, we define a training set of $I$ blocks $\mathbf{v}_i$, $i \in [1, I]$ extracted from $B$ adjacent B-scans associated to label $l = 0$ (i.e., the selected volume does not contain any hyperbola due to buried objects). We then estimate the autoencoder weights by minimizing a loss function $\mathcal{L}$ defined as

$$\mathcal{L} = \frac{1}{I} \sum_{i=1}^{I} |\mathbf{v}_i - \hat{\mathbf{v}}_i|_2^2, \tag{4}$$

where $|\cdot|_2$ represents the $\ell^2$-norm of a vector. In other words, the used loss function is the mean squared error between the input volume $\mathbf{v}_i$ and the autoencoder output $\hat{\mathbf{v}}_i$ averaged over the whole training set. Once the system has converged, we can state that the autoencoder has learnt how to correctly reconstruct background soil information.

*3) System Deployment:* When a volume of soil $\mathbf{V}$ is to be analyzed, we split it into a set of $I$ blocks $\mathbf{v}_i$, $i \in [1, I]$ covering the whole $\mathbf{V}$. Figure 2 explains the anomaly detection procedure. Each block is encoded into its hidden representation $\mathbf{h}_i = \mathcal{E}(\mathbf{v}_i)$. The hidden representation is decoded into $\hat{\mathbf{v}}_i = \mathcal{D}(\mathbf{h}_i)$, which is encoded again into $\hat{\mathbf{h}}_i = \mathcal{E}(\hat{\mathbf{v}}_i)$. We then compare the hidden representations of both the original and autoencoded blocks by means of Euclidean distance, thus computing

$$e_i = |\mathbf{h}_i - \hat{\mathbf{h}}_i|_2. \tag{5}$$

The distance between the hidden representations can be used as anomaly detector. Indeed, blocks containing hyperbola traces produce a hidden representation $\hat{\mathbf{h}}_i$ very different from $\mathbf{h}_i$. On the other hand, the two hidden representations are similar if the samples under analysis are similar to those of the training (i.e., they do not contain traces of buried objects).

(a) Clean B-scan



(b) Mined B-scan



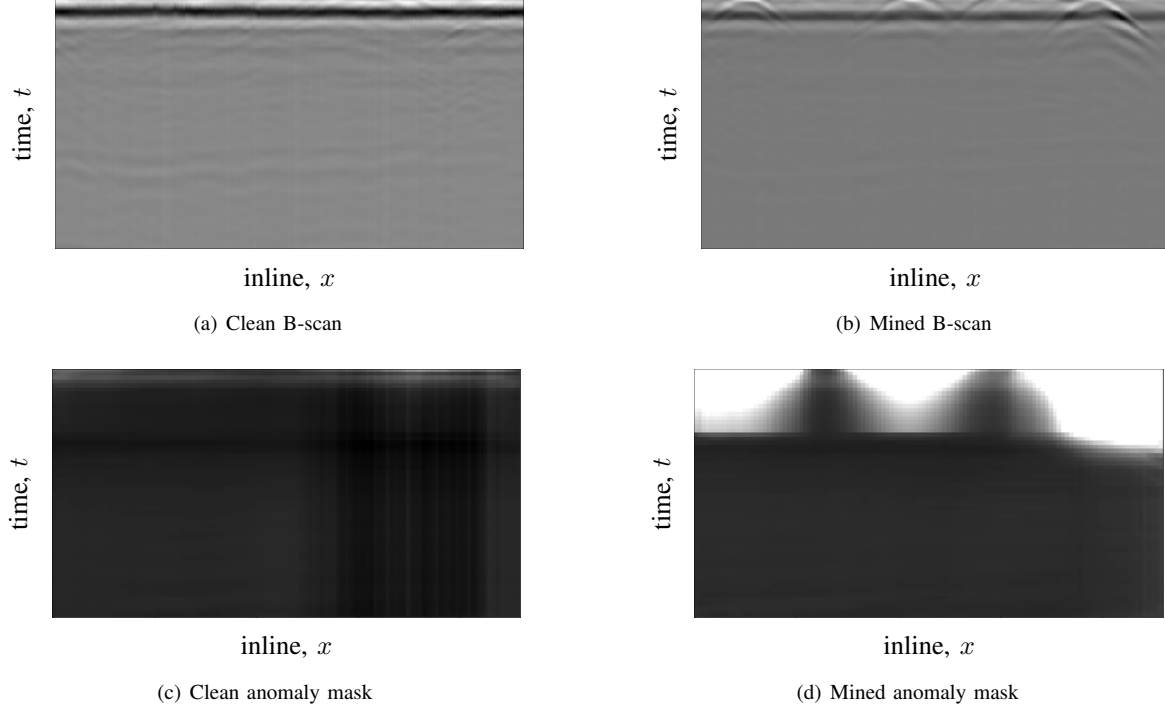(c) Clean anomaly mask



(d) Mined anomaly mask

Fig. 4. Example of background B-scans (landmine-free (a) and mined (b)) with the corresponding anomaly masks (c) and (d) normalized on the same scale of values. It is possible to see that hyperbola-related anomalies are clearly highlighted in white.

## C. Aggregation

For each block $\mathbf{v}_i$ of the input volume $\mathbf{V}$ we have computed the Euclidean distance $e_i$ according to (5). The goal of the aggregation step is to merge all obtained $e_i$ values into a volumetric anomaly mask $\mathbf{M}$ the same size $\mathbf{V}$, indicating which samples of $\mathbf{V}$ should be considered anomalies, thus buried objects.

If we consider the case of overlapped $\mathbf{v}_i$ blocks, a single sample of $\mathbf{V}$ belongs to different blocks $\mathbf{v}_i$. Therefore, we have more $e_i$ values associated to a single sample of the volume $\mathbf{V}$. For this reason, the volumetric anomaly mask $\mathbf{M}$ is constructed by an overlap-and-average operation.

In practice, let us define the set of indexes $i$ of blocks $\mathbf{v}_i$ containing the sample $\mathbf{V}(t, x, y)$ as

$$\mathcal{I}_{t,x,y} = \{i \mid \mathbf{V}(t, x, y) \subset \mathbf{v}_i\}, \tag{6}$$

where, with a slight abuse of notation, we use $\subset$ to express that the sample $\mathbf{V}(t, x, y)$ is contained in block $\mathbf{v}_i$. The mask $\mathbf{M}$ is then defined as

$$\mathbf{M}(t, x, y) = \frac{1}{|\mathcal{I}_{t,x,y}|} \sum_{i \in \mathcal{I}_{t,x,y}} \mathbf{e}_i, \tag{7}$$

where $|\mathcal{I}_{t,x,y}|$ is the cardinality of the set $\mathcal{I}_{t,x,y}$.

Once the anomaly mask $\mathbf{M}$ has been obtained, in order to detect landmines in the $y$-th B-scan, we consider the maximum value in the $y$-th slice of $\mathbf{M}$ and apply the following criterion:

$$\hat{l}(y) = \begin{cases} 1, & \text{if } \max_{t,x} \ \mathbf{M}(t, x, y) > \Gamma, \\ 0, & \text{otherwise}, \end{cases} \tag{8}$$

where $\Gamma$ is a global threshold to be selected upon a set of training B-scans at system tuning time. In other words, if we detect a big anomaly in a B-scan, that B-scan is labeled as suspicious.

