Dictionary Learning for Adaptive GPR Landmine Classification

Fabio Giovanneschi, Kumar Vijay Mishra, *Senior Member, IEEE*, Maria Antonia Gonzalez-Huici, Yonina C. Eldar, *Fellow, IEEE*, and Joachim H. G. Ender, *Fellow, IEEE*

***Abstract*— Ground-penetrating radar (GPR) target detection and classification is a challenging task. Here, we consider online dictionary learning (DL) methods to obtain sparse representations (SR) of the GPR data to enhance feature extraction for target classification via support vector machines. Online methods are preferred because traditional batch DL like K-times singular value decomposition (K-SVD) is not scalable to high-dimensional training sets and infeasible for real-time operation. We also develop Drop-Off MINi-batch Online Dictionary Learning (DOMINODL), which exploits the fact that a lot of the training data may be correlated. The DOMINODL algorithm iteratively considers elements of the training set in small batches and drops off samples which become less relevant. For the case of abandoned anti-personnel landmines classification, we compare the performance of K-SVD with three online algorithms: classical online dictionary learning (ODL), its correlation-based variant, and DOMINODL. Our experiments with real data from L-band GPR show that online DL methods reduce learning time by 36%–93% and increase mine detection by 4%–28% over K-SVD. Our DOMINODL is the fastest and retains similar classification performance as the other two online DL approaches. We use a Kolmogorov–Smirnoff test distance and the Dvoretzky–Kiefer–Wolfowitz inequality for the selection of DL input parameters leading to enhanced classification results. To further compare with the state-of-the-art classification approaches, we evaluate a convolutional neural network (CNN) classifier, which performs worse than the proposed approach. Moreover, when the acquired samples are randomly reduced by 25%, 50%, and 75%, sparse decomposition-based classification with DL remains robust while the CNN accuracy is drastically compromised.**

***Index Terms*— Adaptive radar, deep learning, ground-penetrating radar (GPR), online dictionary learning (ODL), radar target classification, sparse decomposition.**

Manuscript received May 24, 2018; revised April 18, 2019 and June 11, 2019; accepted July 14, 2019. Date of publication August 21, 2019; date of current version November 25, 2019. The work of K. V. Mishra was supported in part by the Lady Davis Post-Doctoral Fellowship and in part by the Andrew and Erna Finci Viterbi Postdoctoral Fellowship. This article was presented at the IEEE International Geoscience and Remote Sensing Symposium, 2017 [12]. *(Corresponding author: Fabio Giovanneschi.)*

F. Giovanneschi and M. A. Gonzalez-Huici are with the Fraunhofer Institute for High Frequency Physics and Radar Techniques, 53343 Wachtberg, Germany (e-mail: fabio.giovanneschi@fhr.fraunhofer.de; maria.gonzalez@fhr.fraunhofer.de).

K. V. Mishra is with the Department of Electrical and Computer Engineering, The University of Iowa, Iowa City, IA 52242 USA (e-mail: kumarvijay-mishra@uiowa.edu).

Y. C. Eldar is with the Weizmann Institute of Science, Rehovot 7610001, Israel (e-mail: yonina.eldar@weizmann.ac.il).

J. H. G. Ender is with the Centre for Sensor Systems (ZESS), University of Siegen, 57076 Siegen, Germany (e-mail: ender@zess.uni-siegen.de).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TGRS.2019.2931134

I. Introduction

AGROUND penetrating radar (GPR) is used for probing the underground by transmitting radio waves from an antenna held closely to the surface and acquiring the echoes reflected from subsurface anomalies or buried objects. As the electromagnetic wave travels through the subsurface, its velocity changes due to the physical properties of the materials in the medium. By recording such changes in the velocity and measuring the travel time of the radar signals, a GPR generates profiles of scattering responses from the subsurface. The interest in GPR is due to its ability to reveal buried objects non-invasively and detect non-metallic scatterers with increased sensitivity to dielectric contrast [1]. From the recordings of the previously observed regions, GPR surveys can also extrapolate subsurface knowledge for inaccessible or unexcavated areas. This sensing technique is therefore attractive for several applications, such as geophysics, archeology, forensics, and defense (see [1], [2] for some surveys). Over the last decade, there has been a spurt in GPR research because of advances in electronics and computing resources. GPR has now surpassed traditional ground applications and has become a more general ultra-wideband (UWB) remote sensing system with proliferation to novel avenues, such as through-the-wall imaging, building construction, food safety monitoring, and vegetation observation.

In this paper, we consider the application of detecting buried landmines using GPR. This is one of the most extensively investigated GPR applications due to its obvious security and humanitarian importance [3]. Mine detection GPR usually operates in the L-band (1–2 GHz) with UWB transmission in order to achieve sufficient resolution to detect small targets (5–10-cm diameter) and penetrate solid media at shallow depths (15–30 cm) [4].

Even though a lot of progress has been made on GPR for landmine detection, discriminating them from natural and man-made clutter remains a critical challenge. In such applications, the signal distortion due to inhomogeneous soil clutter, surface roughness, and antenna ringing hampers target recognition. Moreover, the constituting material of many models of landmines is largely plastic and has a very weak response to radar signals due to its low dielectric contrast with respect to the soil [2]. Finally, a major problem arises due to low radar cross section (RCS) of some landmine models [5]. A variety of signal processing algorithms have been proposed for the

detection of low metal-content landmines in realistic scenarios; approaches based on feature extraction and classification are found to be the most successful [6]–[8], yet false-alarm rates remain very high.

Sparse representation (SR) is effective in extracting the mid- or high-level features in image classification [9], [10]. In the context of anti-personnel landmines recognition using GPR, our prior work [11], [12] has shown that frameworks based on SR improve the performance of support vector machine (SVM) classifiers in distinguishing different types of mines and clutter in highly corrupted GPR signals. In this approach, the signal-of-interest is transformed into a domain where it can be expressed as a linear combination of only a few *atoms* chosen from a collection called the *dictionary* matrix [13], [14]. The dictionary may be learned from the data it is going to represent. *Dictionary learning* (DL) techniques (or *sparse coding* in machine learning parlance) aim to create adaptive dictionaries, which provide the sparsest reconstruction for given training sets, i.e., a representation with a minimum number of constituting atoms. DL methods are critical building blocks in many applications, such as deep learning, image denoising, and super-resolution (see [15]–[17] for further applications.)

Classical DL algorithms, such as method of optimal directions (MOD) [18] and K-times singular value decomposition (K-SVD) [13], operate in batches, dealing with the entire training set in each iteration. Although extremely successful, these methods are computationally demanding and not scalable to high-dimensional training sets. An efficient alternative is the online dictionary learning (ODL) algorithm [19] that has faster convergence than batch DL methods. In this paper, we develop a new approach toward classification based on online DL-SR framework for the specific case of GPR-based landmine identification.

The main contributions of this paper are as follows.

A. Faster DL for GPR Landmine Classification

We investigate the application of DL toward GPR-based landmine classification. To the best of our knowledge, this has not been investigated previously. Furthermore, online DL[1] methods have been studied more generally in GPR. Only one other previous study has employed DL (K-SVD) using GPR signals [20], although for the application of identifying bedrock features. We employ online DL methods and use the coefficients of the resulting sparse vectors as input to an SVM classifier to distinguish mines from clutter. Our comparison of K-SVD and online DL using real data from L-band GPR shows that online DL algorithms present distinct advantages in speed and low false-alarm rates. We propose a new Drop-Off MINi-batch Online Dictionary Learning (DOMINODL), which processes the training data in mini-batches and avoids unnecessary update of the irrelevant atoms in order to reduce the computational complexity. The intuition for the drop-off step comes from the fact that some training samples are highly correlated and, therefore, in the interest of processing time, they can be dropped during training without significantly affecting performance.

[1]We use the term "online DL" to imply any algorithm that operates in online mode. From here on, we reserve the term ODL solely to refer to the method described in [19].

B. Better Statistical Metrics for Improved Classification

Contrary to previous studies [20] which determine DL parameters (the number of iterations, atoms, and so on) based on bulk statistics, such as normalized root-mean-square error (NRMSE), we consider statistical inference for parameter analysis. Our methods based on Kolmogorov–Smirnoff (K-S) test distance [21] and Dvoretzky–Kiefer–Wolfowitz (DKW) inequality [22], [23] are able to fine-tune model selection resulting in improved mine classification performance.

C. Experimental Validation for Different Landmine Sizes

Our comparison of K-SVD with three online DL algorithms—ODL, its correlation-based variant [24], and DOMINODL—shows that online methods successfully detect mines with very small RCS buried deep into clutter and noise. Some recent studies [25]–[28] employ the state-of-the-art deep learning approaches such as a convolutional neural network (CNN) to classify GPR-based mines data. Our comparison with CNN illustrates that it has poorer performance in the detection of small mines than our online DL approaches. This may also be caused by the relatively small dimensions of our training set which, even if perfectly adequate for DL, may not meet the expected requirements for a CNN [29]. We also show that the classification performance of online DL methods does not deteriorate significantly when signal samples are reduced.

The rest of this paper is organized as follows. In Section II, we formally describe the classification problem and GPR specific challenges. In Section III, we explain various DL algorithms used in our methodology and also describe DOMIN-ODL. We provide an overview of the GPR system and field campaign to collect GPR data sets in Section IV. In Section V, we introduce our techniques for DL parameter selection. Section VI presents classification and reconstruction results using real radar data. We conclude in Section VII.

Throughout this paper, we reserve boldface lowercase and uppercase letters for vectors and matrices, respectively. The ith element of a vector $\mathbf{y}$ is $\mathbf{y}_i$, while the (i, j)th entry of the matrix $\mathbf{Y}$ is $\mathbf{Y}_{i,j}$. We denote the transpose by $(\cdot)^T$. We represent the set of real and complex numbers by $\mathbb{R}$ and $\mathbb{C}$, respectively. Other sets are represented by calligraphic letters. The notation $\|\cdot\|_p$ stands for the p-norm of its argument and $\|\cdot\|_F$ is the Frobenius norm. A subscript in the parenthesis such as $(\cdot)_{(t)}$ is the value of the argument in the tth iteration. The convolution product is denoted by $*$. The function diag$(\cdot)$ outputs a diagonal matrix with the input vector along its main diagonal. We use Pr$\{\cdot\}$ to denote probability, E$\{\cdot\}$ is the statistical expectation, and $|\cdot|$ denotes the absolute value. The functions max$(\cdot)$ and sup$(\cdot)$ output the maximum and supremum value of their arguments, respectively.

II. Problem Formulation

A pulsed GPR transmits a signal into the ground and receives its echo for each point in the radar coverage area. The digital samples of this echo constitute a *range profile* designated by the signal vector $\mathbf{y} \in \mathbb{R}^M$, where M is the

number of range cells. Formally, the SR of $\mathbf{y}$ can be described by $\mathbf{y} = \mathbf{D}\mathbf{x}$, where $\mathbf{D} = [\mathbf{d}_1, \cdots, \mathbf{d}_N] \in \mathbb{R}^{M\times N}$ with column vectors or *atoms* $\{\mathbf{d}_i\}$ is a redundant or overcomplete ($M \ll N$) dictionary, and $\mathbf{x} \in \mathbf{R}^N$ is the SR vector. The SR process finds the decomposition that uses the minimum number of atoms to express the signal. If there are L range profiles available, then the SR of the data $\mathbf{Y} = [\mathbf{y}_1, \cdots, \mathbf{y}_L]$ is described as $\mathbf{Y} = \mathbf{D}\mathbf{X}$, where $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_L] \in \mathbb{R}^{N\times L}$.

Our goal is to classify different mines (including the ones with small RCS) and clutter based on the SR of range profiles using fast, online DL methods. The GPR range profiles from successive scans are highly correlated. We intend to exploit this property during the DL process. In the following, we describe the SR-based classification and mention its challenges.

A. GPR Target Classification Method

We use SVM to classify sparsely represented GPR range profiles using the learned dictionary $\mathbf{D}$. Given a predefined collection of labeled observations, SVM searches for a functional $f : \mathbb{R}^n \to \mathbb{R}$ that maps any new observation to a class $c \in \mathbb{R}$. A binary classifier with linearly separable data, for example, would have $c \in \{1, -1\}$. In this paper, we use $\mathbf{X}$, i.e., the sparse decomposition of a given set of signals (the "training" signals) $\mathbf{Y}$ using the learned dictionary $\mathbf{D}$, as a set of labeled observations for the SVM. SVM transforms the data into a high-dimensional feature space, where it is easier to separate between different classes. The kernel function that we use to compute the high-dimensional operations in the feature space is the Gaussian radial basis function (RBF): $\kappa(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma\,(||\mathbf{x}_i - \mathbf{x}_j||)^2 + C)$, where $\gamma > 0$ is the free parameter that decides the influence of the support data vector $\mathbf{x}_j$ on the class of the vector $\mathbf{x}_i$ in the original space and C is the parameter for the soft margin cost function, which controls the influence of each individual support vector [30]. To optimally select the SVM input parameters, we arrange the original classification set into training and validation vectors in ν different ways (ν-fold cross validation with $\nu = 10$) to arrive at a certain mean cross-classification accuracy of the validation vectors. The folds were randomly selected and their number was found empirically by determining the limit above which the accuracy improvement was negligible. We refer the reader to [30] for more details on SVM.

B. Basic Framework of DL

In many applications, the dictionary $\mathbf{D}$ is unknown and has to be learned from the training signals coming from the desired class. A DL algorithm finds an overcomplete dictionary $\mathbf{D}$ that sparsely represents measurements $\mathbf{Y}$ such that $\mathbf{Y} \simeq \mathbf{D}\mathbf{X}$. Each of the vectors $\mathbf{x}_i$ is an SR of $\mathbf{y}_i$ with only K_s nonzero entries. A non-tractable formulation of this problem is

$$\min_{\mathbf{D},\mathbf{X}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}\|_F \quad \text{s.t. } \|\mathbf{x}_i\|_0 \le K_s, \quad 1 \le i \le L. \tag{1}$$

Since both $\mathbf{D}$ and $\mathbf{X}$ are unknown, a common approach is to use alternating minimization in which we start with an initial guess of $\mathbf{D}$ and then obtain the solution iteratively by alternating between two stages—*sparse coding* [31] and *dictionary update* [32]—as follows:

1) *Sparse Coding*: Obtain $\mathbf{X}_{(t)}$ as

$$\mathbf{X}_{(t)} = \min_{\mathbf{X}} \|\mathbf{Y} - \mathbf{D}_{(t-1)}\mathbf{X}\|_F \quad \text{s.t. } \|\mathbf{x}_{i_{t-1}}\|_p \le K_s, \quad 1 \le i \le L \tag{2}$$

where $\mathbf{X}_{(t)}$ is the SR in tth iteration. This can be solved using greedy algorithms such as orthogonal matching pursuit (OMP) ($p = 0$), convex relaxation methods like basis pursuit denoising (BPDN) ($p = 1$), or a focal underdetermined system solver (FOCUSS) ($0 < p < 1$).

2) *Dictionary Update*: Given $\mathbf{X}_{(t)}$, update $\mathbf{D}$ such that

$$\mathbf{D}_{(t)} = \min_{\mathbf{D}\in\mathcal{D}} \|\mathbf{Y} - \mathbf{D}\mathbf{X}_{(t)}\|_F \tag{3}$$

where $\mathcal{D}$ is a set of all dictionaries with unit column norms, $\|\mathbf{d}_j\|_2 = 1$ for $1 \le j \le N$. This subproblem is solved by methods such as singular value decomposition or gradient descent [19], [31].