Figure 4 shows examples of processed B-scans along with the corresponding slices of the anomaly mask. In particular, to a B-scan containing any trace of hyperbolas it correspond a uniformly low-valued anomaly slice; on the other hand, when the B-scan does contain hyperbolas, the correspondent mask clearly shows high values of the anomaly metrics.

## V. EXPERIMENTAL SETUP

In this section we report the details related to all tested CNN architectures, the considered acquisition system, as well as the datasets used during our experimental campaign.

### A. Autoencoder Architectures

In the following, we describe the six CNN architectures we have tested to identify the best candidate solution. Specifically, we considered three different architectures, each one working with either 2D patches, or 3D volume blocks as input. All architectures are symmetric, as to each convolutional layer used at the encoder, corresponds a deconvolutional layer at the decoder. The input size of each network is equal to its output size, as per autoencoder definition. The main difference between the architectures is the hidden representations dimensionality: different architectures compress more (or less) the input data. We tested these different autoencoder architectures to investigate the impact of this reduction.

The first architecture is $\mathcal{A}_1$, and it is composed of:

- One 2D convolutional layer with 16 filters, stride 1×1, size 6×6.
- Three 2D convolutional layer with 16 filters, stride 2×2, size 5×5, 4×4, 3×3, respectively.
- A 2D convolutional layer with 16 filters, stride 2×2, size 2×2. Its output is the hidden representation.
- Four 2D deconvolutional layers with 16 filters, stride 2×2, size 2×2, 3×3, 4×4, 5×5, respectively.
- One 2D deconvolutional layers with 1 filter, stride 1×1, size 6×6, followed by hyperbolic tangent activation.

This architecture shrinks the input by a factor 16 (e.g., a 32×32 image is turned into a 64 element hidden representation).

Architecture $\mathcal{A}_2$ is the same as $\mathcal{A}_1$, but the convolutional layer returning the hidden representation is substituted by one 2D convolutional layer with 8 filters, stride 1×1, size 1×1. This architecture shrinks the input by a factor 32 (e.g., a 32×32 image is turned into a 32 element hidden representation).

Architecture $\mathcal{A}_3$ is the same as $\mathcal{A}_1$, but the convolutional layer returning the hidden representation is substituted by three layers:

- One 2D convolutional layer with 16 filters, stride 2×2, size 2×2.
- One 2D convolutional layer with 16 filters, stride 2×2, size 1×1.
- One 2D deconvolutional layer with 16 filters, stride 2×2, size 2×2.

This architecture shrinks the input by a factor 64 (e.g., a 32×32 image is turned into a 16 element hidden representation).

When these architectures are applied to 2D patches (i.e., blocks $\mathbf{v}_i$ considering the case of $\Delta y = 1$), the input has size $\Delta t \times \Delta x \times 1$. In the 3D scenario (i.e., blocks $\mathbf{v}_i$ considering the case of $\Delta y \neq 1$), the input has size $\Delta t \times \Delta x \times 3$. In other words, the $\mathbf{v}_i$ block is composed by three 2D patches from three adjacent B-scans that are stacked togheter in the third dimension. Thus, the 2D convolution is applied along the $t, x$ coordinates and repeated along the $y$ coordinate. This operation is similar to natural image processing methods, where the first two dimensions are the pixel coordinates and the third dimension is the color channel. We did choose not to build our architectures upon fully 3D convolutions because the acquired data is intrinsically 2D, as $y$ is the B-scans index.

In the following, in the 2D scenario, we refer to these architectures as $\mathcal{A}_1^{\text{2D}}$, $\mathcal{A}_2^{\text{2D}}$ and $\mathcal{A}_3^{\text{2D}}$, respectively. In the 3D scenario, we refer to these networks as $\mathcal{A}_1^{\text{3D}}$, $\mathcal{A}_2^{\text{3D}}$ and $\mathcal{A}_3^{\text{3D}}$, respectively.

All networks have been trained using Adam optimizer [69] with default parameters, until loss function stopped decreasing on a small set of validation data taken from the training set. Network input was always normalized in range $[-1, 1]$.

### B. Acquisition System

The GPR system employed for the measurements consisted of an IDS Aladdin radar (provided by IDS Georadar srl,
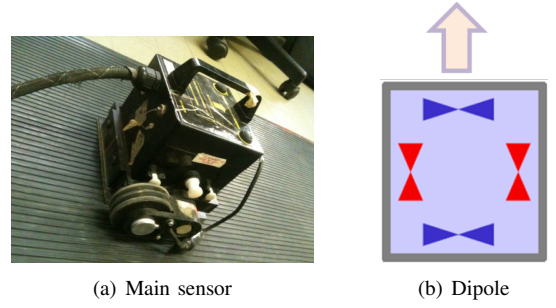


(a) Main sensor      (b) Dipole

Fig. 5. Employed GPR system. (a) Main sensor. (b) Dipoles scheme and geometry.

Figure 5(a)), an impulse device carrying dipole antennas with a central frequency and a bandwidth of 2 GHz. The equipment is composed by two pairs of orthogonally polarized dipole antennas, as shown in Figure 5(b), located such that the reflection center corresponds for both couples and coincides with the geometrical center of the unit.

This configuration guarantees precise matching between the parallel (i.e., vertical) and perpendicular (i.e., horizontal) orientation, in respect to the survey direction, permitting joint orthogonally polarized scans to be acquired in a single pass. Data recording is controlled by an odometric wheel directly connected to the sensor.

Acquisitions were carried out employing a semi-mechanical device (visible in Figure 5(a), [70]), a solution which guarantees accurate data density and regularity, maintaining a precise profile spacing and ensuring a constant antenna orientation during the whole survey.

### C. Acquisition Campaigns

In order to properly evaluate the proposed solution, we acquired a set of real GPR data. The acquisitions were made in two different sand pits of width 1 meter and length 2 meters. The targets were buried at a depth of 5-15 cm, approximately. Specifically, we focused on two different scenarios with incremental landmine detection difficulty.

**Dataset $\mathcal{S}_1$:** The dataset is composed by 9 different objects: 3 real landmines (SB33, PFM1, VS50), 2 printed replica of the VS50, a rubber disk, a crushed aluminum can, a metal sphere and a mortar simulant. The test bed is a indoor confined bay composed of several quadrants (see Figure 6(a)), and filled with a sharp sand material characterized by a very low clay content and a gritty texture, to avoid trench artifacts in the collected data. The material was relatively homogeneous and free of clutter, with an average particle size of less than half centimeter.

Despite the environment humidity, the sand maintained a velocity of $14\,\text{cm/ns}$ and a consequential relative dielectric constant of $4.5$. Although a favorable propagation environment, the material is representative of several mine-affected regions of the world.

**Dataset $\mathcal{S}_2$:** The second scenario consists of 8 different objects: 3 real landmines (PROM1, DM11 and PMN58), a grenade, a metal sphere, an aluminum can, a plastic bottle

(a) $\mathcal{S}_1$



(b) $\mathcal{S}_2$

Fig. 6. Environment details for datasets $\mathcal{S}_1$ and $\mathcal{S}_2$.

TABLE I
ACQUISITION DETAILS FOR DATASETS $\mathcal{S}_1$ AND $\mathcal{S}_2$.

| Parameter | $\mathcal{S}_1$ | $\mathcal{S}_2$ |
|---|---|---|
| Central frequency / Bandwidth | 2 GHz / 2 GHz | 2 GHz / 2 GHz |
| Propagation velocity in soil | 14 cm/ns | 10 cm/ns |
| Time sampling | 0.0117 ns | 0.0117 ns |
| Inline sampling | 0.4 cm | 0.4 cm |
| Crossline sampling | 0.8 cm | 1.6 cm |
| Time window | 15 ns | 20 ns |
| Number of acquired B-scans ($Y$) | 114 | 66 |

and a plant root. as shown in Figure 6(b). In this case, due to the rain falls that occurred few days before the experiment, the sand was not completely dry, providing a propagation velocity of $10\,\mathrm{cm/ns}$ and a resulting dielectric value of 9. The consequence is that the medium may not be strictly considered homogeneous, as the shallow subsurface is expected to be more dry compared to the higher moisture level of the deeper layers.

Acquisition details for the two sites are provided in Table I.

## VI. RESULTS

In this section, we present the experimental results obtained with the proposed methodology. First, we introduce the used detection metric. Then, we describe in detail all the performed tests along with their numerical performances. Finally we provide a few comments about computational resources. The code as well as part of the dataset are available online[3].

[3]https://tinyurl.com/lndmn

### A. Evaluation Metric

The proposed classification method is based on the threshold $\Gamma$ defined in (8). We therefore evaluated our technique by means of receiver operating characteristic (ROC) curve. A ROC curve represents the probability of correct detection (i.e., correctly finding a threat) and probability of false detection (i.e., detecting objects in clear areas) by spanning all possible values of the threshold $\Gamma$. This means that each working point of a ROC curve is determined by a specific $\Gamma$ value. To better understand how ROC curves work, let us consider Figure 7. This figure depicts the maximum values of the Euclidean distance $e_i$ for each B-scan (i.e., $\max_{t,x} \mathbf{M}(t,x,y)$) using three different network configurations on dataset $\mathcal{S}_1$. This figure also reports the ground truth for each B-scan, i.e., the label curve $l(y)$ that assumes value 1 if the $y$-th B-scan contains a threat. Each point of a ROC curve is obtained by comparing the ground truth $l(y)$ with $\max_{t,x} \mathbf{M}(t,x,y)$ binarized with respect to threshold $\Gamma$.

As compact measure of ROC goodness we selected the area under the curve (AUC). This measure ranges between 0 (i.e., estimated labels are inverted with respect to the true ones) and 1 (i.e., perfect result), passing through $0.5$ (i.e., random guess). In order to understand results obtained through the ROC curves, an additional comment is needed. In a real scenario, it is common that traces generated by a single landmine can be observed in multiple B-scans. The way we have decided to evaluate our system, we consider a mis-detection every time we miss a single B-scan, therefore all presented results can than be considered as a lower bound on the performance of the proposed method.

### B. Detection Performance

In order to evaluate the detection performance of the proposed system, we performed a set of tests aiming at validating a specific choice we made. In the following, we discuss all of these tests.

**Block size and stride:** The first test aims at evaluating the impact of the block size and stride. To this purpose, we worked in the 2D scenario. We trained the networks with the first $N = 5$ background B-scans of dataset $\mathcal{S}_1$ using only the horizontal polarization, and tested it on the remaining volume of data.

We choose arbitrary to work with squared blocks, varying the sizes $\Delta t$ and $\Delta x$ between 32, 64, and 128 samples. As the block stride is concerned, we fix $\delta t = \delta x$ and test 4, 16, and 32 samples. From Figure 8, it is possible to observe the block sizes with respect to the input data.

Table II reports the results in terms of AUC values. Concerning the stride, we notice that having a small stride means that blocks are highly overlapped. Therefore, more blocks are extracted from the data volume. This results in a smoother and more detailed anomaly mask, thus making a small stride preferable. Regarding the block size, 128 samples are an oversized dimension, resulting in mediocre outcomes. Conversely, 32 and 64 have almost the same performance.

In the rest of our experiment, we set $\Delta t = \Delta x = 64$ samples for two reasons:
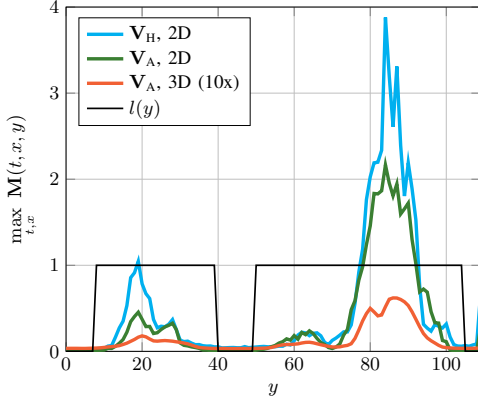
Fig. 7. Examples of maximum anomaly values (colored) related to architectures $\mathcal{A}_3$ applied to dataset $\mathcal{S}_1$. These curves are thresholded and compared with the ground truth (black) in order to obtain the ROC curves.
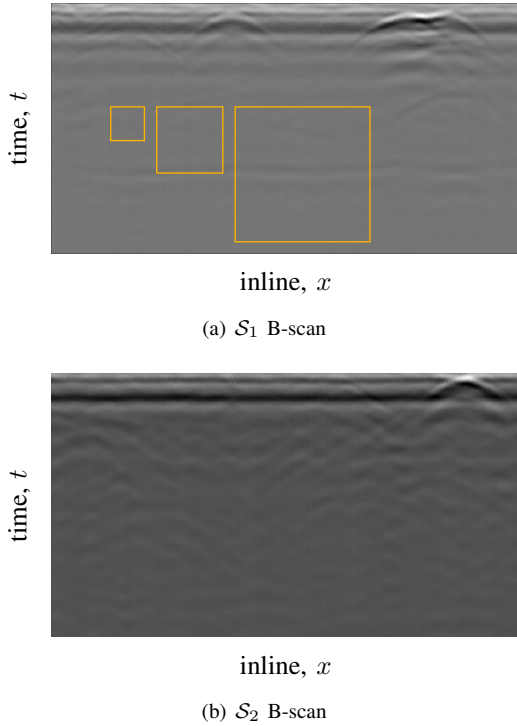


(a) $\mathcal{S}_1$ B-scan



(b) $\mathcal{S}_2$ B-scan

Fig. 8. Examples of B-scans taken from $\mathcal{S}_1$ (a) and $\mathcal{S}_2$ (b). In yellow, three different squared patches of 32, 64, and 128 samples respectively.

- It contains almost the whole hyperbola as suggested in [45].
- It involves less memory usage in terms of CNN training and test (i.e., less patches to analyze without significant loss in accuracy).

In terms of stride, we select $\delta t = \delta x = 4$ as it provides smoother anomaly masks overall, still granting accurate results.

**Training set size:** Having selected the patch size $\Delta t = \Delta x = 64$ and stride $\delta t = \delta x = 4$, we analyze the impact of the number of B-scans needed to train the architectures. For this test, again we consider 2D data from $\mathcal{S}_1$ using only the horizontal polarization.