Classical methods such as MOD [18] and K-SVD [13] retain a guess for $\mathbf{D}$ and $\mathbf{X}$ and iteratively update either $\mathbf{X}$ using basis pursuit/matched pursuit (BP/MP) or $\mathbf{D}$ using a least squares solver. Both MOD and K-SVD operate in batches, i.e., they deal with the entire training set in each iteration, and solve the same DL model but differ in the optimization method. Since the initial guesses of $\mathbf{D}$ or $\mathbf{X}$ can be far removed from the actual dictionary, the BP step may behave erratically. While there are several state-of-the-art results that outline DL algorithms with concrete performance guarantees [16], [33], [34], they require stronger assumptions on the observed data. In practice, heuristic DL, such as MOD and K-SVD, does yield overcomplete dictionaries although provable guarantees for such algorithms are difficult to come by.

Several extensions of batch DL have been proposed, e.g., label consistent (LC) K-SVD [35] and discriminative K-SVD [36] introduced label information into the procedure of learning dictionaries to make them more discriminative. The performance of K-SVD can be improved in terms of both computational complexity and obtaining an incoherent dictionary if the learning process enforces constraints such as hierarchical tree sparsity [37], structured group sparsity (StructDL) [38], Fisher discrimination (FDDL) [39], and low-rank-and-Fisher ($D^2L^2R^2$) [40]. Often objects belonging to different classes have common features. This has been exploited in improving K-SVD to yield methods such as DL with structured incoherence and shared features (DLSI) [41], separating the commonality and the particularity (COPAR) [42], convolutional sparse DL (CSDL) [43], shift-invariant DL [44], principal component analysis DL [45], convolutional DL [46], and low-rank shared DL (LRSDL) [47]. A recent review of various DL algorithms can be found in [48].

In general, batch DL methods are computationally demanding at test time and not scalable to high-dimensional training sets. On the other hand, online methods such as ODL [19] converge fast and process small sets. A few improvements to ODL have already been proposed. For example, [49] considered a faster online sparse dictionary learning (OSDL)

to efficiently handle bigger training set dimensions using a double-sparsity model. A recent study [24] notes that even though online processing reduces computational complexity compared to batch methods, ODL performance can be further improved if the useful information from previous data is not ignored in updating the atoms. In this paper, a new online DL called correlation-based weighted least square update (CBWLSU) was proposed, which employs only part of the previous data correlated with current data for the update step. The CBWLSU is relevant to GPR because the latter often contains highly correlated range profiles.

In this paper, our focus is to investigate such fast DL methods in the context of GPR-based landmine detection and classification. We also propose a new online DL method that exploits range profile correlation as in CBWLSU but is faster than both ODL and CBWLSU. Our inspiration is the K-SVD variant called incremental structured DL (ISDL) that was used earlier in the context of SAR imaging [50]. In ISDL, at each iteration, a small batch of samples is randomly drawn from the training set. Let $\mathcal{R}_t \subset \{1, \cdots, L\}$ be the set of indices of the mini-batch training elements chosen uniformly at random at the tth iteration. Then, ISDL updates the dictionary $\mathbf{D}$ using the mini-batch $\mathbf{Y}_{\mathcal{R}_t} = \{\mathbf{y}_l : l \in \mathcal{R}_t\}$ and the corresponding representation coefficient $\mathbf{X}$. The fast iterative shrinkage-thresholding algorithm (FISTA) [51] and block coordinate descent methods solve the sparse coding and dictionary update, respectively. As we will see in Section VI, the mini-batch strategy that we employed in our DOMINODL reduces computational time without degrading performance.

III. Online DL

We now describe the DL techniques used for GPR target classification and then develop DOMINODL in order to address the challenges of long training times in the context of our problem.

A. K-SVD, LRSDL, ODL, and CBWLSU

As mentioned earlier, the popular K-SVD algorithm [52] sequentially updates all the atoms during the dictionary update step using all training set elements. For the sparse coding step at iteration t, K-SVD employs OMP with the formulation

$$\begin{aligned} &\min_{\mathbf{x}_i} \ \|\mathbf{x}_i\|_0 \\ &\text{s.t.} \ \ \|\mathbf{y}_i - \mathbf{D}_{(\mathbf{t-1})}\mathbf{x}_i\|_2^2 \leq \delta, \quad \forall 1 \leq i \leq L \end{aligned} \tag{4}$$

where δ is the maximum residual error used as a stopping criterion. For the dictionary update at iteration t, K-SVD solves the global minimization problem (3) via K sequential minimization problems, wherein every column $\mathbf{d}_k$ of $\mathbf{D}$ and its corresponding row of coefficients $\mathbf{X}_{\text{row},k}$ of $\mathbf{X}$ are updated as follows:

$$\begin{aligned} &\{\mathbf{X}_{\text{row},k_{(t)}}, \mathbf{d}_{k_{(t)}}\} \\ &= \min_{\mathbf{X}_{\text{row},k_{(t-1)}}, \mathbf{d}_k} \left\| \mathbf{Y} - \sum_{l \neq k} \mathbf{d}_{l_{t-1}} \mathbf{X}_{\text{row},l_{(t-1)}} - \mathbf{d}_k \mathbf{X}_{\text{row},k_{(t-1)}} \right\|_F . \end{aligned} \tag{5}$$

The update process employs SVD to find the closest rank-1 approximation (in Frobenius norm) of the error term $\mathbf{Y} - \sum_{l \neq k} \mathbf{d}_{l_{t-1}} \mathbf{X}_{\text{row},l_{(t-1)}}$ subject to the constraint $\|\mathbf{d}_{k_{(t)}}\|_2 = 1$.

Another recent batch method of interest is the LRSDL [47]. This is a discriminative batch DL algorithm (others being D-KSVD [36] and LC-KSVD [35]) that learns by promoting the generation of a dictionary $\mathbf{D}$ which is separated in blocks of atoms associated with different classes as $\mathbf{D} = [\mathbf{D}_1, \cdots, \mathbf{D}_C] \in \mathbb{R}^{M \times N}$, where C is the number of classes present in the training set $\mathbf{Y}$. The resultant coefficient matrix $\mathbf{X}$ has a block diagonal structure. The assumption of non-overlapping subspaces is often unrealistic in practice. Techniques, such as COPAR [42], joint dictionary learning [53], and CSDL [43], exploit common patterns among different classes even though different objects possess distinct class-specific features. These methods produce an additional constituent $\mathbf{D_0}$ which is shared among all classes so that $\mathbf{D} = [\mathbf{D}_1, \cdots, \mathbf{D}_C, \mathbf{D}_0] \in \mathbb{R}^{M \times N}$. The drawback of these strategies is that the shared dictionary may also contain class-discriminative features. To avoid this problem, LRSDL requires that the shared dictionary must have a low-rank structure and that its sparse coefficients have to be almost similar. Once the data are sparsely represented with such dictionaries, an SR-based classifier (SRC) is used to predict the class of new data. The LRSDL update process employs alternating direction method of multipliers (ADMM) [54] and FISTA for the sparse decomposition step.

ODL is an interesting alternative for inferring a dictionary from large training sets or ones which change over time [19]. ODL also updates the entire dictionary sequentially, but uses one element of training data at a time for the dictionary update. Assuming that the training set $\mathbf{Y}$ is composed of independent and identically distributed samples of a distribution $p(\mathbf{Y})$, ODL first draws an example of the training set $\mathbf{y}_t$ from $\mathbf{Y}$. Then, the sparse coding is the Cholesky-based implementation of the LARS-LASSO algorithm [55]. The latter solves a ℓ_1-regularized least-squares problem for each column of $\mathbf{Y}$. In the dictionary update, we consider all the training set elements analyzed so far, namely, $\mathbf{y}_i$ with $i = 1 \ldots t$

$$\mathbf{x}_t = \min_{\mathbf{x} \in \mathbb{C}^n} \frac{1}{2} ||\mathbf{y}_t - \mathbf{D}_{(\mathbf{t-1})}\mathbf{x}||_2^2 + \lambda ||\mathbf{x}||_1. \tag{6}$$

In the next step, each column of $\mathbf{D}$ is sequentially updated via gradient descent using the dictionary computed in the previous iteration. Before receiving the next training data, the dictionary update is repeated multiple times for convergence.

CBWLSU is an online method that introduces an interesting alternative for the dictionary update step [24]. Like ODL, CBWLSU evaluates one new training data $\mathbf{y}_i$. However, to update the dictionary, it searches among all previous training data and uses only the ones which share the same atoms with $\mathbf{y}_i$. Let $\mathbf{Y}_{\mathcal{Q}_t} = \{\mathbf{y}_l : l \in \mathcal{Q}_t\}$ with $\mathcal{Q}_t = \{\mathbf{y}_l : 1 \leq l < t-1\}$ be the set of previous training elements at iteration i. Define $\mathcal{N}_t = \{l : 1 < l < t, \langle \mathbf{x}_l^T \cdot \mathbf{x}_t \rangle \neq 0\} \subset \mathcal{Q}_t$ as the set of indices of all previous training elements that are correlated with the new element such that $|\mathcal{N}_t| = N_{p_t}$. The new training set is $\mathbf{Y}_{\mathcal{N}_t} = \{\mathbf{y}_l : l \in \mathcal{N}_t\} \cup \mathbf{y}_t$. Then, CBWLSU employs a weighting matrix $\mathbf{W}(\mathbf{y}_t)$ to evaluate the influence of the

selected previous elements $\mathbf{y}_t$ for the dictionary update step and solves the optimization problem therein via weighted least squares. Unlike K-SVD and ODL, CBWLSU does not require the dictionary pruning step to replace the unused or rarely used atoms with the training data. The sparse coding in CBWLSU is achieved via batch OMP.

B. Drop-Off Mini-Batch Online Dictionary Learning

We now introduce our DOMINODL approach for online DL, which not only leads to a dictionary ($\mathbf{D}$) that is tuned to sparsely represent the training set ($\mathbf{Y}$) but is also faster than other online algorithms. The key idea of DOMINODL is as follows: when sequentially analyzing the training set, it is pertinent to leverage the memory of previous data in the dictionary update step. However, algorithms such as CBWLSU consider *all* previous elements. Using all previous training set samples is computationally expensive and may also slow down convergence. The samples that have already contributed in the dictionary update do not need to be considered again. Moreover, in some real-time applications (such as highly correlated range profiles of GPR), their contribution may not be relevant anymore for updating the dictionary.

In DOMINODL, we save computations by considering only a small batch of previous elements that are correlated with the new elements. The two sets are defined correlated if, in their sparse decomposition, they have at least one common nonzero element. The time gained from considering fewer previous training elements is used to consider a mini-batch of new training data (instead of a single element as in ODL and CBWLSU). The sparse coding step of DOMINODL employs batch OMP, selecting the maximal residual error δ in (4) using a data-driven entropy-based strategy as described later in this section. At the end of each iteration, DOMINODL also drops-off those previous training set elements that have not been picked up after a certain number of iterations, N_u. The mini-batch drawing combined with dropping off training elements and entropy-based criterion to control sparsity results in an extremely fast online DL algorithm that is beneficial for real-time radar operations.

We initialize the dictionary $\mathbf{D}$ using a collection of K training set samples that are randomly chosen from $\mathbf{Y}$. We then perform a sparse decomposition of $\mathbf{Y}$ with the dictionary $\mathbf{D}$. Let the iteration count t indicate the tth element of the training set. We define the mini-batch of N_b new training elements as $\mathbf{Y}_{\mathcal{B}_t} = \{\mathbf{y}_l : l \in \mathcal{B}_t\}$, where $\mathcal{B}_t = \{\mathbf{y}_l : i \leq l < t + N_b\}$ with $|\mathcal{B}_t| = N_b$. When $t > L - N_b$, we simply take the remaining new elements to constitute this mini-batch.[2] We store the set of dictionary atoms participating in the SR of the signals in $\mathbf{Y}_{\mathcal{B}_t}$ as $\mathbf{D}_{\mathcal{B}_t}$. We define $\mathbf{Y}_{\mathcal{Q}_t} = \{\mathbf{y}_l : l \in \mathcal{Q}_t\}$ with $\mathcal{Q}_t = \{\mathbf{y}_l : 1 \leq l < t - 1\}$ as the set of previous training elements at iteration t. We consider a randomly selected mini-batch $\mathbf{Y}_{\mathcal{M}_t} = \{\mathbf{y}_l : l \in \mathcal{M}_t\}$ with $\mathcal{M}_t \subset \mathcal{Q}_t$ such that $|\mathcal{M}_t| = N_r$. Let $\mathbf{Y}_{\mathcal{A}_t} = \{\mathbf{y}_l : l \in \mathcal{A}_t\}$ where $\mathcal{A}_t \subset \mathcal{M}_t$ such that $\mathcal{A}_t = \{l : l \in \mathcal{M}_t, \langle \mathbf{x}_l^T \cdot \mathbf{x}_t \rangle \neq 0\}$ be a subset of previous training elements that are correlated with the mini-batch of new elements. In order to avoid multiple occurrences of the same element in consecutive mini-batches, DOMINODL ensures that $\mathcal{M}_t \cap \mathcal{M}_{t-1} = \emptyset$. Let $\mathbf{D}_{\mathcal{A}_t}$ be the set of dictionary atoms used for SR of $\mathbf{Y}_{\mathcal{A}_t}$. Our new training set is $\mathbf{Y}_{\mathcal{C}_t} = \mathbf{Y}_{\mathcal{A}_t} \cup \mathbf{Y}_{\mathcal{B}_t}$. Both mini-batches of new and previous elements are selected such that the entire training set size $(N_b + N_r)$ is still smaller than that of CBWLSU where it is $N_{p_t} + 1$.

The dictionary update subproblem reduces to considering only the sets $\mathbf{Y}_{\mathcal{C}_t}$, $\mathbf{D}_{\mathcal{C}_t}$, and $\mathbf{X}_{\mathcal{C}_t}$

$$\hat{\mathbf{D}}_{\mathcal{C}_t} = \min_{\mathbf{D}_{\mathcal{C}_t} \in \mathcal{D}} ||\mathbf{Y}_{\mathcal{C}_t} - \mathbf{D}_{\mathcal{C}_t}\mathbf{X}_{\mathcal{C}_t}||_F^2. \tag{7}$$

Assume that the sparse coding for each example is known and define the errors as

$$\mathbf{E}_{\mathcal{C}_t} = \mathbf{Y}_{\mathcal{C}_t} - \mathbf{D}_{\mathcal{C}_t}\mathbf{X}_{\mathcal{C}_t} = [\mathbf{e}_1, \ldots, \mathbf{e}_{N_r}]. \tag{8}$$

We can update $\mathbf{D}_{\mathcal{C}_t}$, such that the above error is minimized, with the assumption of fixed $\mathbf{X}_{\mathcal{C}_t}$. A similar problem is considered in MOD, where error minimization is achieved through least squares. Here, we employ weighted least squares inspired by the fact that it has shown improvement in convergence over standard least squares [24]. We compute the weighting matrix $\mathbf{W}_{\mathcal{C}_t}$ using the SR error $\mathbf{E}_{\mathcal{C}_t}$

$$\mathbf{W}_{\mathcal{C}_t} = \text{diag}\left(\frac{1}{||\mathbf{e}_1||_2^2}, \ldots, \frac{1}{||\mathbf{e}_{N_r}||_2^2}\right). \tag{9}$$

We then solve the optimization problem

$$\hat{\mathbf{D}}_{\mathcal{C}_t} = \min_{\mathbf{D}_{\mathcal{C}_t} \in \mathcal{D}} ||(\mathbf{Y}_{\mathcal{C}_t} - \mathbf{D}_{\mathcal{C}_t}\mathbf{X}_{\mathcal{C}_t})\mathbf{W}_{\mathcal{C}_t}^{\frac{1}{2}}||_F^2. \tag{10}$$

This leads to the weighted least squares solution

$$\hat{\mathbf{D}}_{\mathcal{C}_t} = \mathbf{Y}_{\mathcal{C}_t}\mathbf{W}_{\mathcal{C}_t}\mathbf{Y}_{\mathcal{C}_t}^T\left(\mathbf{Y}_{\mathcal{C}_t}\mathbf{W}_{\mathcal{C}_t}\mathbf{Y}^{\mathbf{T}}{}_{\mathcal{C}_t}\right)^{-1}. \tag{11}$$

The dictionary $\mathbf{D}$ is then updated with the atoms $\hat{\mathbf{D}}_{\mathcal{C}_t}$ and its columns are normalized by their ℓ_2-norms.

$\mathbf{D}$ is next used for updating the sparse coding of $\mathbf{Y}_{\mathcal{C}_t}$ using batch OMP. Selecting a value for the maximal residual error δ in (4) is usually not straightforward. This value can be related to the amount of noise in the observed data but this information is not known. The samples of our training set can be seen as the realizations of a statistical process with an unknown distribution; therefore, one can associate with these realizations as the concept of *statistical entropy*. We compute the normalized entropy of the mean vector of all the training set samples as

$$E(\boldsymbol{\mu}_{\mathbf{Y}}) = -\sum_{i=1}^{M} P(\boldsymbol{\mu}_{\mathbf{y}_t}) \log P(\boldsymbol{\mu}_{\mathbf{y}_t}) \tag{12}$$

where $\boldsymbol{\mu}_{\mathbf{Y}}$ is the mean vector of all training samples, M is the number of features for each training sample, and $P(\cdot)$ is the probability mass function. In our case, $P(\cdot)$ is obtained as the normalized histogram of $\boldsymbol{\mu}_{\mathbf{Y}}$. Here, $E(\boldsymbol{\mu}_{\mathbf{Y}})$ is an indicator of the *randomness* of the data due to noise. We use $E(\boldsymbol{\mu}_{\mathbf{Y}})$ as the maximal residual error δ while applying batch OMP in DOMINODL. Algorithm 1 summarizes all major steps of DOMINODL.

[2] In numerical experiments, we observed that the condition $t > L - N_b$ rarely occurs because DOMINODL updates the dictionary and converges in very few iterations. The algorithm also ensures that the number of previous samples $\geq 2N_r$ before the dictionary update. If this condition is not fulfilled, then it considers all previous training samples.

TABLE I
COMPARISON OF DL STEPS

DL step	K-SVD	LRSDL	ODL	CBWLSU	DOMINODL
Training method	Batch	Batch	Online	Online	Online
Sparse coding method	OMP	FISTA	LARS	Batch OMP	Entropy-thresholded batch OMP
Dictionary update	Entire $\mathbf{D}$ atom-wise	Entire $\mathbf{D}$	Entire $\mathbf{D}$ group-wise	Entire $\mathbf{D}$ atom-wise	Partial $\mathbf{D}$ adaptively group-wise
Training samples per iteration	Entire $\mathbf{Y}$	Entire $\mathbf{Y}$	$\mathbf{Y}_t$	$\mathbf{Y}_{\mathcal{N}_t}$	$\mathbf{Y}_{\mathcal{C}_t}$
Optimization method	SVD	ADMM	Gradient descent	Weighted least squares	Weighted least squares
Post-update dictionary pruning	Yes	No	Yes	No	No
Training-set drop-off	No	No	No	No	Yes

Algorithm 1 DOMINODL

Input: Training set ($\mathbf{Y}$), number of trained atoms (K), mini-batch dimension for new training data (N_b), mini-batch dimension for previous training data (N_r), drop-off value (N_u), convergence threshold ($\chi \in \mathbb{R}$)

Output: Learned dictionary ($\mathbf{D}$), sparse decomposition of the training set ($\mathbf{X}$)

Generate initial dictionary $\mathbf{D}$ of dimension K using training samples
Normalize the columns of $\mathbf{Y}$ and $\mathbf{D}$ by their ℓ_2-norms
Sparsely decompose $\mathbf{Y}$ with the initial dictionary using entropy-thresholded batch OMP
Loop
  Gather a mini-batch of new training set elements $\mathbf{Y}_{\mathcal{B}_t} = \{\mathbf{y}_l : l \in \mathcal{B}_t\}$ with $\mathcal{B}_t = \{\mathbf{y}_l : i \leq l < i + N_b\}$ and $|\mathcal{B}_t| = N_b$
  SR of $\mathbf{Y}_{\mathcal{B}_t}$ with the dictionary $\mathbf{D}$ using entropy-thresholded batch OMP
  Store the set of atoms $\mathbf{D}_{\mathcal{B}_t}$ participating in the SR of $\mathbf{Y}_{\mathcal{B}_t}$
  Randomly select a mini-batch of previous training set elements $\mathbf{Y}_{\mathcal{M}_t} = \{\mathbf{y}_l : l \in \mathcal{M}_t\}$ with $|\mathcal{M}_t| = N_r$ and $\mathcal{M}_t \cap \mathcal{M}_{t-1} = \emptyset$
  Consider a subset $\mathbf{Y}_{\mathcal{A}_t} = \{\mathbf{y}_l : l \in \mathcal{A}_t\}$ where $\mathcal{A}_t \subset \mathcal{M}_t$ such that $\mathcal{A}_t = \{l : l \in \mathcal{M}_t, \langle \mathbf{x}_l^T \cdot \mathbf{x}_t \rangle \neq 0\}$
  Form $\mathbf{Y}_{\mathcal{C}_t} = \mathbf{Y}_{\mathcal{A}_t} \cup \mathbf{Y}_{\mathcal{B}_t}$ and store $\mathbf{D}_{\mathcal{C}_t}$ the atoms of $\mathbf{D}$ shared by $\mathcal{B}_t$ and $\mathcal{M}_t$
  Compute the errors $\mathbf{E}_{\mathcal{C}_t} = \mathbf{Y}_{\mathcal{C}_t} - \mathbf{D}_{\mathcal{C}_t}\mathbf{X}_{\mathcal{C}_t} = [\mathbf{e}_1, \ldots, \mathbf{e}_{N_r}]$
  Form the weighting matrix $\mathbf{W}_{\mathcal{A}_t} = \text{diag}\left(\frac{1}{||\mathbf{e}_1||_2^2}, \ldots, \frac{1}{||\mathbf{e}_{N_r}||_2^2}\right)$
  Update $\hat{\mathbf{D}}_{\mathcal{C}_t}$: $\hat{\mathbf{D}}_{\mathcal{C}_t} = \mathbf{Y}_{\mathcal{C}_t}\mathbf{W}_{\mathcal{C}_t}\mathbf{Y}_{\mathcal{C}_t}^T(\mathbf{Y}_{\mathcal{C}_t}\mathbf{W}_{\mathcal{C}_t}\mathbf{Y}_{\mathcal{C}_t}^T)^{-1}$ and normalize its columns
  Replace the updated atoms $\mathbf{D}_{\mathcal{C}_t}$ into $\mathbf{D}$ and normalize its columns
  Perform SR of selected signals used in the previous step using entropy-thresholded batch OMP
  Eliminate previous training set elements which have not been used for the last N_u iterations
  if $||(\mathbf{Y}_{\mathcal{C}_t} - \mathbf{D}_t\mathbf{X}_{\mathcal{C}_t})(\mathbf{W}_i)^{0.5}||_F^2 < \chi$ **then** break
EndLoop

Table I summarizes the important differences between DOMINODL and other related algorithms. Like MOD and CBWLSU, DOMINODL uses a weighted least squares solution in the dictionary update. The proof of convergence for the alternating minimization method in MOD was provided in [34], where it is shown that alternating minimization converges linearly as long as the following assumptions hold true: sparse coefficients have bounded values, sparsity level is on the order of $\mathcal{O}(M^{1/6})$, and the dictionary satisfies the RIP property. In [24], these assumptions have been applied for CBWLSU convergence. Compared with CBWLSU, the improvements in DOMINODL include mini-batch-based data selection and data reduction via a drop-off strategy but the update algorithms remain the same. Numerical experiments in Section VI suggest that DOMINODL usually converges in far fewer iterations than CBWLSU.

Although we developed and tested DOMINODL on a highly correlated GPR data set (see Section IV), this technique may be employed in other applications where real-time learning is necessary and the signals are correlated. Our tests demonstrate that DOMINODL converges faster than other online DL approaches (see Section VI) because of the combined strategy of drawing more new elements for each iteration, considering less previous elements in search for correlation and dropping off the unused previous elements. The entropy-based calculation of δ, although not exclusive for DL applications as mentioned above, also helps in improving the SR of the data, thus learning a more representative dictionary.

Computational complexity of DOMINODL is very low compared with other online approaches. As mentioned earlier, there are K atoms in the dictionary. Assume that every signal is represented by a linear combination of K_s atoms, $K_s \ll K$. Empirically, among all possible combinations of K_s atoms from K, the probability to have a common atom in the SR is K_s/K. Given L training elements, the number of training data which have a specific atom in their representation is proportional to LK_s/K. Suppose our mini-batch has elements that reduce the number of training data by a factor $\beta < 1$ (depending on the values of N_r and N_b). Furthermore, assume that the dropping off step reduces the training set elements by a factor $\rho < 1$. The number of training data L_i in the tth iteration is proportional to $\beta\rho i K_s/K$. Then, the worst estimate of DOMINODL's computational complexity is due to the sparse coding batch OMP which is of order $\mathcal{O}(L_t K^2) = \mathcal{O}(\beta\rho t K_s K) \approx \mathcal{O}(\beta\rho t K)$. This is much smaller than the complexity of ODL ($= \mathcal{O}(K^3)$) or CBWLSU ($= \mathcal{O}(tK)$).
Fig. 1 shows the computational complexity of online DL approaches. Fig. 1(a) shows that, for a fixed number of iterations ($t = 60$), the general trend of complexity with respect to the increase in the number of atoms (K) is similar for all algorithms. However, the complexity of ODL is higher

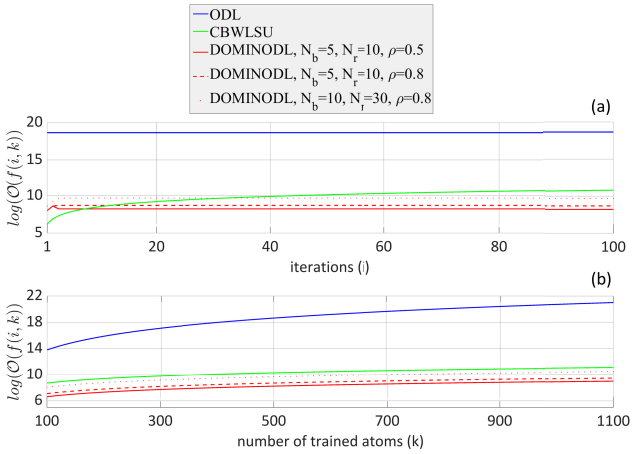

Fig. 1. Computational complexity of online DL strategies for an increasing number of (a) iterations and (b) trained atoms.

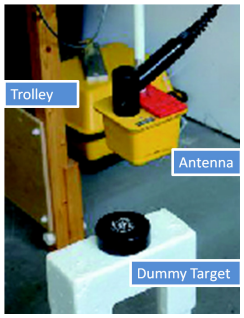

Fig. 2. L-band GPR system is attached to a movable trolley platform. It is mounted along a rail system and scans the target from above.

than CBWLSU and DOMINODL; the latter being the least complex. When the number of iterations is increased, the complexity of ODL and CBWLSU has a similar increasing trend [see Fig. 1(b)]. In the case of DOMINODL, its complexity is similar to the increasing trend of CBWLSU and determined largely by N_b. When DOMINODL iterations begin accounting for N_r previous elements, its complexity stays constant. The value of β changes for every iteration, while ρ depends on the data itself. In general, after a few dozens of iterations, DOMINODL's complexity always stays lower than CBWLSU.

IV. Measurement Campaign

In this section, we first provide details of our GPR system and the field measurement campaign. We then describe the procedure to organize the entire data set for our application.

A. Ground Penetrating Radar System

Our GPR (see Fig. 2) is the commercially available SPRScan system manufactured by Electrical Research Association (ERA) Technology. It is an L-band, impulse waveform, UWB radar that is mounted on a movable trolley platform. Pulsed GPRs are more effective in terms of offering penetration depth and wide bandwidth with respect to the standard stepped-frequency continuous wave (SFCW) systems. The former is also more robust to electronic interference and does not suffer from unequal balancing of antenna signals [56].

TABLE II
Technical Characteristics of Impulse GPR

Parameter	Value
Operating frequency	2 GHz
Pulse repetition frequency	1 MHz
Pulse length	0.5 ns
Sampling time	25 ps
Spatial sampling along the beam	1 cm
Cross-beam resolution	4 cm
Antenna height	5-9 cm
Antenna configuration	Perpendicular broadside
Samples/A-scan	512

Table II lists the salient technical parameters of the system. The radar uses an 8×8 cm dual bow-tie dipole antenna for both transmit (Tx) and receive (Rx) sealed in a metallic shielding filled with an internal absorber. The central frequency of the system (f_c) and its bandwidth (Δf) is 2 GHz. The pulse repetition frequency (PRF) and the sampling of the receiver ADC are 1 MHz. The scanning system has a resolution of 1 cm toward the perpendicular broadside (or X-direction) and 4 cm toward the cross-beam (Y-direction). In our field campaigns, the SPRScan system moves along the survey area over a rail system, which allows accurate positioning of the sensor head in order to obtain the aforementioned resolution in X and Y (see also Section IV).

The transmit pulse of the GPR system is a *monocycle*. Given the Gaussian waveform

$$s_G(t) = Ae^{-2\pi^2 f_c^2 (t-\tau)^2}, \quad t \in [0, +\infty] \tag{13}$$

where f_c is the central frequency, A is the peak amplitude, and $\tau = 1/f_c$, and the monocycle waveform is its first derivative [57]

$$s_T(t) = -4\pi^2 f_c^2 A(t-\tau)e^{-2\pi^2 f_c^2 (t-\tau)^2}, \quad t \in [0, +\infty]. \tag{14}$$

In these UWB systems, both the central frequency and the bandwidth are approximately the reciprocal of the pulselength.

The scattering of UWB radar signals from complex targets that are composed of a finite number of scattering centers can be described in terms of the channel impulse response (CIR). Here, the CIR is considered as a linear, time invariant, causal system, which is a function of the target shape, size, constituent materials, and scan angle. The CIR $h(t)$ of a GPR target, with M scatterers, is expressed as a series of time-delayed and weighted Gaussian pulses [58]

$$h(t) = \sum_{m=1}^{M} \alpha_m e^{-4\pi[(t-t_m)/\Delta T_m]^2} \tag{15}$$

where each scatterer located at range r_m from the radar is characterized by the reflectivity α_m, duration ΔT_m, and relative time shift $t_m = 2r_m/v_s$, where $v_s = c/\sqrt{\epsilon_r}$ is the speed of the electromagnetic wave in the soil, $c = 3\times10^8$ m/s is the speed of light, and ϵ_r is the dielectric constant which depends on the soil composition and moisture.