Table III reports the results of this scenario. Notice that the

TABLE II
IMPACT OF DIFFERENT BLOCK SIZES AND STRIDES IN THE 2D SCENARIO IN TERMS OF AUC.

| Parameters | | $\mathcal{A}_1^{2D}$ | $\mathcal{A}_2^{2D}$ | $\mathcal{A}_3^{2D}$ |
|---|---|---|---|---|
| $\Delta t = \Delta x = 32$ | $\delta t = \delta x = 4$ | 0.9335 | 0.9470 | 0.9400 |
| | $\delta t = \delta x = 16$ | 0.9480 | 0.9455 | 0.9340 |
| | $\delta t = \delta x = 32$ | 0.9255 | 0.9285 | 0.9205 |
| $\Delta t = \Delta x = 64$ | $\delta t = \delta x = 4$ | 0.9375 | 0.9360 | 0.9115 |
| | $\delta t = \delta x = 16$ | 0.9475 | 0.9320 | 0.9550 |
| | $\delta t = \delta x = 32$ | 0.9385 | 0.9130 | 0.8816 |
| $\Delta t = \Delta x = 128$ | $\delta t = \delta x = 16$ | 0.9100 | 0.8646 | 0.8701 |

TABLE III
IMPACT OF DIFFERENT NUMBERS OF B-SCANS ON WHICH THE ARCHITECTURES ARE TRAINED IN TERMS OF AUC.

| Training b-scans | $\mathcal{A}_1^{2D}$ | $\mathcal{A}_2^{2D}$ | $\mathcal{A}_3^{2D}$ |
|---|---|---|---|
| $N = 1$ | 0.9468 | 0.9344 | 0.9247 |
| $N = 3$ | 0.8956 | 0.9329 | 0.9195 |
| $N = 5$ | 0.9320 | 0.9360 | 0.9230 |

system is very robust with respect to this parameter, as there are no significant accuracy losses while changing $N$. However, as there is no reason for not using training data when available, we fix $N = 5$ hereinafter. Moreover, this enables to fairly compare 2D and 3D scenarios, as more than one B-scan is needed to generate volumetric data.

**Different polarizations:** The experiment aims at demonstrating the improvements introduced by exploiting both the horizontal and vertical GPR polarizations, as described in section IV-A. This is done by comparing results obtaining using only horizontal polarization, or both.

Table IV shows the impact of the preprocessed dataset using all the three 2D architectures on $\mathcal{S}_1$. Notice that architecture $\mathcal{A}_3^{2D}$ outperforms all the others, showing an AUC value of 96.5%. This confirms the importance of using both polarizations.

**Volumetric data:** The objective of the following experiment is to compare the 2D architectures against their 3D-extended versions. Results have been obtained on dataset $\mathcal{S}_1$.

Table V reports these results. For all the architectures, adding the third dimension allows to gain from 2 to 5 percentage points. Moreover, the performances reflect the complexity of the networks. The deeper the architecture, the better the results, despite the increased dimensionality reduction rate. In particular, architecture $\mathcal{A}_3^{3D}$ obtains 98% of AUC. This result is well expected, as the exploitation of the high interdependence between adjacent B-scans could only help in terms of detection.

Figure 9 plots the ROC curves which shows the diagnostic ability of the binary classifier while varying the threshold $\Gamma$ applied to the anomaly detection metrics, as described in section VI-A. In particular, we choose to show the improvements in terms of detection ability due to the adoption of multi-polarization and volumetric data. It is worth noticing also that

TABLE IV
IMPACT OF THE PRE-PROCESSING APPLIED TO GPR POLARIZATIONS IN TERMS OF AUC.

| GPR Polarization | $\mathcal{A}_1^{2D}$ | $\mathcal{A}_2^{2D}$ | $\mathcal{A}_3^{2D}$ |
|---|---|---|---|
| $\mathbf{V}_H$ | 0.9320 | 0.9360 | 0.9105 |
| $\mathbf{V}_V$ | 0.8691 | 0.9400 | 0.9635 |
| $\mathbf{V}_A$ | 0.9170 | 0.9240 | **0.9650** |

TABLE V
COMPARISON OF THE DETECTION PERFORMANCE OF BOTH THE 2D AND 3D ARCHITECTURES ON PRE-PROCESSED DATA FROM $\mathcal{S}_1$.

| Architecture | AUC |
|---|---|
| $\mathcal{A}_1^{2D}$ | 0.9170 |
| $\mathcal{A}_2^{2D}$ | 0.9240 |
| $\mathcal{A}_3^{2D}$ | **0.9650** |
| $\mathcal{A}_1^{3D}$ | 0.9555 |
| $\mathcal{A}_2^{3D}$ | 0.9750 |
| $\mathcal{A}_3^{3D}$ | **0.9755** |



Fig. 9. ROC curves for three setups on dataset $\mathcal{S}_1$: $\mathcal{A}_3^{2D}$ applied on horizontal (blue) and vertical (yellow) polarization data; green, $\mathcal{A}_3^{2D}$ applied on preprocessed data; red, $\mathcal{A}_3^{3D}$ applied on preprocessed data.

TABLE VI
RESULTS OF CROSS-TESTS BETWEEN $\mathcal{S}_1$ AND $\mathcal{S}_2$ WITH PREPROCESSED DATA AND ARCHITECTURE $\mathcal{A}_3^{3D}$.

| | | TEST | |
|---|---|---|---|
| | | $\mathcal{S}_1$ | $\mathcal{S}_2$ |
| TRAIN | $\mathcal{S}_1$ | 0.9755 | 0.8095 |
| | $\mathcal{S}_2$ | 0.8426 | 0.9338 |

the 3D extension on multi-polarization data (the red line) can achieve a true positive rate of almost 90% at a false positive rate of 0%. If considering buried objects rather than B-scans, the system is able to detect threats without skipping any buried objects.

These results confirm the improvement provided by using volumetric data.

**Cross-dataset:** In computer vision and machine learning literatures, CNNs should not be extremely overfitted to training data. Instead, the architectures should provides good performance also on datasets which have not been used for training. It is indeed a common practice to train an architecture in order to be as general as possible, and then fine-tuning it depending on the specific target data [71], [9].

In our scenario, we trained the best proposed architecture (i.e., $\mathcal{A}_3^{2D}$) on dataset $\mathcal{S}_1$, and then tested it on $\mathcal{S}_2$, and vice-versa. Table VI reports the results in terms of AUC value, which stays strictly above 93% when the same dataset is used for training and test, and above 81% in cross-dataset scenarios. It can be noted that, the proposed method is robust against cross-training, thus the system is not strongly conditioned by the kind of soil used during training.

These cross-dataset results suggest that our architecture could be pre-trained on a generic dataset and then refined and deployed on a specific soil by acquiring a few background B-scans on a safe area.

**Computational resources:** Let us briefly comment on the effort required by the proposed workflow. Figure 10 shows the curves of the loss values (4) with respect to the training epoch (i.e., the number of iterations on the whole training dataset). In particular we can state that the proposed methodology converges reasonably fast, as after 12 epochs, the network does not improve over validation anymore. It is therefore possible to retrain the network for each specific kind of soil just before deployment time. As a matter of fact, all tests were
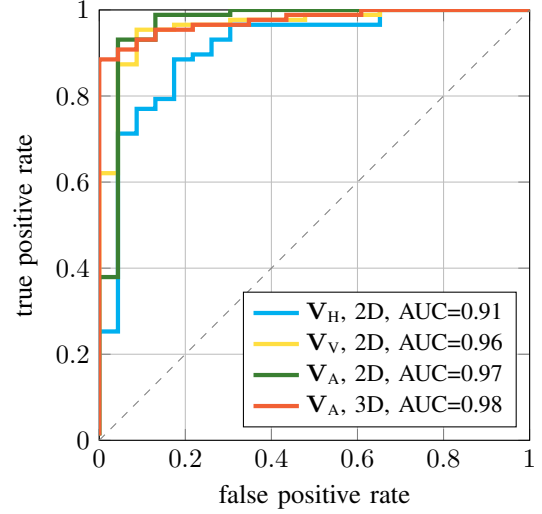
performed on a workstation equipped with a Nvidia Titan X GPU reaching convergence in a few minutes.