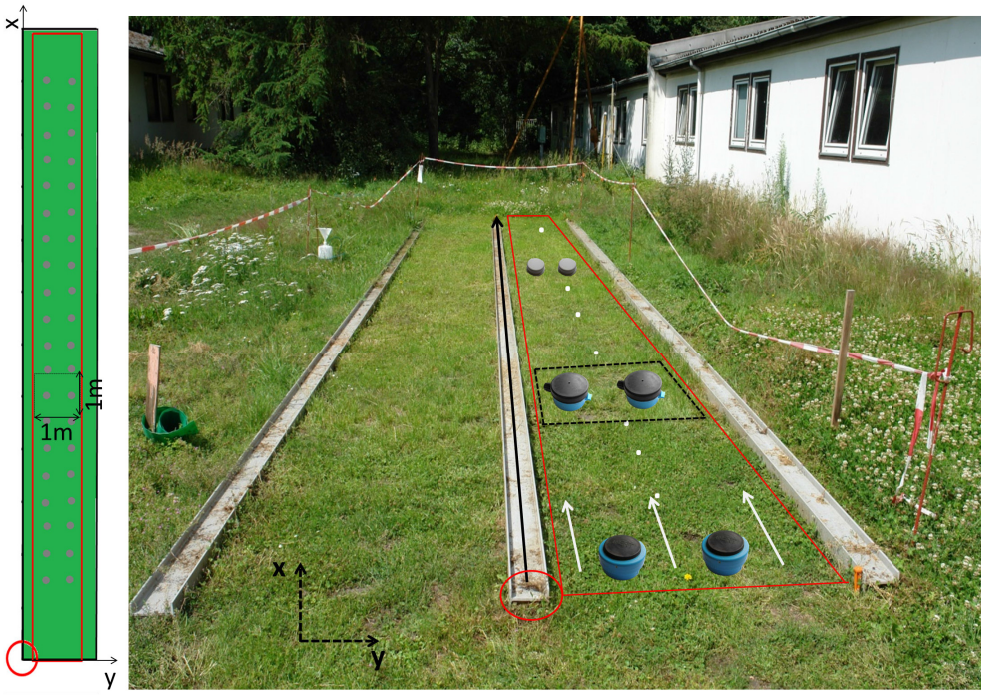

Fig. 3. (Right) LIAG test field in Hannover along with its (Left) layout. The scan directions X and Y of the radar are indicated on the photograph and layout. The radar coverage region is indicated by solid red lines with a red circle showing the origin of the scan. The white arrows in the photograph indicate that specific lanes scanned in the X-direction are separated in the Y-direction by 4 cm. In the layout, each gray dot represents the location of a buried test target. An individual survey area unit of 1 m $\times$ 1 m that contains two targets is also indicated on the layout (solid black lines) and the photograph (dotted black lines). The solid black arrow over the middle rail in the photograph is where the SPRScan was mounted.

The response of the target to the Gaussian monocycle is the received signal

$$y(t) = s_T(t) * h(t) \tag{16}$$

also regarded as the target image or *range profile*. For each X/Y position, the system receives a radar echo (range profile) from the transmitted pulse. In order to deal with the exponential signal attenuation during the propagation through the soil medium, the dynamic range of the signal is enhanced via stroboscopic sampling [2], [59], [60]. This technique comprises integrating N receiver samples (generated by transmitting a sequence of N pulses) at the ADC receiver sampling rate but with a small time offset δ for each of them. To achieve the desired stroboscopic sampling rate T_s, the time offset must be selected accordingly, i.e., $\delta = T_s/N$ [60]. Our GPR system employs stroboscopic sampling to reach a pseudo sampling frequency of $f_s = 1/T_s = 40$ GHz (much above the Nyquist rate) to yield the discrete-time signal $y[n] = y(nT_s)$.

The receiver has the ability to acquire a maximum of 195 profiles per second, each one consisting of 512 range samples. Prior to the A/D conversion, the signal is averaged to improve the signal-to-noise ratio (SNR). A time-varying gain correction can be applied to compensate for the soil attenuation and increase the overall dynamic range of the system. The receiver averages 100 range profiles for each antenna position.

B. Test Field Measurements

We evaluated the proposed approach with the measurement data from a 2013 field campaign at the Leibniz Institute for Applied Geophysics (LIAG) in Hannover, Germany [6]; Fig. 3 shows the test field for detailed ground truth information. The soil texture was sandy and highly inhomogeneous (due to the presence of material such as organic matter and stones), thereby leading to a high variability in the electrical parameters. We measured the dielectric constant at three different locations of the testbed with a time-domain reflectometer (TDR) to obtain an estimate of its mean value and variability. The average value oscillated between 4.6 and 10.1 with 15% standard deviation and *correlation length* [6] of 20 cm. These large variations in soil dielectric characteristics pose difficulties in mine detection.

During the field tests, the SPRScan system moved on two plastic rails with the scan resolution in the X- and Y-directions being 1 and 4 cm, respectively. The entire survey lane was divided into 1 $\times$ 1 m sections (see Fig. 3), each containing two targets in the center. The targets on the left-hand and right-hand sides of the lane were buried at approximately 10- and 15-cm depths, respectively.

Our testbed contains standard test targets (STT) and simulant landmines (SIM) of different sizes and shapes. An STT is a surrogate target used for testing landmine detection equipment. It is intended to interact with the equipment in an identical manner as a real landmine does. An SIM has the representative characteristics of a specific landmine class although it is not a replica of any specific model. In this paper, we study three STTs (PMA2, PMN, and Type-72) and one SIM (ERA). All of these test objects are buried at a depth of 10–15 cm in the test field [61]. For classification purposes,

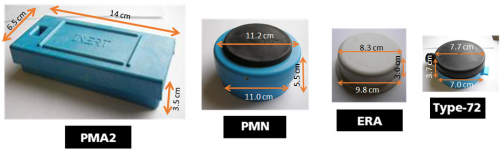

Fig. 4. Details of the SIM and the STT buried in the test field.

we group PMN and PMA2 together as the largest targets while T72 mines are the smallest (Fig. 4).

C. Data Set Organization

The entire LIAG data set consists of 27 aforementioned survey sections (or simply, "surveys") of size 1×1 m. Every survey consists of 2500 range profiles. We arranged the data into the training set ($\mathbf{Y}$) to be used for both DL and classification (as explained in Section II-A) and a test set ($\mathbf{Y}_{\text{TEST}}$) to evaluate the performance of the proposed algorithms.

The training set $\mathbf{Y} \in \mathbb{R}^{M\times L}$ is a matrix whose L columns $\{\mathbf{y}_i\}_{i=1}^{L}$ consist of sampled range profiles $\mathbf{y}_i = \left[y[0], \cdots, y[M-1]\right]^T$ of M range profiles each. The profiles are selected from different surveys and contain almost exclusively either a particular class of landmine or clutter. In total, we have 463, 168, 167, and 128 range profiles for clutter, PMA2/PMN, ERA, and Type-72, respectively. An accurate separation of these classes was very challenging because of the contributions from the non-homogeneous soil clutter that often masked the target responses completely. A poor selection would lead the DL to learn a dictionary that is appropriate for sparsely representing clutter, instead of landmines. The test set $\mathbf{Y}_{\text{TEST}} \in \mathbb{R}^{M\times J}$ is a matrix with $J = 15000$ columns $\{\mathbf{y}_{\text{TEST}_i}\}_{i=1}^{J}$ that correspond to sampled range profiles from six surveys, two for each target class. The test and training sets contain data from separate surveys to enable fair assessment of the classification performance.

We denote by the matrices $\mathbf{X} \in \mathbb{R}^{K\times L}$ and $\mathbf{X}_{\text{TEST}} \in \mathbb{R}^{K\times J}$ the SRs of $\mathbf{Y}$ and $\mathbf{Y}_{\text{TEST}}$, respectively and K by the number of atoms of the learned dictionary $\mathbf{D} \in \mathbb{R}^{M\times K}$.

V. Parametric Analysis

In practice, the SR-based classification performance is sensitive to the input parameters of DL algorithms thereby making it difficult to directly apply DL with arbitrary parameter values. Previous works set these parameters through hit-and-trial or resorting to metrics that are unable to discriminate the influence of different parameters [20]. In this section, we propose methods to investigate the effect of the various input parameters on the learning performance and then preset the parameter to *optimal* values that yield the dictionary $\mathbf{D}$ (for each DL method) optimized to sparsely represent our GPR data, therefore improving the quality of the features for classification (i.e., the sparse coefficients).

Table III lists these parameters (see Section III): the number of iterations N_t, the number of trained atoms K, and DOMINODL parameters N_b, N_r, and N_u. We applied K-SVD, LRSDL, ODL, CBWLSU, and DOMINODL separately on the training set for different combinations of parameter values.

TABLE III
DL Parameters

DL algorithm	Input parameters
K-SVD	N_t, K
LRSDL	N_t, K
ODL	N_t, K
CBWLSU	K
DOMINODL	K, N_b, N_r, N_u

In order to compare the dictionaries obtained from various DL algorithms, we use a *similarity measure* that quantifies the closeness of the original training set $\mathbf{Y}$ with the reconstructed set $\hat{\mathbf{Y}}$ obtained using the sparse coefficients of the learned dictionary $\mathbf{D}$. From these similarity values, empirical probability density functions (EPDFs) for any combination of parameter values are obtained; we evaluate these EPDFs using statistical metrics described in Section V-B. These metrics efficiently characterize the similarity between $\mathbf{Y}$ and $\hat{\mathbf{Y}}$ and lead us to an optimal selection of various DL input parameters for our experimental GPR data set.

A. Similarity Measure

Consider the cross correlation between the original training set vector $\mathbf{y}_i$ and its reconstruction $\hat{\mathbf{y}}_i$: $\mathbf{r}_{\mathbf{y}_i,\hat{\mathbf{y}}_i}[l] = \sum_{n=-\infty}^{+\infty} \mathbf{y}_i[n]\hat{\mathbf{y}}_i[n+l]$. The normalized cross correlation is defined as

$$\overline{\mathbf{r}_{\mathbf{y}_i,\hat{\mathbf{y}}_i}}[l] = \frac{\mathbf{r}_{\mathbf{y}_i,\hat{\mathbf{y}}_i}[l]}{\sqrt{\mathbf{r}_{\mathbf{y}_i,\mathbf{y}_i}[0]\mathbf{r}_{\hat{\mathbf{y}}_i,\hat{\mathbf{y}}_i}[0]}}. \tag{17}$$

For the vector $\mathbf{y}_i$, we define the similarity measure s_i as

$$s_i = \max|\overline{\mathbf{r}_{\mathbf{y}_i,\hat{\mathbf{y}}_i}(l)}| \tag{18}$$

where a value of s_i closer to unity demonstrates greater similarity of the reconstructed data with the original training set. We compute $\{s_i\}_{i=1}^{L}$ for all vectors $\{\mathbf{y}_i\}_{i=1}^{L}$ and then obtain the normalized histogram or EPDF $p_{s_{\text{DL}}}$ of all similarity measures. Here, the subscript DL represents the algorithm used for learning $\mathbf{D}$, e.g., "K","O","C," and "D" for K-SVD, ODL, CBWLSU, and DOMINODL, respectively. Various parameter combinations for a specific DL method result in a collection of EPDFs. For a given DL method, our goal is to compare the EPDFs by varying these parameters and arrive at the thresholds of parameter values after which the changes in $p_{s_{\text{DL}}}$ are only incremental. For instance, Fig. 5 shows the EPDFs of $\{s_i\}_{i=1}^{L}$ obtained from the GPR mines data, where optimal parameters for different DL methods were determined using the statistical methods described in Section V-B. We note that the online DL approaches (p_{s_O}, p_{s_C}, and p_{s_D}) yield distributions that are more skewed toward unity than K-SVD (p_{s_K}).

B. Statistical Metrics

We are looking for parameter values for which $p_{s_{\text{DL}}}$ is skewed toward unity and has small variance. The individual comparisons of mean (μ) and standard deviation (σ), as used in previous GPR DL studies [20], are not sufficient to quantify

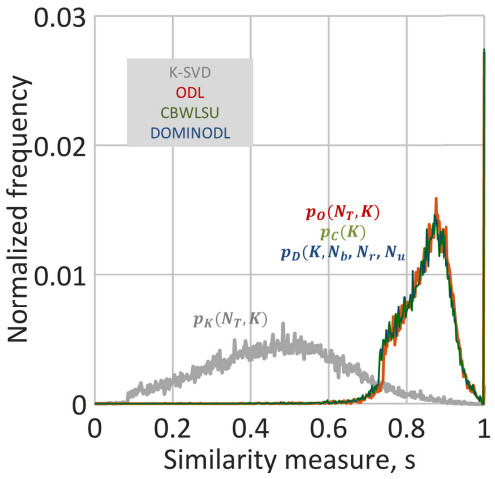

Fig. 5. Normalized histograms of similarity measure using the following optimal parameters for the DL algorithm: $N_t = 100$, $K = 640$, $N_b = 30$, $N_r = 10$, and $N_u = 10$. See Section V-C on the process to select these optimal values.

the observed dispersion in the EPDFs obtained by varying any of the parameter values. Some DL studies [20], [50], [62] rely on bulk statistics such as NRMSE but these quantities are insensitive to large changes in parameter values and, therefore, unhelpful in fine-tuning the algorithms. For this evaluation, we will use three different metrics: the coefficient of variation, the two-sample K-S distance, and the DKW inequality.

1) Coefficient of Variation: We choose to simultaneously compare both (μ) and variance (σ) of a single EPDF by using the *coefficient of variation*, CV $= \sigma/\mu$; in our analysis, it represents the extent of variability in relation to the mean of the similarity values.

2) Two-Sample Kolmogorov–Smirnov Distance: In the context of our application, it is more convenient to work with the cumulative distribution functions (CDFs) rather than with PDFs because the well-developed statistical inference theory allows for convenient comparison of CDFs. Therefore, our second metric to compare similarity measurements obtained by successive changes in parameter values is the *two-sample K-S distance* [21], which is the maximum distance between two given empirical CDFs (ECDFs). Larger values of this metric indicate that samples are drawn from different underlying distributions. Given two random variables s_1 and s_2, suppose $\hat{F}_{s_1}$ and $\hat{G}_{s_2}$ are their ECDFs of the same length and correspond to their EPDFs $\hat{f}_{s_1}$ and $\hat{g}_{s_2}$, respectively. Then, the K-S distance is

$$d_{ks}(\hat{F}_{s_1}, \hat{G}_{s_2}) = \sup_{1 \le i \le L} |\hat{F}_{s_1}(i) - \hat{G}_{s_2}(i)| \tag{19}$$

where sup denotes the supremum over all distances and L is the number of i.i.d. observations (or samples) to evaluate both distributions. In our case, L is the number of range profiles in the training set. We first compute a reference ECDF ($\hat{G}_{s_{\text{ref}}}$) for each DL algorithm and fixed parameter values. For our purposes, this reference ECDF will be obtained by a particular combination of input parameters of the selected DL algorithm. Then, we vary parameter values from this reference and obtain the corresponding ECDF $\hat{F}_{s_{\text{test}}}$ of similarity measure. Finally, we calculate the K-S distance d_{ks} of $\hat{F}_{s_{\text{test}}}$ with respect to $\hat{G}_{s_{\text{ref}}}$ as

$$d_{ks} = d_{ks}(\hat{F}_{s_{\text{test}}}, \hat{G}_{s_{\text{ref}}}) = \sup_{1 \le i \le L} |\hat{F}_{s_{\text{test}}}(i) - \hat{G}_{s_{\text{ref}}}(i)|. \tag{20}$$

For our evaluation, d_{ks} states how much the selection of certain input parameters of DL changes the ECDFs of similarity values (i.e., how different is the result of DL) with respect to the reference.

3) Dvoretzky–Kiefer–Wolfowitz Inequality Metric: As a third metric, we exploit the DKW inequality [22], [23], which precisely characterizes the rate of convergence of an ECDF to a corresponding exact CDF (from which the empirical samples are drawn) for any finite number of samples. Let $d_{ks}(\hat{G}_s, F_s)$ be the K-S distance between ECDF $\hat{G}_s$ and the continuous CDF F_s for a random variable s and L samples. Since $\hat{G}_s$ changes with the change in the L random samples, $d_{ks}(\hat{G}_s, F_s)$ is also a random variable. We are interested in the conditions that provide desired confidence in verifying if F and G are the same distributions for a given finite L. If the two distributions are indeed identical, then the DKW inequality bounds the probability that d_{ks} is greater than any number ϵ, with $0 < \epsilon < 1$ as follows[3]:

$$\Pr\{d_{ks}(\hat{G}_s, F) > \epsilon\} \le 2e^{-2L\epsilon^2}. \tag{21}$$

Consider a binary hypothesis testing framework where we use (21) to test the null hypothesis $\mathcal{H}_0 : F = \hat{G}$ for a given ϵ. The probability of rejecting the null hypothesis when it is true is called the p-value of the test and is bounded by the DKW inequality. Assuming the p-value is smaller than a certain confidence level α, the following inequality must hold with probability at least $1 - \alpha$:

$$d_{ks}(\hat{G}_s, F) \le \sqrt{-\frac{1}{2L}\ln\left(\frac{\alpha}{2}\right)}. \tag{22}$$

Our goal is to use the DKW inequality to compare two ECDFs $\hat{F}_{s_{\text{test}}}$ and $\hat{G}_{s_{\text{ref}}}$ as in (20) to verify if they are drawn from the same underlying CDF. By the triangle inequality, the K-S distance satisfies

$$d_{ks}(\hat{F}_{s_{\text{test}}}, \hat{G}_{s_{\text{ref}}}) = d_{ks}(\hat{F}_{s_{\text{test}}}, F) + d_{ks}(\hat{G}_{s_{\text{ref}}}, F) \tag{23}$$

where G and F are the underlying CDFs corresponding to $\hat{G}$ and $\hat{F}$. We now bound the right-hand side using DKW

$$\begin{aligned} d_{ks}(\hat{F}_{s_{\text{test}}}, \hat{G}_{s_{\text{ref}}}) &\le \sqrt{-\frac{1}{2L}\ln\left(\frac{\alpha}{2}\right)} + \sqrt{-\frac{1}{2L}\ln\left(\frac{\alpha}{2}\right)} \\ &= \sqrt{-\frac{2}{L}\ln\left(\frac{\alpha}{2}\right)} \end{aligned} \tag{24}$$

which is the maximum distance for which $\hat{F}_{s_{\text{test}}}$ and $\hat{G}_{s_{\text{ref}}}$ are identical with probability $1 - \alpha$. The DKW metric is the difference

$$d_{dkw} = \sqrt{-\frac{2}{L}\ln\left(\frac{\alpha}{2}\right)} - d_{ks}(\hat{F}_{s_{\text{test}}}, \hat{G}_{s_{\text{ref}}}). \tag{25}$$

[3]The corresponding asymptotic result that as $L \to \infty$, $d_{ks} \to 0$ with probability 1 is due to the Glivenko–Cantelli theorem [63], [64].

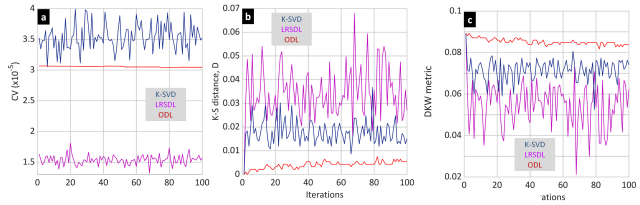

Fig. 6. (a) CV, (b) K-S distance, and (c) DKW metric for K-SVD, LRSDL, and ODL parameter analyses as a function of the number of iterations N_t.

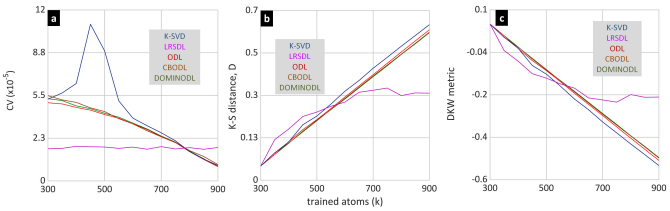

Fig. 7. (a) CV, (b) K-S distance, and (c) DKW metric for various DL algorithms as a function of the number of trained atoms K.

Larger values of this metric imply greater similarity between the two ECDFs; a negative value implies that the null hypothesis is not true.

C. Parametric Evaluation

We evaluated the performance of the aforementioned DL algorithms by analyzing the influence of the various DL input parameters using the metrics introduced in Section V-B for the reconstruction of the training set $\mathbf{Y}$. There are various soil types and scenarios for a landmine contaminated site. The LIAG test data provide an accurate representation of a practical scenario. Our metrics are general and derived from widely accepted statistical studies. Thus, their relevance to similar scenarios is very likely. As shown in Table III, the number of iterations N_t is not relevant to CBWLSU and DOMINODL, while the latter requires additional parameters to specify the mini-batch dimensions and the iterations required to drop-off unused training set elements. We compute the K-S distance and the DKW metric for all methods with respect to a reference distribution p_{ref}. This reference, different for each DL algorithm, is obtained using the following parameters as applicable: $N_t = 1$, $K = 300$, $N_b = 30$, $N_r = 10$, and $N_u = 10$.

1) Number of Iterations: Fig. 6(a)–(c) shows the effect of N_t on the CV, K-S test distance d_{ks}, and the DKW metric d_{W_L} for K-SVD, ODL, and LRSDL. We have skipped CBWLSU and DOMINODL from this analysis because they do not accept N_t as an input. For ODL, the CV remains relatively unchanged with an increase in N_t. However, the K-SVD CV exhibits an oscillating behavior and generally high values. In case of the K-S distance, ODL shows slight increase in d_{ks} while K-SVD oscillates around a mean value that is higher than ODL. The DKW metric provides better insight: even though the ODL distributions differ from p_{ref} with an increase in the iterations, the null hypothesis always holds because d_{dkw} remains positive. The d_{dkw} for K-SVD is also positive but much smaller than ODL. It also does not exhibit any specific trend with an increase in iterations. We also observed a similar behavior with the mean of similarity values. The influence of the number of iterations in LRSDL had the same oscillating behavior as in K-SVD but with larger variation. We conclude that the number of iterations N_t does not significantly influence the metrics for these algorithms and choose $N_t = 100$.

2) Number of Trained Atoms: Fig. 7(a)–(c) compares all three metrics with change in the number of trained atoms K, a parameter that is common to all DL methods. We observe that CV generally decreases with an increase in K. This indicates an improvement in the similarity between the reconstructed and the original training set. K-SVD shows an anomalous pattern for lower values of K but later converges to a trend that is identical to other DL approaches. The K-S distance exhibits a linear change in the distributions with respect to the reference. Since d_{ks} quantifies the difference between the distributions rather than stating which one is better, combining its behavior with CV makes it evident that an increase in K leads to better distributions of similarity values. The DKW metric d_{dkw}, calculated with the same reference, also expectedly shows a linear change. It is clear that even a slight change in K leads to more negative values of d_{dkw}, implying that the null hypothesis does not hold true. This shows the significant influence of the parameter K on the distributions. It was interesting to see a slight improvement in the coefficient of variation when using LRSDL with respect to the other strategies. However, KS-distance and DKW metric indicated that the distributions of similarity values for LRSDL were sensitive to the number of trained atoms only up to a certain value.

3) DOMINODL Parameters : It is difficult to evaluate DOMINODL EPDFs by varying all four parameters together. Instead, we fix the parameter that is common to all algorithms, i.e., the number of trained atoms K, and then determine optimal values of N_b, N_r, and N_u. Fig. 8 shows the coefficient of variation CV of the distribution of similarity values as a function of DOMINODL parameters. The drop-off value N_u appears to have a greater influence than mini-batch dimensions N_b and N_r. Our analysis of the computational times of DOMINODL showed that it is essentially independent of N_r and N_u but slightly increases with N_b. This is expected because we also increased the number of steps for sparse decomposition (see Algorithm 1), which is the source of bulk of computations in DL algorithms [24]. Furthermore, in order to ensure that the correlation and the drop-off steps kick off from the very first iteration, DOMINODL should admit several new samples for each iteration, thereby increasing N_b as well as the number of previous elements accordingly. Taking into account these observations, we choose $N_b = 30$, $N_r = 10$, and $N_u = 10$.

According to the results of the parametric evaluation, we choose the following combination of "optimal" parameters for testing our DL strategies: $N_t = 100$, $K = 640$, $N_b = 30$, $N_r = 10$, and $N_u = 10$.

VI. Experimental Results and Analysis

After selecting the input parameters of the proposed DL strategies, we proceed with the trained dictionaries for sparse

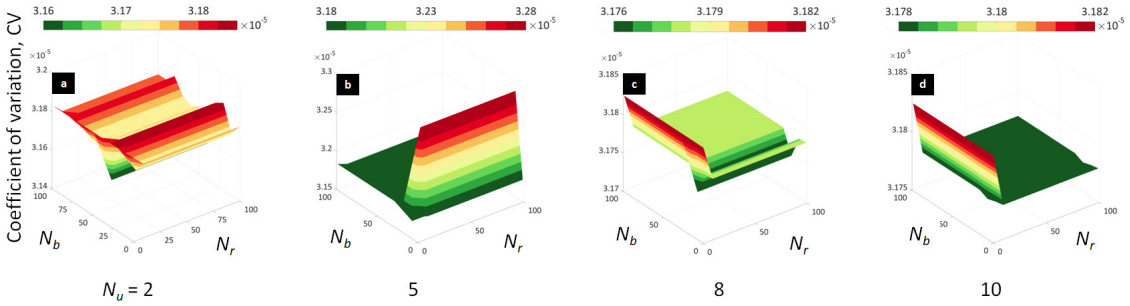

Fig. 8. CV as a function of DOMINODL input parameters for $k = 640$ and N_u as (a) 2, (b) 5, (c) 8, and (d) 10.

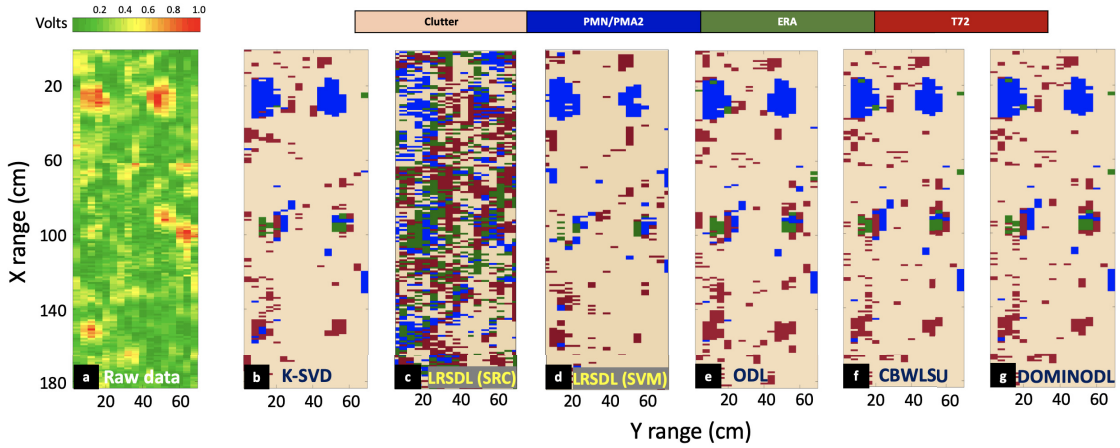

Fig. 9. (a) Raw data at 15-cm depth. The classification maps of the same area containing six buried landmines using an SR-based approach with dictionary learned using (b) K-SVD, (c) LRSDL (SRC), (d) LRSDL (SVM), (e) ODL, (f) CBWLSU, and (g) DOMINODL algorithms and optimally selected input parameters.

decomposition of both training and test sets. The resulting sets of sparse coefficients are the input to the SVM classifier. As mentioned in Section II-A, the threshold C and the kernel function parameter γ for SVM have been selected through cross validation. Our key objective is to demonstrate that online DL algorithms may lead to an improvement in the classification performance over batch learning strategies. In particular, we want to analyze the performance of DOMINODL in terms of classification accuracy and learning speed. As a comparison with a popular state-of-the-art classification method, we also show the classification results with a deep-learning approach based on CNN. Finally, we demonstrate classification performance when the original samples of the range profiles are randomly reduced.

A. Classification With Optimal Parameters

For a comprehensive analysis of the classification performance, we provide both classification maps and confusion matrices for the test set $\mathbf{Y}_{\text{TEST}}$ using the optimal DL input parameters that we selected following our parametric evaluation in Section V. The classification maps depict the predicted class of each range profile of the survey under test. The pixel dimension of these maps is dictated by the sampling of the GPR in X- and Y-directions (see Table II). We stitched together three of the six surveys from the test set $\mathbf{Y}_{\text{TEST}}$, where each survey had two buried landmines from a specific target class (PMN/PMA2, ERA, and Type-72).

Fig. 9 shows the classification maps for different DL methods along with the raw data at the depth of 15 cm. The selected survey area covers a total of 2880 range profiles. The raw data in Fig. 9(a) show that only four of the six mines exhibit a strong reflectivity while the other two mines have echoes so weak that they are not clearly visible in the raw data. Fig. 9(b)–(d) shows the results of the SR-based classification approaches using DL. All methods clearly detect and correctly classify the large PMN/PMA2 mines. In the case of the medium-size ERA, the echoes are certainly detected as non-clutter but some of its constituent pixels are incorrectly classified as another mine. It is remarkable that the left ERA mine is recognized by our method even though it cannot be discerned visually in the raw data. Most of the false alarms in the map belong to the smallest Type-72 mines. This is expected because their small sizes produce echoes very similar to the ground clutter. On the other hand, when T-72 is the ground truth, it is correctly identified.

Using accurate ground truth information, we defined *target halos* as the boundaries of the buried landmines. The dimension of the target halos varied depending on the mine size. Let the number of pixels and the declared mine pixels inside the target halo be n_t and n_m, respectively. Similarly, we denote the number of true and declared clutter pixels outside the target halo by n_c and n_d, respectively. Then, the probabilities of

TABLE IV
CONFUSION MATRIX WITH OPTIMAL DL INPUT PARAMETER SELECTION

		Clutter	PMN/PMA2	ERA	Type-72
K-SVD	Clutter	0.892	0.044	0.25	0.37
	PMN/PMA2	0.022	0.938[1]	0.166	0.074
	ERA	0.021	0.017	0.472	0.018
	Type-72	0.064	0	0.111	0.537
LRSDL (SRC)	Clutter	0.435	0.061	0.111	0.351
	PMN/PMA2	0.155	0.289	0.319	0.259
	ERA	0.172	0.372	0.361	0.278
	Type-72	0.237	0.272	0.208	0.111
LRSDL (SVM)	Clutter	0.889	0.114	0.333	0.463
	PMN/PMA2	0.026	0.877	0.186	0.185
	ERA	0.027	0	0.444	0.926
	Type-72	0.058	0.008	0.041	0.426
ODL	Clutter	0.871	0	0.194	0.333
	PMN/PMA2	0.022	0.973	0.139	0
	ERA	0.018	0.026	0.583	0.018
	Type-72	0.088	0	0.083	0.648
CBWLSU	Clutter	0.872	0.017	0.181	0.314
	PMN/PMA2	0.023	0.973	0.153	0
	ERA	0.025	0.008	0.528	0
	Type-72	0.08	0	0.138	0.685
DOMINODL	Clutter	0.876	0.017	0.167	0.315
	PMN/PMA2	0.023	0.974	0.138	0
	ERA	0.027	0.008	0.58	0
	Type-72	0.077	0	0.11	0.685

[1] Gray denotes the $\mathbf{P_{CC}}$ value for a specified class and DL algorithm

correct classification (P_{CC}) for each target class and clutter are, respectively

$$P_{\mathrm{CC}_{\mathrm{mines}}} = \frac{n_m}{n_t}, \text{ and } P_{\mathrm{CC}_{\mathrm{clutter}}} = \frac{n_d}{n_c}. \tag{26}$$

The P_{CC} being the output of a classifier should not be mistaken as the radar's probability of detection P_d, which is the result of a detector. A detector declares the presence of a mine when only a few pixels inside the halo have been declared as mine; P_{CC} provides a fairer and more accurate evaluation of the classification result. This per-pixel information can be easily used to improve the final detection result. For instance, the operator could set a threshold for the minimum number of pixels to be detected in a cluster so that a circle with center at the cluster centroid could be used as the detected mine. However, such a circle may exclude some of the mine pixels leading to a potential field danger. The per-pixel classification is then employed to determine the guard area around the mine circle.