### C. Comparison with State-of-the-Art Methodologies

From the experimental campaign just presented, we conclude that the best proposed configuration consists of architecture $\mathcal{A}_3^{3D}$, applied to pre-processed volumetric data $\mathbf{V}_A$, split into 3D blocks characterized by $(\Delta t, \Delta x, \Delta y) = (64, 64, 3)$ and $(\delta t, \delta x, \delta y) = (4, 4, 1)$. This justify both the use of multiple polarization and volumetric data.

In order to compare the proposed solution against other methods proposed in the literature, we consider the following algorithms:

1) First, a simple energy detector. Specifically, we use the maximum value of the energy of the B-scan under analysis as anomaly detector. Indeed, the presence of a buried object can generate a peak in the reflected GPR wave. Figure 11(a) and Figure 11(b) depict the ROC curves on datasets $\mathcal{S}_1$ and $\mathcal{S}_2$, respectively; in particular the best AUC score is 0.8563. However, its capabilities are insufficient for landmine detection purposes in noisy and gritty GPR acquisitions. We believe that the proposed methodology is more effective since it exploits also geometric patterns in the B-scans, leveraging the absence of evident hyperbolas from the training dataset.

2) As a slightly more advanced technique compared to the energy detector, we also implemented a standard
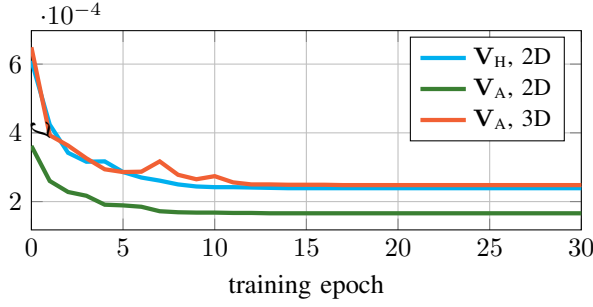
Fig. 10. Examples of validation loss function during the training stage. The convergence is reached in a few iterations.

constant false alarm rate (CFAR) technique. From the ROC curves reported in Figure 11(a) and Figure 11(b) it is possible to notice that the CFAR detector outperforms the energy one on average. However, the best AUC score it achieves is only 0.8508.

3) We re-implemented the model-based methodology proposed in [10]. The authors devise a hyperbola fitting scheme based on energy binarization, clustering and cross-correlation with respect to a reference hyperbola. The methodology is quite fast and would need a little fine-tuning (e.g., the construction of the reference hyperbola knowing the soil properties). However, the detection performance are lower than the autoencoder scheme: on our dataset $\mathcal{S}_1$ we obtained an AUC of 0.6752. Figure 11(a) reports the ROC curve.

4) A classical computer vision approach, proposed in [41], in which a random forest classifier discriminates hyperbolas from background through the different histograms of oriented gradients (HOG). Figure 11(a) reports the ROC curve: the AUC score is 0.8695.

5) From the same authors of [41], a CNN approach to the same problem is proposed in [42]. Here, a deep CNN is responsible for both the features extraction and the classification. However, due to the larger amount of parameters to be trained on a few data, the performances are quite poor in our setup with just a few data. Figure 11(a) reports the ROC curve: the AUC score is 0.6519. We expect this results to improve whether a large amount of GPR acquisition is available.

6) A CNN approach proposed in [12] in which a deep binary classifier is trained with synthetic GPR images of both classes (i.e., target/background) alongh with real background-only B-scans. Figure 11(a) reports the ROC curve when applied to dataset $\mathcal{S}_1$: the AUC score is 0.9659. However, this approach fails where the real data are too different from the synthetic ones. Figure 11(b) reports the ROC curve when applied to dataset $\mathcal{S}_2$: the AUC score is 0.5952.

7) Our basic version of the autoencoder scheme [13], in which neither the multi-polarization nor the 3D architecture are present.

The former exploits a two-class CNN trained jointly on synthetic and real data. The latter is a simpler version of the autoencoder-based algorithm proposed in this paper, which
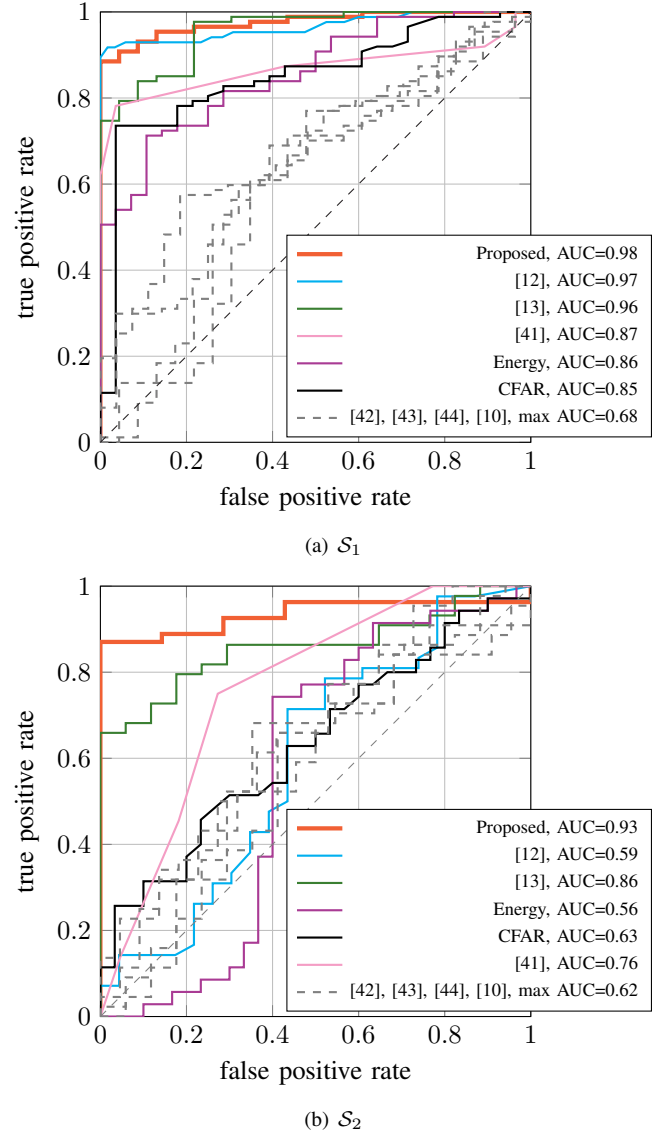


(a) $\mathcal{S}_1$



(b) $\mathcal{S}_2$

Fig. 11. ROC curves of the proposed methodology compared with state-of-the-art solutions on dataset $\mathcal{S}_1$ (a) and $\mathcal{S}_2$ (b).

does not take into account neither multi-polarization, nor volumetric data. Both algorithms have been trained as explained in the original papers, using their best parameters and the same training data.

From this comparison we can conclude that the proposed architecture has competing performances against other state-of-the-art solutions. This is an additional pro, considering that we do not even need a great amount of training data.