A confusion matrix is a quantitative representation of the classifier performance. The matrix lists the probability of classifying the ground truth as a particular class. The classes listed column-wise in the confusion matrix are the ground truths, while the row-wise classes are their predicted labels. Therefore, the diagonal of the matrix is P_{CC} while off-diagonal elements are the probabilities of misclassification.

For the classification map of Fig. 9, Table IV shows the corresponding confusion matrices for each DL-based classification approach. In general, we observe an excellent classification of PMN/PMA2 landmines ($\sim$98%), implying that almost every range profile in the test set which belongs to this class is correctly labeled. The P_{cc} for the clutter is also quite high ($\sim$90%). This can also be concluded from the classification maps where the false alarms within the actual clutter regions are very sparse (i.e., they do not form a cluster) and, therefore, unlikely to be interpreted as an extended target. As noted previously, most of the clutter misclassification is associated with the Type-72 class. The ERA test targets show some difficulty with correct classification. But most of the pixels within its target halo are declared, at least, as some type of mine (which is quite useful in terms of issuing safety warnings in the specific field area). This result can be explained by the fact that ERA test targets do not represent a specific mine but have general characteristics common to most landmines. The Type-72 mines exhibit P_{cc}, which is slightly higher with respect to ERA targets. This is a remarkable result because Type-72 targets were expected to be the most challenging to classify due to their small size.

Conventionally, as mentioned in [47], LRSDL is used with an SR-based classification. However, applying this approach to our problem resulted in very low accuracy (an average of $\sim$20% across all classes as evident from Table IV) and semi-random classification maps (Fig. 9). This can be explained by the extreme similarity between the training set examples of different classes; mines and clutter are only slightly dissimilar in their responses and mine responses are generally hidden in the ground reflections. Each learned "block" D_c differed only slightly from the other; therefore, poor classification results are achieved with this data set. On the other hand, when we used the dictionary learned with LRSDL with our SVM-based technique, we obtained better classification accuracy (see Table IV and Fig. 9). However, this performance is still inferior to K-SVD and, hence, even worse than the other online DL approaches.

All DL algorithms used for our sparse classification approach show very similar results for the clutter and PMN/PMA2 classes. However, online DL methods show higher P_{CC} for the ERA and Type-72 targets than K-SVD. From Table IV, the detection enhancement using the best of the online DL algorithms for PMN/PMA2 over K-SVD is $((0.974 - 0.938) \times 100)/0.938 \approx 4\%$. The improvements for ERA and T-72 are computed similarly as 23% and 28%, respectively.

B. Classification With Non-Optimal Parameters

In order to demonstrate how the quality of the learned dictionary affects the final classification, we now show the confusion matrices for a non-optimal selection of input parameters in different DL algorithms. Our goal is to emphasize the importance of learning a good dictionary by selecting the optimal parameters rather than specifying how each parameter affects the final classification result. We arbitrarily selected the number of trained atoms K to be only 300 for all DL approaches and reduced the number of iterations to 25 for ODL and KSVD and, for DOMINODL, we use $N_r = 30$, $N_b = 5$, and $N_u = 2$. Table V shows the resulting confusion matrix. While the clutter classification accuracy is almost the same as in Table IV, the P_{cc} for PMN/PMA2 landmines decreased by $\sim$10% for most of the algorithms except ODL where it remains unchanged. The classification accuracy for ERA and Type-72 mines is only slightly worse for online

TABLE V
CONFUSION MATRIX WITH NON-OPTIMAL DL INPUT PARAMETER SELECTION

		Clutter	PMN/PMA2	ERA	Type-72
K-SVD	Clutter	0.853	0.07	0.305	0.222
	PMN/PMA2	0.037	0.851	0.222	0.111
	ERA	0.032	0	0.194	0.241
	Type-72	0.077	0.078	0.277	0.426
ODL	Clutter	0.86	0.017	0.181	0.444
	PMN/PMA2	0.016	0.973	0.097	0
	ERA	0.022	0.008	0.638	0
	Type-72	0.1	0	0.083	0.555
CBWLSU	Clutter	0.887	0.078	0.319	0.352
	PMN/PMA2	0.019	0.877	0.097	0
	ERA	0.018	0.043	0.541	0
	Type-72	0.074	0	0.042	0.648
DOMINODL	Clutter	0.888	0.078	0.319	0.352
	PMN/PMA2	0.019	0.877	0.097	0
	ERA	0.018	0.043	0.54	0
	Type-72	0.074	0	0.042	0.648

TABLE VI
COMPUTATIONAL TIMES FOR DL ALGORITHMS

	DOMINODL	CBWLSU	ODL	K-SVD	LRSDL
Time (seconds)	1.75[1]	16.49	5.75	25.8	1057

[1] Blue denotes the best performance among all DL algorithms

DL approaches. However, in the case of K-SVD, the P_{CC} reduces by ∼30% and ∼10% for ERA and Type-72, respectively. Clearly, the reconstruction and correct classification of range profiles using batch algorithms such as K-SVD is strongly affected by a non-optimal choice of DL input parameters. As discussed earlier in Section V-C, this degradation is likely due to the influence of K rather than N_t.

C. Computational Efficiency

We used MATLAB 2016a platform on an 8-Core CPU Windows 7 desktop PC to clock the times for DL algorithms. The ODL algorithm from [19] is implemented as `mex` executable and, therefore, already fine-tuned for speed. For K-SVD, we employed the efficient implementation from [52] to improve computational speed. Table VI lists the execution times of the five DL approaches. Here, the parameters were optimally selected for all the algorithms. The LRSDL is the slowest of all, while ODL is more than four times faster than K-SVD. The CBWLSU provided better classification results but is three times slower than ODL. This could be because the dictionary update step always considers the entire previous training set elements that correlate with only one new element (i.e., there is no mini-batch strategy). This makes the convergence in CBWLSU more challenging.

The DOMINODL is the fastest DL method clocking three times speeder than ODL and 15 times than K-SVD. This is because the DOMINODL updates the dictionary by evaluating only a mini-batch of previous elements (instead of all of them as in CBWLSU) that correlate with a mini-batch of several new elements (CBWLSU uses just one new element). Furthermore, DOMINODL drops out the unused elements leading to a faster convergence. We note that, unlike ODL and K-SVD implementations, we did not use `mex` executables of DOMINODL, which can further shorten current execution times. From Table VI, the reduction in DOMINODL computational time over K-SVD is $((25.8-1.75)\times 100)/25.8 \approx 93\%$. The reduction for ODL and CBWLSU is computed similarly as 8% and 36%, respectively.

The computational bottleneck of mines classification lies in the training times. In comparison, the common steps of sparse decomposition and SVM-based classification during testing take just 0.4 and 1 s, respectively, for an entire survey (1 m × 1 m area with 2500 range profiles). Thus, time taken per range profile is ∼0.59 ms. The average scan rate of our GPR system is 0.19 m/s (or 1 cm/52.1 ms). This can go as high as 2.7 m/s (or 1 cm/3.61 ms) in other GPRs used for landmines application. Therefore, the test times do not impose much computational cost.

TABLE VII
CONFUSION MATRICES USING A COARSE SELECTION FOR $\mathbf{Y}$ INCLUDING SRC

		Clutter	PMN/PMA2	ERA	Type-72
SRC	Clutter	0.912	0.017	0.25	0.925
	PMN/PMA2	0.037	0.456[1]	0.042	0
	ERA	0.025	0.526	0.278	0.074
	Type-72	0.025	0	0.43	0
DOMINODL	Clutter	0.729	0.017	0.083	0.241
	PMN/PMA2	0.041	0.982	0.139	0
	ERA	0.054	0	0.667	0.074
	Type-72	0.176	0	0.111	0.685
CBWLSU	Clutter	0.584	0	0.125	0.185
	PMN/PMA2	0.063	0.982	0.153	0.11
	ERA	0.106	0	0.625	0.185
	Type-72	0.247	0.017	0.097	0.518
ODL	Clutter	0.71	0.035	0.153	0.259
	PMN/PMA2	0.036	0.912	0.069	0.074
	ERA	0.088	0.008	0.667	0.074
	Type-72	0.165	0.044	0.111	0.593
K-SVD	Clutter	0.617	0	0.111	0.148
	PMN/PMA2	0.044	0.982	0.194	0.056
	ERA	0.113	0	0.667	0.241
	Type-72	0.226	0.017	0.027	0.444

[1] Gray denotes the $\mathbf{P_{CC}}$ value for a specified class and DL algorithm

D. Comparison With Sparse-Representation-Based Classification

We compared our proposed DL-based approach with the SRC method proposed in [11]. The SRC needs a labeled dictionary but the dictionary that we learn from $\mathbf{Y}$ does not have label information anymore thereby making SRC infeasible here. Therefore, we adopt the following steps for a reasonable comparison of the two methods. We feed SRC with $\mathbf{Y}$ as the dictionary $\mathbf{D}$. As indicated in Section IV-C, a meticulously selected collection of mines/clutter responses as $\mathbf{Y}$ is meaningful for comparing different DL approaches. But it does not highlight the benefits of employing DL *per se*. Therefore, we generate a coarser selection, i.e., more profiles than the handpicked case, as $\mathbf{Y}$ for both approaches. Table VII shows the confusion matrix for the residual-based classification along with the proposed DL-based approaches. The DL-based mine

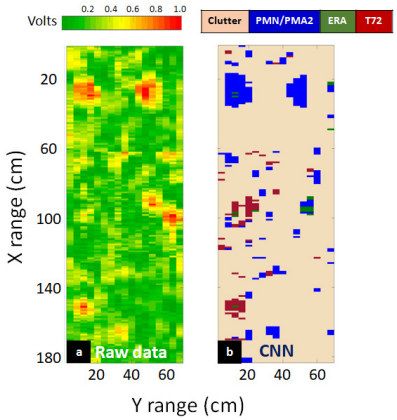

Fig. 10. (a) Raw data at 15-cm depth. (b) Classification maps of the same area containing six buried landmines using CNN-based classification.

classification is consistent with previous results—even better with DOMINODL—but with some tradeoff of decreasing clutter accuracy. The accuracy of the residual-based classifier is severely degraded for all mine classes, dropping by at least 45%, 39%, and 55% for PMA/PMA2, ERA, and T72, respectively. This renders the increase in clutter classification accuracy of this method not usable.

E. Deep-Learning-Based Classification

The core idea of SR-based classification is largely based on the assumption that signals are linear combinations of a few atoms. In practice, this is often not the case. This has led to a few recent works [65] that suggest employing deep learning for radar target classification. However, these techniques require significantly large data sets for training.

We compared classification results of our methods with a deep learning approach. In particular, we constructed a CNN because these networks are known to efficiently exploit structural or locational information in the data and yield comparable learning potential with far fewer parameters [66]. We modeled our proposed CNN framework as a classification problem wherein each class denotes the type of mine or clutter. The training data set for our CNN structure is the matrix $\mathbf{Y}$ (see Section IV). Building up a synthetic database is usually an option for creating (or extending) a training set for deep learning applications. However, accurately modeling a GPR scenario is still an ongoing challenge in the GPR community because of the difficulties in accurately reproducing the soil inhomogeneities (and variabilities), the surface and underground clutter, the antenna coupling and ringing effects, and so on. Even though some applications have been promising [67], this remains a cumbersome task.

The input layer of our CNN took 1-D sample set of size 211. It was followed by two convolutional layers with 20 and 5 filters of size 20 and 10, respectively. The output layer consisted of four units wherein the network classifies the given input data as clutter or one of the three mines. There were rectified linear units (ReLU) after each convolutional layer; the ReLU function is given by $\text{ReLU}(x) = \max(x, 0)$ [68]. The architecture of the CNN was selected through an arduous process of testing many combinations of layers/filters and hyperparameters, which would lead to better accuracy during training. A deeper network slightly increased the accuracy in the training phase but led to poorer performance when classifying new data (i.e., the test set $\mathbf{Y}_{\text{test}}$). Since our data are limited, adding more layers (i.e., more weights) only led to overfitting and made the network incapable to generalize on new data sets. A multi-dimensional CNN formed by clustering 2-D and 3-D data would have further reduced the training set. Augmenting the data was also envisioned but commonly used transformations such as scaling/rotations are not useful in our case because the mines were always in the same inclination and their dimension defines the class itself. We also attempted adding different levels of noise but this did not lead to better results, considering that the available data are already very noisy.

TABLE VIII
CONFUSION MATRIX FOR CNN-BASED CLASSIFICATION

	Clutter	PMN/PMA2	ERA	Type-72
Clutter	0.909	0.14	0.016	0.574
PMN/PMA2	0.032	0.807	0.181	0
ERA	0	0.053	0.319	0.315
Type-72	0.033	0	0.111	0.370

We trained the network with the labeled training set $\mathbf{Y}$, selecting ∼20% of the training data for validation. Specifically, the validation set employed 100, 25, 25, and 25 range profiles for clutter, PMN/PMA2, ERA, and Type-72, respectively. We used a stochastic gradient descent algorithm for updating the network parameters with the learning rate of 0.001 and mini-batch size of 20 samples for 2000 epochs.

We realized the proposed network in TensorFlow on a Windows 7 PC with 8-core CPU. The network training took 3.88 min. Fig. 10 shows the classification map obtained using the CNN. The corresponding confusion matrix is listed in Table VIII. We note that the CNN classifier shows worse P_{CC} than our SR-based techniques, particularly for ERA and Type-72 target classes.

F. Classification With Reduced Range Samples

We now analyze the robustness of our DL-based adaptive classification method to the reduction of the number of samples in the raw data. Assuming that the collected data $\mathbf{Y}_{\text{TEST}}$ are sparse in dictionary $\mathbf{D}$, we undersampled the original raw data $\mathbf{Y}_{\text{TEST}}$ in range to obtain its row-undersampled version $\widetilde{\mathbf{Y}}_{\text{TEST}}$ by randomly reducing the samples. We then applied the same random sampling pattern to the dictionary $\mathbf{D}$ for obtaining the sparse coefficients. We also analyzed the CNN classifier when the signals are randomly reduced in the same way. Fig. 11 shows the classification map for all DL approaches when the sampling is reduced by 50%. Table IX clubs together the confusion matrices when undersampling by 25%, 50%, and 75%.

In comparison to the results in Table IV which used all samples of the raw data, the DL approaches maintain similar classifier performance even when we reduce the samples by

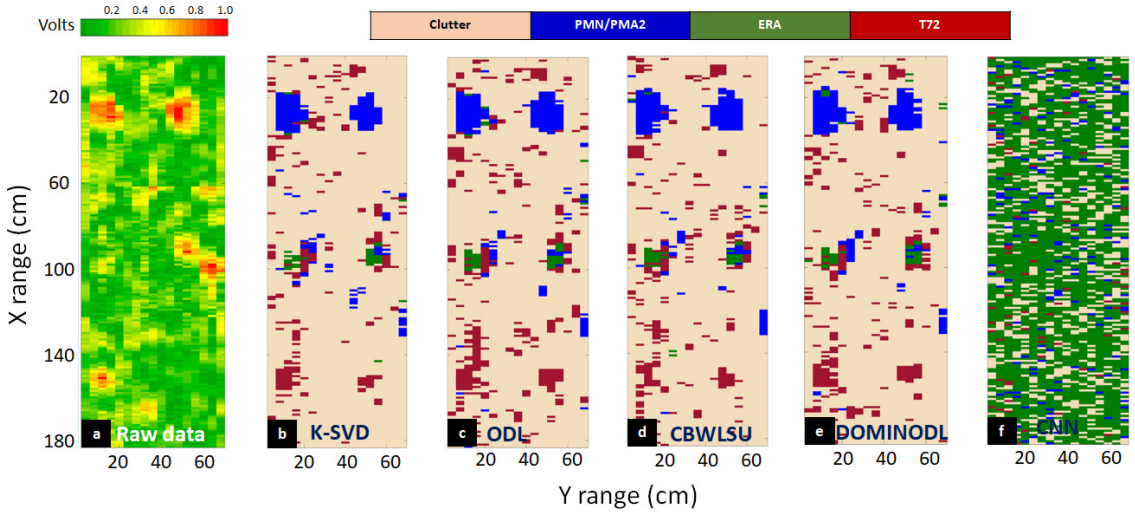

Fig. 11. (a) Raw data at 15-cm depth. The classification maps of the same area containing six buried landmines using an SR-based approach with dictionary learned using (b) K-SVD, (c) ODL, (d) CBWLSU, and (e) DOMINODL algorithms. The input parameters were optimally selected and the number of samples were reduced by 50%. (f) Corresponding result with reduced samples for CNN-based classification.

TABLE IX
CONFUSION MATRICES FOR DIFFERENT DL ALGORITHMS AND CNN WITH REDUCED SIGNAL SAMPLES

		25% Reduction				50% Reduction				75% Reduction			
		Clutter	PMN/PMA2	ERA	Type-72	Clutter	PMN/PMA2	ERA	Type-72	Clutter	PMN/PMA2	ERA	Type-72
K-SVD	Clutter	0.892	0.078	0.319	0.389	0.882	0.026	0.291	0.37	0.877	0.061	0.402	0.426
	PMN/PMA2	0.021	0.921	0.153	0.055	0.018	0.947	0.153	0.037	0.02	0.912	0.125	0.074
	ERA	0.021	0	0.486	0.018	0.021	0.026	0.5	0	0.021	0.026	0.333	0.018
	Type-72	0.065	0	0.041	0.537	0.078	0	0.055	0.592	0.08	0	0.138	0.481
ODL	Clutter	0.872	0.088	0.208	0.315	0.868	0	0.208	0.333	0.862	0.02	0.319	0.296
	PMN/PMA2	0.021	0.973	0.152	0	0.021	0.965	0.18	0.018	0.023	0.964	0.138	0.018
	ERA	0.018	0.017	0.527	0.018	0.018	0.035	0.5	0	0.021	0.008	0.416	0.074
	Type-72	0.087	0	0.111	0.666	0.09	0	0.111	0.648	0.091	0	0.125	0.611
CBWLSU	Clutter	0.871	0.026	0.194	0.351	0.872	0.017	0.25	0.40	0.855	0.088	0.388	0.370
	PMN/PMA2	0.024	0.956	0.139	0	0.023	0.973	0.111	0	0.024	0.974	0.111	0
	ERA	0.025	0.017	0.541	0	0.02	0.008	0.541	0	0.027	0.017	0.333	0.018
	Type-72	0.79	0	0.125	0.648	0.083	0	0.097	0.592	0.091	0	0.125	0.611
DOMINODL	Clutter	0.88	0.017	0.236	0.277	0.868	0.035	0.194	0.296	0.864	0.035	0.278	0.444
	PMN/PMA2	0.022	0.964	0.138	0	0.023	0.929	0.138	0	0.027	0.938	0.152	0
	ERA	0.018	0.017	0.527	0	0.024	0.035	0.527	0.018	0.026	0.026	0.5	0
	Type-72	0.078	0	0.097	0.722	0.083	0	0.138	0.685	0.082	0	0.069	0.556
CNN	Clutter	0.708	0.359	0.236	0.407	0.265	0.166	0.645	0.148	0.162	0.105	0.647	0.129
	PMN/PMA2	0.026	0.41	0.097	0.018	0.062	0.096	0.069	0.018	0.015	0.061	0.013	0
	ERA	0	0.21	0.5	0.426	0	0.72	0.73	0.75	0	0.71	0.708	0.759
	Type-72	0.029	0.017	0.069	0.148	0.027	0.088	0.014	0.074	0.17	0.12	0.12	0.11

75% (i.e., just 52 samples in total). In contrast, the CNN classifier result that is already heavily compromised with a reduction of 25% fails completely for 50% and 75% sampling rate. Reducing the number of signal samples when using a dictionary which minimizes the number of nonzero entries in the SR still assures an exact reconstruction of the signal itself and, consequently, its correct classification. The features for classifying the traces are robust to the reduction of the original samples. Deep learning strategies use the signal samples directly as classification features. They also require enormous amount of data for training. Therefore, the degradation in their performance is expected. The confusion matrix in Table IX indicates that CNN has the highest P_{CC} for ERA. This is a false trail because the network mis-classified almost every pixel as ERA. Overall, DOMINODL and CBWLSU provide excellent results for small mines. However, as seen earlier, CBWLSU is not very well-suited for real-time operation because of longer execution times.

We also assessed the performance of different methods when, instead of sampling fewer range samples per profile, we include all samples in every range profile but reduce the overall number of training set elements randomly from 926 to 694, 464, and 232 range profiles (which, respectively, correspond to 25%, 50%, and 75% reduction). From the corresponding confusion matrices listed in Table X, we note that CNN-based classification results have improved with respect to Table IX. However, the classification accuracy of CNN is still poorer than the DL-based classification. In general, all

TABLE X
CONFUSION MATRICES FOR DIFFERENT DL ALGORITHMS AND CNN WITH REDUCED TRAINING SET ELEMENTS

		25% Reduction				50% Reduction				75% Reduction			
		Clutter	PMN/PMA2	ERA	Type-72	Clutter	PMN/PMA2	ERA	Type-72	Clutter	PMN/PMA2	ERA	Type-72
K-SVD	Clutter	0.887	0.044	0.236	0.370	0.868	0	0.417	0.556	0.885	0.097	0.444	0.426
	PMN/PMA2	0.019	0.912	0.139	0	0.023	0.912	0.153	0.037	0.019	0.746	0.014	0.130
	ERA	0.023	0.044	0.528	0.019	0.010	0.070	0.306	0.019	0.012	0.088	0.319	0
	Type-72	0.071	0	0.097	0.611	0.099	0.018	0.125	0.389	0.084	0.070	0.222	0.444
ODL	Clutter	0.865	0.009	0.167	0.315	0.865	0.009	0.375	0.611	0.857	0.070	0.375	0.352
	PMN/PMA2	0.016	0.939	0.181	0	0.023	0.991	0.111	0	0.027	0.851	0.056	0.093
	ERA	0.020	0.053	0.583	0.019	0.011	0	0.431	0	0.016	0.079	0.444	0
	Type-72	0.096	0	0.069	0.667	0.101	0	0.083	0.389	0.101	0	0.125	0.556
CBWLSU	Clutter	0.865	0.035	0.250	0.426	0.874	0.009	0.347	0.407	0.834	0.035	0.500	0.685
	PMN/PMA2	0.021	0.912	0.167	0	0.033	0.904	0.069	0.241	0.020	0.851	0.056	0
	ERA	0.026	0.053	0.542	0.019	0.014	0.088	0.500	0	0.015	0.105	0.403	0
	Type-72	0.088	0	0.042	0.556	0.080	0	0.083	0.352	0.131	0.089	0.042	0.315
DOMINODL	Clutter	0.841	0.026	0.222	0.296	0.864	0.053	0.333	0.574	0.815	0.175	0.431	0.519
	PMN/PMA2	0.021	0.921	0.167	0	0.023	0.868	0.069	0.074	0.027	0.746	0.069	0
	ERA	0.026	0.053	0.542	0.019	0.014	0.079	0.486	0	0.017	0.079	0.486	0
	Type-72	0.112	0	0.070	0.685	0.099	0	0.111	0.352	0.142	0	0.014	0.482
CNN	Clutter	0.838	0.026	0.040	0.333	0.819	0.088	0.047	0.315	0.644	0.035	0.109	0.463
	PMN/PMA2	0.062	0.868	0.208	0.537	0.059	0.851	0.319	0.463	0.080	0.693	0.181	0.019
	ERA	0	0.053	0.250	0.074	0	0	0.153	0.130	0	0.193	0.306	0.019
	Type-72	0.106	0.053	0.208	0.556	0.120	0.061	0.250	0.500	0.105	0.079	0.097	0

methods show performance degradation as the training set elements are reduced. Among the online DL methods, ODL is more robust to the range profile reduction than K-SVD.

VII. CONCLUSION

In this paper, we proposed effective online DL strategies for sparse decomposition of GPR traces of buried landmines. The online methods outperform K-SVD, thereby making them a good candidate for SR-based classification. Our algorithm DOMINODL is always the fastest providing near real-time performance and high clutter rejection while also maintaining classifier performance that is comparable to other online DL algorithms. DOMINODL and CBWLSU generally classify smaller mines better than ODL and K-SVD. Unlike previous works that rely on RMSE, we used metrics based on statistical inference to tune the DL parameters for enhanced operation.

Fast ODL computations pave the way toward cognition [69]–[71] in GPR operation, wherein the system uses previous measurements to optimize the processing performance and is capable of sequential sampling adaptation [72] based on the learned dictionary. For example, in a realistic landmine clearance campaign, an operator could gather the training measurements over a safe area next to the contaminated site, hypothetically placing some buried landmine simulants over it in order to have a faithful representation of the soil/targets interaction beneath the surface. In other words, this paper allows the operator to *calibrate* the acquisition by providing a good training set to learn the dictionary.

ACKNOWLEDGMENT

The authors would like to thank D. Mateos-Núñez for the valuable assistance in Section VI-E. They would also like to thank the colleagues at the Leibniz Institute for Applied Geophysics (LIAG), Hannover, Germany, for their support during the measurement campaign.

REFERENCES

[1] H. M. Jol, Ed., *Ground Penetrating Radar Theory and Applications*. Amsterdam, The Netherlands: Elsevier, 2009.
[2] D. J. Daniels, "Ground penetrating radar," in *Encyclopedia of RF and Microwave Engineering*. Edison, NJ, USA: IET, 2004.
[3] *Landmine Monitor 2017*, Int. Campaign Ban Landmines—Cluster Munition Coalition, Geneva, Switzerland, 2017.
[4] F. Giovanneschi, M. A. González-Huici, and U. Uschkerat, "A parametric analysis of time and frequency domain GPR scattering signatures from buried landmine-like targets," *Proc. SPIE*, vol. 8709, Jun. 2013, Art. no. 870914.
[5] M. A. González-Huici, I. Catapano, and F. Soldovieri, "A comparative study of GPR reconstruction approaches for landmine detection," *IEEE J. Sel. Topics Appl. Earth Observat. Remote Sens.*, vol. 7, no. 12, pp. 4869–4878, Dec. 2014.
[6] M. A. González-Huici and F. Giovanneschi, "A combined strategy for landmine detection and identification using synthetic GPR responses," *J. Appl. Geophys.*, vol. 99, pp. 154–165, Dec. 2013.
[7] 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 Trans. Geosci. Remote Sens.*, vol. 52, no. 3, pp. 1539–1550, Mar. 2014.
[8] I. Giannakis, A. Giannopoulos, and A. Yarovoy, "Model-based evaluation of signal-to-clutter ratio for landmine detection using ground-penetrating radar," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3564–3573, Jun. 2016.
[9] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.
[10] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.
[11] F. Giovanneschi and M. A. González-Huici, "A preliminary analysis of a sparse reconstruction based classification method applied to GPR data," in *Proc. Int. Workshop Adv. Ground Penetrating Radar*, Jul. 2015, pp. 1–4.
[12] F. Giovanneschi, K. V. Mishra, M. A. González-Huici, Y. C. Eldar, and J. H. G. Ender, "Online dictionary learning aided target recognition in cognitive GPR," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2017, pp. 4813–4816.

[13] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[14] Y. C. Eldar, *Sampling Theory: Beyond Bandlimited Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2015.

[15] M. Elad, *Sparse and Redundant Representations—From Theory to Applications in Signal and Image Processing*. New York, NY, USA: Springer-Verlag, 2010.

[16] S. Arora, R. Ge, and A. Moitra, "New algorithms for learning incoherent and overcomplete dictionaries," in *Proc. Conf. Learn. Theory*, 2014, pp. 779–806.

[17] Y. C. Eldar and G. Kutyniok, *Compressed Sensing: Theory and Applications*. Cambridge, U.K.: Cambridge Univ. Press, 2012.

[18] K. Engan, S. O. Aase, and J. H. Husoy, "Method of optimal directions for frame design," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 5, Mar. 1999, pp. 2443–2446.

[19] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proc. 26th Annu. Int. Conf. Mach. Learn.*, 2009, pp. 689–696.

[20] W. Shao, A. Bouzerdoum, and S. L. Phung, "Sparse representation of GPR traces with application to signal classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 7, pp. 3922–3930, Jul. 2013.

[21] I. M. Chakravarti, R. G. Laha, and J. Roy, *Handbook of Methods of Applied Statistics*, vol. 1. Hoboken, NJ, USA: Wiley, 2004.

[22] A. Dvoretzky, J. Kiefer, and J. Wolfowitz, "Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator," *Ann. Math. Statist.*, vol. 27, no. 3, pp. 642–669, 1956.

[23] P. Massart, "The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality," *Ann. Probab.*, vol. 18, no. 3, pp. 1269–1283, 1990.

[24] Y. Naderahmadian, S. Beheshti, and M. A. Tinati, "Correlation based online dictionary learning algorithm," *IEEE Trans. Signal Process.*, vol. 64, no. 3, pp. 592–602, Feb. 2016.

[25] J. N. Wilson, P. Gader, W.-H. Lee, H. Frigui, and K. C. Ho, "A large-scale systematic evaluation of algorithms using ground-penetrating radar for landmine detection and discrimination," *IEEE Trans. Geosci. Remote Sens.*, vol. 45, no. 8, pp. 2560–2572, Aug. 2007.

[26] L. Robledo, M. Carrasco, and D. Mery, "A survey of land mine detection technology," *Int. J. Remote Sens.*, vol. 30, no. 9, pp. 2399–2410, 2009.

[27] S. Lameri, F. Lombardi, P. Bestagini, M. Lualdi, and S. Tubaro, "Landmine detection from GPR data using convolutional neural networks," in *Proc. Eur. Signal Process. Conf.*, 2017, pp. 508–512.

[28] L. E. Besaw and P. J. Stimac, "Deep convolutional neural networks for classifying GPR B-scans," *Proc. SPIE*, vol. 9454, May 2015, Art. no. 945413.

[29] E. D. Sontag, "VC dimension of neural networks," *NATO ASI F Comput. Syst. Sci.*, vol. 168, pp. 69–96, Jan. 1998.

[30] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 27:1–27:27, 2011.

[31] M. Aharon, M. Elad, and A. Bruckstein, "K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[32] M. S. Lewicki and T. J. Sejnowski, "Learning overcomplete representations," *Neural Comput.*, vol. 12, no. 2, pp. 337–365, 2000.

[33] D. A. Spielman, H. Wang, and J. Wright, "Exact recovery of sparsely-used dictionaries," in *Proc. Conf. Learn. Theory*, 2012, pp. 37.1–37.18.

[34] A. Agarwal, A. Anandkumar, P. Jain, P. Netrapalli, and R. Tandon, "Learning sparsely used overcomplete dictionaries," in *Proc. Conf. Learn. Theory*, 2014, pp. 123–137.