## VII. CONCLUSION AND FUTURE WORKS

In this paper, a novel technique for landmine detection by means of multi-polarization GPR acquisitions has been introduced. The proposed system (i.e., architecture $\mathcal{A}_2^{\text{3D}}$ trained on multiple polarization $\mathbf{V}_A$ data) is able to detect buried threats in a volumetric dataset with an AUC of 98%. In particular, it reaches a detection probability of 90% also when the desired false positive rate is zero (i.e., we do not detect any non-mine object). A special convolutional neural network,

the *autoencoder*, is employed as anomaly detector. This is trained to reconstruct background soil only, i.e., without any hyperbola whose electro-magnetic signature is similar to those of landmines. The training dataset can be very small (i.e., just a few B-scans) and the convergence of the training is reached after a few epochs (i.e., a few minutes on a modern GPU). Moreover, the system has shown promising performance also in a cross-dataset scenario (i.e., it has been trained on a dataset and then deployed on another dataset). As a suggestion, the proposed architecture can be trained on a dataset and then refined on the area of investigation, whose conditions may vary day by day due to weather conditions and other external causes.

The proposed study enabled us to investigate different aspects about the use of CNNs for GPR data analysis within the context of buried targets detection. An interesting finding is that it is possible to solve the considered problem using an autoencoder in an anomaly detection fashion, rather than a classical CNN designed for classification purpose. This is important as the autoencoder enabled us to work with a limited amount of labeled data. Moreover, only one class of data needs to be labeled (i.e., the background not containing any target). Another interesting finding is that the use of 3D data actually helped improving detection performance compared to a similar solution exploiting 2D data only. This may further motivates the development of tools enabling full 3D acquisitions (C-scans), rather than splicing together multiple 2D ones (B-scans). Finally, we also investigated the effect of using multiple-polarization acquisitions. Concerning this aspect, it appears that using both horizontal and vertical polarization can improve the achieved accuracy, albeit it is not a fundamental requirement. As a matter of fact, the improved performance probably has to do mainly with the geometric characteristics of the buried objects, which can be non perfectly aligned with respect to a single polarization. Therefore, "observing" the buried landmine with different polarization can be beneficial, but only to a certain extent.

This work can be considered as an additional step toward automated humanitarian demining systems. In this scenario, we are currently improving the proposed methodology for counting and localizing the threats. Once geometric information about the buried threats is available, a classifier could be used to discriminate the landmines from inert buried objects.

## REFERENCES

[1] I. C. to ban Landmines, "Landmine Monitor 2017," 2017.
[2] "International Mine Action Standards," 2001. [Online]. Available: https://www.mineactionstandards.org/
[3] J. N. Wilson, P. Gader, W. Lee, H. Frigui, and K. C. Ho, "A large-scale systematic evaluation of algorithms using ground-penetrating radar for landmine detection and discrimination," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 8, pp. 2560–2572, Aug 2007.
[4] J. M. Malof, D. Reichman, A. Karem, H. Frigui, K. C. Ho, J. N. Wilson, W. Lee, W. J. Cummings, and L. M. Collins, "A large-scale multi-institutional evaluation of advanced discrimination algorithms for buried threat detection in ground penetrating radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 57, no. 9, pp. 6929–6945, Sep. 2019.
[5] X. Zhou, H. Chen, and J. Li, "An Automatic GPR B-Scan Image Interpreting Model," *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 56, no. 6, pp. 3398–3412, 2018.
[6] A. Dell'Acqua, A. Sarti, S. Tubaro, and L. Zanzi, "Detection of linear objects in GPR data," *Signal Processing*, vol. 84, no. 4, pp. 785–799, 2004.
[7] H. Liu, Z. Long, B. Tian, F. Han, G. Fang, and Q. H. Liu, "Two-Dimensional Reverse-Time Migration Applied to GPR with a 3-D-to-2-D Data Conversion," *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing (J-STARS)*, vol. 10, no. 10, pp. 4313–4320, 2017.
[8] C. Maas and J. Schmalzl, "Using pattern recognition to automatically localize reflection hyperbolas in data from ground penetrating radar," *Computers & Geosciences*, vol. 58, pp. 116–125, 2013.
[9] J. M. Malof, D. Reichman, and L. M. Collins, "How do we choose the best model? The impact of cross-validation design on model evaluation for buried threat detection in ground penetrating radar," in *Detection and Sensing of Mines, Explosive Objects, and Obscured Targets*, 2018.
[10] Q. Dou, L. Wei, D. R. Magee, and A. G. Cohn, "Real-Time Hyperbola Recognition and Fitting in GPR Data," *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 55, no. 1, pp. 51–62, 2017.
[11] X. Núñez-Nieto, M. Solla, P. Gómez-Pérez, and H. Lorenzo, "GPR signal characterization for automated landmine and UXO detection based on machine learning techniques," *Remote Sensing*, vol. 6, no. 10, pp. 9729–9748, 2014.
[12] S. Lameri, F. Lombardi, P. Bestagini, M. Lualdi, and S. Tubaro, "Landmine Detection from GPR Data Using Convolutional Neural Networks," in *European Signal Processing Conference (EUSIPCO)*, 2017.
[13] F. Picetti, G. Testa, F. Lombardi, P. Bestagini, M. Lualdi, and S. Tubaro, "Convolutional Autoencoder for Landmine Detection on GPR Scans," in *IEEE Telecommunications and Signal Processing (TSP)*, 2018.
[14] J. He, S. Misra, and H. Li, "Comparative Study of Shallow Learning Models for Generating Compressional and Shear Traveltime Logs," *Petrophysics*, vol. 59, no. 6, 2018.
[15] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, http://www.deeplearningbook.org.
[16] G. Alregib, M. Deriche, Z. Long, H. Di, Z. Wang, Y. Alaudah, M. A. Shafiq, and M. Alfarraj, "Subsurface Structure Analysis Using Computational Interpretation and Learning," *IEEE Signal Processing Magazine (SPM)*, vol. 35, no. 2, pp. 82–98, 2018.
[17] T. Zhao, V. Jayaram, A. Roy, and K. J. Marfurt, "A comparison of classification techniques for seismic facies recognition," *Interpretation*, vol. 3, no. 4, 2015.
[18] P. Bestagini, V. Lipari, and S. Tubaro, "A machine learning approach to facies classification using well logs," in *Society of Exploration Geophysicists International Exposition and Annual Meeting (SEG)*, 2017.
[19] H. Li, J. He, and S. Misra, "Data-Driven In-Situ Geomechanical Characterization in Shale Reservoirs," in *Society of Petroleum Engineers (SPE) Annual Technical Conference and Exhibition*, 2018.
[20] T. Zhao and P. Mukhopadhyay, "A fault-detection workflow using deep learning and image processing," in *SEG Technical Program Expanded Abstracts 2018*, 2018, pp. 1966–1970.
[21] M. A. Shafiq, M. Prabhushankar, H. Di, and G. AlRegib, "Towards understanding common features between natural and seismic images," in *SEG Technical Program Expanded Abstracts 2018*, 2018, pp. 2076–2080.
[22] S. M. Mousavi, W. Zhu, W. Ellsworth, and G. Beroza, "Unsupervised clustering of seismic signals using deep convolutional autoencoders," *IEEE Geoscience and Remote Sensing Letters*, pp. 1–5, 2019.
[23] Y. Jin, X. Wu, J. Chen, Z. Han, and W. Hu, "Seismic data denoising by deep-residual networks," in *SEG Technical Program Expanded Abstracts 2018*, 2018, pp. 4593–4597.
[24] K. Wang, G. Zhang, Y. Leng, and H. Leung, "Synthetic aperture radar image generation with deep generative models," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 6, pp. 912–916, June 2019.
[25] J. Geng, J. Fan, H. Wang, X. Ma, B. Li, and F. Chen, "High-resolution sar image classification via deep convolutional autoencoders," *IEEE Geoscience and Remote Sensing Letters*, vol. 12, no. 11, pp. 2351–2355, Nov 2015.
[26] K. H. Jin, M. T. McCann, E. Froustey, and M. Unser, "Deep Convolutional Neural Network for Inverse Problems in Imaging," *IEEE Transactions on Image Processing (TIP)*, vol. 26, no. 9, pp. 4509–4522, 2017.
[27] M. T. McCann, K. H. Jin, and M. Unser, "Convolutional neural networks for inverse problems in imaging: A review," *IEEE Signal Processing Magazine (SPM)*, vol. 34, no. 6, pp. 85–95, 2017.
[28] F. Picetti, V. Lipari, P. Bestagini, and S. Tubaro, "A generative adversarial network for seismic imaging applications," in *Society of Exploration Geophysicists International Exposition and Annual Meeting (SEG)*, 2018.