[35] Z. Jiang, Z. Lin, and L. S. Davis, "Label consistent K-SVD: Learning a discriminative dictionary for recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 11, pp. 2651–2664, Nov. 2013.

[36] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2691–2698.

[37] K. R. Varshney, M. Çetin, J. W. Fisher, and A. S. Willsky, "Sparse representation in structured dictionaries with application to synthetic aperture radar," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3548–3561, Aug. 2008.

[38] Y. Suo, M. Dao, U. Srinivas, V. Monga, and T. D. Tran, "Structured dictionary learning for classification," 2014, *arXiv:1406.1943*. [Online]. Available: https://arxiv.org/abs/1406.1943

[39] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based Fisher discrimination dictionary learning for image classification," *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 209–232, Sep. 2014.

[40] L. Li, S. Li, and Y. Fu, "Learning low-rank and discriminative dictionary for image classification," *Image Vis. Comput.*, vol. 32, no. 10, pp. 814–823, 2014.

[41] I. Ramirez, P. Sprechmann, and G. Sapiro, "Classification and clustering via dictionary learning with structured incoherence and shared features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 3501–3508.

[42] S. Kong and D. Wang, "A dictionary learning approach for classification: Separating the particularity and the commonality," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 186–199.

[43] S. Gao, I. W.-H. Tsang, and Y. Ma, "Learning category-specific dictionary and shared dictionary for fine-grained image categorization," *IEEE Trans. Image Process.*, vol. 23, no. 2, pp. 623–634, Feb. 2014.

[44] C. Rusu, "On learning with shift-invariant structures," 2018, *arXiv:1812.01115*. [Online]. Available: https://arxiv.org/abs/1812.01115

[45] L. H. Nguyen and T. D. Tran, "Separation of radio-frequency interference from SAR signals via dictionary learning," in *Proc. IEEE Radar Conf.*, Apr. 2018, pp. 908–913.

[46] C. Garcia-Cardona and B. Wohlberg, "Convolutional dictionary learning: A comparative review and new algorithms," *IEEE Trans. Comput. Imag.*, vol. 4, no. 3, pp. 366–381, Sep. 2018.

[47] T. H. Vu and V. Monga, "Fast low-rank shared dictionary learning for image classification," *IEEE Trans. Image Process.*, vol. 26, no. 11, pp. 5160–5175, Nov. 2017.

[48] B. Dumitrescu and P. Irofti, *Dictionary Learning Algorithms and Applications*. New York, NY, USA: Springer-Verlag, 2018.

[49] J. Sulam, B. Ophir, M. Zibulevsky, and M. Elad, "Trainlets: Dictionary learning in high dimensions," *IEEE Trans. Signal Process.*, vol. 64, no. 12, pp. 3180–3193, Jun. 2016.

[50] J. Chen, L. Jiao, W. Ma, and H. Liu, "Unsupervised high-level feature extraction of SAR imagery with structured sparsity priors and incremental dictionary learning," *IEEE Geosci. Remote Sens. Lett.*, vol. 13, no. 10, pp. 1467–1471, Oct. 2016.

[51] A. Beck and M. Teboulle, "A fast iterative shrinkage-thresholding algorithm for linear inverse problems," *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[52] R. Rubinstein, M. Zibulevsky, and M. Elad, "Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit," *CS Technion*, vol. 40, no. 8, pp. 1–15, 2008.

[53] N. Zhou and J. Fan, "Jointly learning visually correlated dictionaries for large-scale visual recognition applications," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 4, pp. 715–730, Apr. 2014.

[54] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Found. Trends Mach. Learn.*, vol. 3, no. 1, pp. 1–122, Jan. 2011.

[55] M. R. Osborne, B. Presnell, and B. A. Turlach, "A new approach to variable selection in least squares problems," *IMA J. Numer. Anal.*, vol. 20, no. 3, pp. 389–403, 2000.

[56] J. Leckebusch, "Comparison of a stepped-frequency continuous wave and a pulsed GPR system," *Archaeol. Prospection*, vol. 18, no. 1, pp. 15–25, 2011.

[57] C. Warren, A. Giannopoulos, and I. Giannakis, "An advanced GPR modelling framework: The next generation of gprMax," in *Proc. IEEE Int. Workshop Adv. Ground Penetrating Radar*, Jul. 2015, pp. 1–4.

[58] M. G. M. Hussain, "Principles of high-resolution radar based on nonsinusoidal waves. III. Radar-target reflectivity model," *IEEE Trans. Electromagn. Compat.*, vol. 32, no. 2, pp. 144–152, May 1990.

[59] D. Pasculli and G. Manacorda, "Real-time, pseudo real-time and stroboscopic sampling in time-domain GPRs," in *Proc. IEEE Int. Workshop Adv. Ground Penetrating Radar*, Jul. 2015, pp. 1–4.

[60] A. Bystrov and M. Gashinova, "Analysis of stroboscopic signal sampling for radar target detectors and range finders," *IET Radar, Sonar Navigat.*, vol. 7, no. 4, pp. 451–458, 2013.

[61] M. A. González-Huici, "Accurate ground penetrating radar numerical modeling for automatic detection and recognition of antipersonnel landmines," Ph.D. dissertation, Dept. Geosci., Univ. Landesbibliothek Bonn, Bonn, Germany, 2013.

[62] H. Hongxing, J. M. Bioucas-Dias, and V. Katkovnik, "Interferometric phase image estimation via sparse coding in the complex domain," *IEEE Trans. Geosci. Remote Sens.*, vol. 53, no. 5, pp. 2587–2602, May 2015.

[63] V. Glivenko, "Sulla determinazione empirica della legge di probabilità," (in Italian), *Giornale Dell'Istituto Italiano Degli Attuari*, no. 4, p. 9299, 1933.

[64] F. P. Cantelli, "Sulla determinazione empirica della legge di probabilità," (in Italian), *Giornale Dell'Istituto Italiano Degli Attuari*, no. 4, 1933, Art. no. 221424.
[65] T. H. Vu, L. Nguyen, T. Guo, and V. Monga, "Deep network for simultaneous decomposition and classification in UWB-SAR imagery," in *Proc. IEEE Radar Conf.*, Apr. 2018, pp. 553–558.
[66] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 1440–1448.
[67] I. Giannakis, A. Giannopoulos, and C. Warren, "A realistic FDTD numerical modeling framework of ground penetrating radar for landmine detection," *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, vol. 9, no. 1, pp. 37–51, Jan. 2016.
[68] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 1929–1958, 2014.
[69] K. V. Mishra and Y. C. Eldar, "Sub-Nyquist radar: Principles and prototypes," in *Compressed Sensing in Radar Signal Processing*, A. D. Maio, Y. C. Eldar, and A. Haimovich, Eds. Cambridge, U.K.: Cambridge Univ. Press, 2019.
[70] K. V. Mishra *et al.*, "Cognitive sub-Nyquist hardware prototype of a collocated MIMO radar," in *Proc. IEEE Int. Workshop Compressed Sens. Theory Appl. Radar, Sonar Remote Sens.*, Sep. 2016, pp. 56–60.
[71] K. V. Mishra and Y. C. Eldar, "Performance of time delay estimation in a cognitive radar," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2017, pp. 3141–3145.
[72] K. V. Mishra, A. Kruger, and W. F. Krajewski, "Compressed sensing applied to weather radar," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2014, pp. 1832–1835.

Fabio Giovanneschi received the M.Sc. degree in telecommunications engineering from the University of Pisa, Pisa, Italy, in 2010. He is currently concluding his Ph.D. project in the University of Siegen, Siegen, Germany, with a work entitled "Online Dictionary Learning for Classification of Antipersonnel Landmines using Ground Penetrating Radar" carried within the cooperation with the Technion Institute of Technology, Haifa, Israel.

In 2011, he was an Applied Geophysicist with the company IDS, Pisa. In 2012, he with the Ultra Wide Band Team, Fraunhofer FHR, Wachtberg, Germany, dealing with buried landmine classification using GPR. In 2015, he joined the Cognitive Radar Department, Fraunhofer FHR, and become a member of the Adaptive Sensing and Perception (ASP) Group. From 2015 to 2018, he was involved in CS-based MIMO radar imaging for the Radar for Warning and Information System (RAWIS) Project. He is actually contributing in upcoming IET books about dictionary learning and micro-Doppler analysis. His research interests include sparse representation, compressive sensing, dictionary learning, machine learning, classification, micro-Doppler analysis, and MIMO radar signal processing

Mr. Giovanneschi serves as a voluntary reviewer for IET and TGRS journals. He has been nominated for the best paper award at the International Ground Penetrating Radar Workshop (IWAGPR2015) for his paper entitled "A Preliminary Analysis of a Sparse Representation Based Classification Method Applied to GPR Data."

Kumar Vijay Mishra (S'08–M'15–SM'18) received the B.Tech. degree *(summa cum laude) (Hons.)* in electronics and communications engineering from the National Institute of Technology, Hamirpur, India, in 2003, the M.S. degree in electrical and computer engineering from Colorado State University, Fort Collins, CO, USA, in 2012, and the Ph.D. degree in electrical and computer engineering and the M.S. degree in mathematics from The University of Iowa, Iowa City, IA, USA, in 2015, while working on the NASA Global Precipitation Mission (GPM) Ground Validation (GV) weather radars.

He has over 15 years of experience in research and development of various radar systems. From 2003 to 2007, he was involved in military surveillance radars as a Research Scientist at the Electronics and Radar Development Establishment (LRDE), Defence Research and Development Organisation (DRDO), Bengaluru, India. From 2015 to 2017, he was an Andrew and Erna Finci Viterbi and Lady Davis Post-Doctoral Fellow with the Viterbi Faculty of Electrical Engineering, Technion–Israel Institute of Technology, Haifa, Israel. He is currently a Visiting Scholar with The University of Iowa and a Technical Advisor to the automotive radar start-up Hertzwell, Singapore. His research interests include radar systems, signal processing, remote sensing, and electromagnetics.

Dr. Mishra was a recipient of the Royal Meteorological Society Quarterly Journal Editor's Prize in 2017, the Lady Davis Fellowship from 2016 to 2017, the Andrew and Erna Finci Viterbi Fellowship (twice awarded) in 2015 and 2016, the Technion Excellent Undergraduate Adviser Award in 2017, the DRDO LRDE Scientist of the Year Award in 2006, the NITH Director's Gold Medals in 2003, and the NITH Best Student Award in 2003.

Maria Antonia Gonzalez-Huici received the M.Sc. degree in theoretical physics from the Autonomous University of Madrid, Madrid, Spain, in 2002, and the Ph.D. degree *(cum laude)* in geophysics from the University of Bonn, Bonn, Germany, in 2013. Her Ph.D. thesis was entitled Accurate Ground Penetrating Radar (GPR) Numerical Modeling for Automatic Detection and Recognition of Antipersonnel Landmines.

From 2002 to 2005, she participated in the Bonn International Graduate School and carried out postgraduate studies in cosmology, astrophysics, and applied geophysics. In May 2005, she joined the Fraunhofer FHR, Wachtberg, Germany, where she has been leading the Adaptive Sensing and Perception (ASP) Team, Cognitive Radar Department, since March 2015. Her research interests include array processing and beamforming, MIMO radar and waveform design, statistical signal processing, machine learning, and compressive sensing. In these fields, she has published more than 25 conference and journal papers and has filed two patents.

Dr. Gonzalez-Huici is also a member of the scientific committees of the COSERA and IWAGPR workshops and serves as a voluntary reviewer for a number of international journals. She has been involved in the research team of a number of national and international projects, some of them as a PI, including projects funded by the German Ministry of Defense, the German Ministry of Education and Research, the European Defense Agency, as well as several industry projects.

Yonina C. Eldar (S'98–M'02–SM'07–F'12) received the B.Sc. degree in physics and the B.Sc. degree in electrical engineering from Tel-Aviv University (TAU), Tel-Aviv, Israel, in 1995 and 1996, respectively, and the Ph.D. degree in electrical engineering and computer science from the Massachusetts Institute of Technology (MIT), Cambridge, MI, USA, in 2002.

From January 2002 to July 2002, she was a Post-Doctoral Fellow with the Digital Signal Processing Group, MIT. She is currently a Professor with the Department of Mathematics and Computer Science, Weizmann Institute of Science, Rehovot, Israel. Previously, she was a Professor with the Department of Electrical Engineering, Technion, Haifa, Israel, where she held the Edwards Chair in Engineering. She is also a Visiting Professor with MIT, a Visiting Scientist with the Broad Institute, Cambridge, MA, USA, and an Adjunct Professor with Duke University, Durham, NC, USA, and was a Visiting Professor with Stanford University, Stanford, CA, USA. She has authored the book *Sampling Theory: Beyond Bandlimited Systems* and coauthored the books *Compressed Sensing* and *Convex Optimization Methods in Signal Processing and Communications* (all published by Cambridge University Press).

Dr. Eldar is a member of the Israel Academy of Sciences and Humanities in 2017 and a EURASIP Fellow. She was a member of the Young Israel Academy of Science and Humanities and the Israel Committee for Higher Education. She has received numerous awards for excellence in research and teaching, including the IEEE Signal Processing Society Technical Achievement Award in 2013, the IEEE/AESS Fred Nathanson Memorial Radar Award in 2014, and the IEEE Kiyo Tomiyasu Award in 2016. She was a Horev Fellow of the Leaders in Science and Technology Program at the Technion and an Alon Fellow. She received the Michael Bruno Memorial Award from the Rothschild Foundation, the Weizmann Prize for Exact Sciences, the Wolf Foundation Krill Prize for Excellence in Scientific Research, the Henry Taub Prize for Excellence in Research (twice), the Hershel Rich Innovation Award (three times), the Award for Women with Distinguished Contributions, the Andre and Bella Meyer Lectureship, the Career Development Chair at the Technion, the Muriel & David Jacknow Award for Excellence in Teaching, and the Technion's Award for Excellence in Teaching (twice). She has also received several best paper awards and best demo awards together with her research students and colleagues, including the SIAM outstanding Paper Prize and the IET Circuits, Devices and Systems Premium Award, and was selected as one of the 50 most influential women in Israel. She is the Editor-in-Chief of *Foundations and Trends in Signal Processing* and a member of the IEEE Sensor Array and Multichannel Technical Committee and serves on several other IEEE committees.

Joachim H. G. Ender (F'14) received the Dr.-Ing. degree.

He has held the positions as the Director of the Fraunhofer-Institute for High Frequency Physics and Radar Technology FHR, Wachtberg, Germany, and as the Chair for High Frequency Sensors and Radar Techniques at the University of Siegen, Siegen, Germany, until July 2016. After his diploma in mathematics/physics, he performed research on various topics of radar science for more than 40 years. He started his carrier as a Young Scientist in 1976, was appointed as the Head of the department in 1999, and finally took over the function as the Director of FHR in 2003. After his retirement, he is still active as a Senior Scientist with FHR and as a Professor with the Center for Sensor Systems (ZESS), University Siegen. He has authored or co-authored numerous papers.

Dr. Ender was named an IEEE fellow in 2014 for contributions to multi-channel synthetic aperture radar and radar array signal processing. Among other prizes, e.g., awarded the GRS-S Transactions Prize Paper Award of the year 2006, he received from EURASIP the Group Technical Achievement Award—for contributions to array signal processing and multichannel synthetic aperture radar and the Best Paper Award for a journal publication on compressive sensing. He acted as a Chief Guest Editor for the special issue of the *IEEE Signal Processing Magazine* "Recent Advances in Radar Imaging." He was one of the founder members of the European Conference on Synthetic Aperture Radar (EUSAR) which takes place every two years since 1996 and the International Workshop on Compressed Sensing applied to Radar, Multimodal Sensing and Imaging (CoSeRa), starting in 2012.