[29] S. Mandelli, F. Borra, V. Lipari, P. Bestagini, A. Sarti, and S. Tubaro, "Seismic Data Interpolation Through Convolutional Autoencoder," in *Society of Exploration Geophysicists International Exposition and Annual Meeting (SEG)*, 2018.

[30] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *AACM Computing Surveys (CSUR)*, vol. 41, no. 3, pp. 15:1–15:58, 2009.

[31] T. Smith and S. Treitel, "Self-organizing artificial neural nets for automatic anomaly identification," *Society of Exploration Geophysicists Technical Program Expanded Abstracts*, vol. 3, no. 2, pp. 1403–1407, 2010.

[32] Y. Koizumi, S. Saito, H. Uematsu, Y. Kawachi, and N. Harada, "Unsupervised detection of anomalous sound based on deep learning and the neyman–pearson lemma," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 1, pp. 212–224, Jan 2019.

[33] A. Dairi, F. Harrou, Y. Sun, and M. Senouci, "Obstacle Detection for Intelligent Transportation Systems Using Deep Stacked Autoencoder and $k$ -Nearest Neighbor Scheme," *IEEE Sensors Journal*, vol. 18, no. 12, pp. 5122–5132, June 2018.

[34] D. Singh and C. K. Mohan, "Deep spatio-temporal representation for detection of road accidents using stacked autoencoder," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 3, pp. 879–887, March 2019.

[35] T. Wang, M. Qiao, Z. Lin, C. Li, H. Snoussi, Z. Liu, and C. Choi, "Generative neural networks for anomaly detection in crowded scenes," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 5, pp. 1390–1399, May 2019.

[36] D. Park, Y. Hoshi, and C. C. Kemp, "A multimodal anomaly detector for robot-assisted feeding using an lstm-based variational autoencoder," *IEEE Robotics and Automation Letters*, vol. 3, no. 3, pp. 1544–1551, July 2018.

[37] S. Park, M. Kim, and S. Lee, "Anomaly detection for http using convolutional autoencoders," *IEEE Access*, vol. 6, pp. 70 884–70 901, 2018.

[38] P. Seeböck, S. M. Waldstein, S. Klimscha, H. Bogunovic, T. Schlegl, B. S. Gerendas, R. Donner, U. Schmidt-Erfurth, and G. Langs, "Unsupervised identification of disease marker candidates in retinal oct imaging data," *IEEE Transactions on Medical Imaging*, vol. 38, no. 4, pp. 1037–1047, April 2019.

[39] Y. Su, A. Marinoni, J. Li, J. Plaza, and P. Gamba, "Stacked nonnegative sparse autoencoders for robust hyperspectral unmixing," *IEEE Geoscience and Remote Sensing Letters*, vol. 15, no. 9, pp. 1427–1431, Sep. 2018.

[40] Y. Yamashita, T. Kamoshita, Y. Akiyama, H. Hattori, Y. Kakishita, T. Sadaki, and H. Okazaki, "Improving efficiency of cavity detection under paved road from gpr data using deep learning method," in *The 13th SEGJ International Symposium, Tokyo, Japan, 12-14 November 2018*, 2019, pp. 526–529.

[41] P. A. Torrione, K. D. Morton, R. Sakaguchi, and L. M. Collins, "Histograms of oriented gradients for landmine detection in ground-penetrating radar data," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 52, no. 3, pp. 1539–1550, 2014.

[42] R. T. Sakaguchi, K. D. Morton, L. M. Collins, and P. A. Torrione, "Recognizing subsurface target responses in ground penetrating radar data using convolutional neural networks," in *SPIE 9454, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XX,*, 2015.

[43] L. E. Besaw and P. J. Stimac, "Deep learning algorithms for detecting explosive hazards in ground penetrating radar data," in *SPIE 9072, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIX*, 2014.

[44] ——, "Deep convolutional neural networks for classifying gpr b-scans," in *SPIE 9454, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XX*, 2015.

[45] D. Reichman, L. M. Collins, and J. M. Malof, "Some good practices for applying convolutional neural networks to buried threat detection in ground penetrating radar," in *IEEE International Workshop on Advanced Ground Penetrating Radar (IWAGPR)*, 2017.

[46] V. Kafedziski, S. Pecov, and D. Tanevski, "Detection and Classification of Land Mines from Ground Penetrating Radar Data Using Faster R-CNN," in *IEEE Telecommunications Forum (TELFOR)*, 2018.

[47] c. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016.

[48] H. Li and S. Misra, "Long Short-Term Memory and Variational Autoencoder With Convolutional Neural Networks for Generating NMR T2 Distributions," *IEEE Geoscience and Remote Sensing Letters*, vol. 16, no. 2, pp. 192–195, 2018.

[49] ——, "Prediction of Subsurface NMR T2 Distributions in a Shale Petroleum System Using Variational Autoencoder-Based Neural Networks," *IIEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 12, pp. 2395–2397, 2017.

[50] Y. Han and S. Misra, "Bakken Petroleum System Characterization Using Dielectric-Dispersion Logs," *Petrophysics*, vol. 59, no. 2, pp. 201–217, 2018.

[51] M. Lualdi and F. Lombardi, "Significance of GPR polarisation for improving target detection and characterisation," *Nondestructive Testing and Evaluation*, vol. 29, no. 4, pp. 345–356, 2014.

[52] A. Villela and J. M. Romo, "Invariant properties and rotation transformations of the GPR scattering matrix," *Journal of Applied Geophysics*, vol. 90, pp. 71–81, 2013.

[53] F. Lehmann, D. E. Boerner, K. Holliger, and A. G. Green, "Multicomponent georadar data: Some important implications for data acquisition and processing," *Geophysics*, vol. 65, no. 5, pp. 1542–1552, 2000.

[54] F. Lombardi and M. Lualdi, "Multi-azimuth ground penetrating radar surveys to improve the imaging of complex fractures," *Geosciences*, vol. 8, no. 11, 2018. [Online]. Available: https://www.mdpi.com/2076-3263/8/11/425

[55] S. J. Radzevicius and J. J. Daniels, "Ground penetrating radar polarization and scattering from cylinders," *Journal of Applied Geophysics*, vol. 45, no. 6, pp. 111–125, 2000.

[56] P. Lutz, S. Garambois, and H. Perroud, "Influence of antenna configurations for GPR survey: information from polarization and amplitude versus offset measurements," *Geological Society Special Publications*, vol. 211, pp. 299–313, 2003.

[57] M. Lualdi and F. Lombardi, "Utilities detection through the sum of orthogonal polarization in 3D georadar surveys," *Near Surface Geophysics*, vol. 13, no. 1, pp. 73–81, 2015.

[58] ——, "Combining orthogonal polarization for elongated target detection with GPR," *Journal of Geophysics and Engineering*, vol. 11, no. 5, p. 055006, 2014.

[59] J. J. Daniels, L. Wielopolski, S. J. Radzevicius, and J. Bookshar, "3D GPR Polarization Analysis for Imaging Complex Objects," in *Symposium on the Application of Geophysics to Engineering and Environmental Problems (SAGEEP)*, 2003.

[60] G. P. Tsoflias and A. Hoch, "Investigating multi-polarization GPR wave transmission through thin layers: Implications for vertical fracture characterization," *Geophysical Research Letters*, vol. 33, no. 20, 2006.

[61] R. Roberts, J. J. Daniels, and L. Peters, "Improved GPR Interpretation from Analysis of Buried Target Polarization Properties," in *Symposium on the Application of Geophysics to Engineering and Environmental Problems (SAGEEP)*, 1992.

[62] J. Leckebursch, "Problems and Solutions with GPR Data Interpretation: Depolarization and Data Continuity," *Archaeological Prospection*, vol. 18, no. 4, pp. 303–308, 2011.

[63] M. Lualdi and F. Lombardi, "Effects of antenna orientation on 3-D ground penetrating radar surveys: an archaeological perspective," *Geophysical Journal International*, vol. 196, no. 2, pp. 818–827, 2014.

[64] T. Wang, J. M. Keller, P. D. Gader, and O. Sjahputera, "Frequency subband processing and feature analysis of forward-looking ground-penetrating radar signals for land-mine detection," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 45, no. 3, pp. 718–729, March 2007.

[65] J. A. Camilo, L. M. Collins, and J. M. Malof, "A large comparison of feature-based approaches for buried target classification in forward-looking ground-penetrating radar," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 547–558, Jan 2018.

[66] R. Bello, "Literature Review on Landmines and Detection Methods," *Frontiers in Science*, vol. 3, no. 1, pp. 27–42, 2013.

[67] D. Cozzolino and L. Verdoliva, "Single-image splicing localization through autoencoder-based anomaly detection," in *IEEE International Workshop on Information Forensics and Security (WIFS)*, 2016.

[68] S. K. Yarlagadda, D. Güera, P. Bestagini, F. M. Zhu, S. Tubaro, and E. J. Delp, "Satellite Image Forgery Detection and Localization Using GAN and One-Class Classifier," *IS&T Electronic Imaging (EI)*, 2018.

[69] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," in *International Conference on Learning Representations (ICLR)*, 2015.

[70] M. Lualdi and L. Zanzi, "The PSG, a New Positioning System to Execute 3D GPR Surveys for Utility Mapping," in *Symposium on the Application of Geophysics to Engineering and Environmental Problems (SAGEEP)*, 2003.

[71] D. Reichman, L. M. Collins, and J. M. Malof, "On Choosing Training and Testing Data for Supervised Algorithms in Ground-Penetrating Radar Data for Buried Threat Detection," *IEEE Transactions on Geoscience and Remote Sensing (TGRS)*, vol. 56, no. 1, pp. 497–507, 2017.

**Paolo Bestagini** (M'11) was born in Novara, Italy, on February 22, 1986. He received the M.Sc. degree in Telecommunications Engineering and the Ph.D. degree in Information Technology from the Politecnico di Milano, Italy, in 2010 and 2014, respectively. He is currently an Assistant Professor at the Image and Sound Processing Group, Politecnico di Milano. His research interests focus on multimedia forensics and acoustic signal processing for microphone arrays. He is an elected member of the IEEE Information Forensics and Security Technical Committee, and a co-organizer of the IEEE Signal Processing Cup 2018.

**Federico Lombardi** (GSM'17) was born in Piacenza, Italy on August 18, 1987. He received the B.Sc. and the M.Sc. degree in Telecommunications Engineering from the Politecnico di Milano, Italy, in 2009 and 2011, respectively. He joined the Radar research group at University College of London, United Kingdom, in 2015 as a Ph.D. candidate and he is currently defending his thesis on the topic "Novel Radar Techniques for Humanitarian Demining. Since July, he is a Research Fellow at the Department of Civil and Environmental Engineering working on Ground Penetrating Radar investigation for soil characterization. His research interests focus on non destructive testing of structures and near surface geophysical investigation. He is a nominated member of the IEEE AESS Board of Governors where he serves as the Graduate Student Representative and a member of the SEG Italian Section.

**Maurizio Lualdi** was born in Busto Arsizio, Italy in 1973. He completed his studies in Environmental Engineering at the Politecnico di Milano, Italy. He is Associate Professor of Applied Geophysics at the Dipartimento di Ingegneria Civile e Ambientale of Politecnico di Milano. In the last years he has focused his research interests on the development of full polarimetric 3D Georadar surveys.

**Francesco Picetti** (S'18) was born in Milan, Italy, in 1992. He received the B.Sc. degree in Electronic Engineering and M.Sc. degree in Computer Science and Engineering from the Politecnico di Milano, Italy, in 2015 and 2017, respectively. He is currently a Ph.D. student in Information Technology working with the Image and Sound Processing Group, Politecnico di Milano. His research interests focus on signal processing techniques for geophysical imaging.

**Stefano Tubaro** (SM'01) was born in Novara, Italy, in 1957. He completed his studies in Electronic Engineering at the Politecnico di Milano, Milan, Italy, in 1982. He then joined the Dipartimento di Elettronica, Informazione e Bioingegneria of the Politecnico di Milano, first as a Researcher of the National Research Council, and then (in November 1991) as an Associate Professor. Since December 2004, he has been appointed as a Full Professor of telecommunication at the Politecnico di Milano. His current research interests include advanced algorithms for video and sound processing. He is the author of more than 180 scientific publications on international journals and congresses and the coauthor of more than 15 patents. In the past few years, he has focused his interest on the development of innovative techniques for image and video tampering detection and, in general, for the blind recovery of the "processing history" of multimedia objects. He coordinates the research activities of the Image and Sound Processing Group at the Dipartimento di Elettronica, Informazione e Bioingegneria, Politecnico di Milano. He had the role of Project Coordinator of the European Project ORIGAMI (A new paradigm for high-quality mixing of real and virtual) and of the research project ICT-FET-OPEN REWIND (REVerse engineering of audio-VIsual coNtent Data). This last project was aimed at synergistically combining principles of signal processing, machine learning, and information theory to answer relevant questions on the past history of such objects. He is a member the IEEE Multimedia Signal Processing Technical Committee and of the IEEE SPS Image Video and Multidimensional Signal Technical Committee. He was in the organization committee of a number of international conferences including the IEEE MMSP 2004/2013, IEEE ICIP 2005, IEEE AVSS 2005/2009, IEEE ICDSC 2009, IEEE MMSP 2013, IEEE ICME 2015. From May 2012 to April 2015, he was an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING, and is currently an Associate Editor of the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY.