Communications on Analysis and Computation

doi:10.3934/cac.2024012

DETECTION OF DISPERSIVE TARGETS IN SYNTHETIC APERTURE RADAR IMAGES WITH THE HELP OF DEEP LEARNING

Sai Naga Vamshi Chidara[1], Mikhail Gilman[2], Rageeni Sah[3] and Semyon Tsynkov[*2]

[1]Amazon.com, Inc., WA, USA

[2]Department of Mathematics, North Carolina State University, Raleigh, NC, USA

[3]American Express Company, NY, USA

(Communicated by Rongfang Gong)

Abstract. Detection of dispersive targets in SAR is obstructed by ambiguity between the scattering delay and increase of the signal travel time. To overcome this ambiguity, we design image classifiers based on deep learning that can discriminate between the instantaneous and delayed targets in coordinate-delay SAR images (cdSAR). Both targets and images are modeled stochastically. This allows one to account for the speckle effect but still leaves a possibility of misclassification.

Previously, we considered a simplified setup where the cdSAR images contained only two coordinates: range and delay, while the cross-range coordinate of the scatterer to be classified was known. In this work, we incorporate a third coordinate and determine the tolerance of the previously developed 2D classifier to inaccuracies in the cross-range position of the scatterer. We found this tolerance to be shorter than the cross-range resolution. Hence, increasing the robustness of the classifier appears of merit.

We increase robustness of the classifiers by training them on 3D "slabs" of cdSAR images that span a finite cross-range interval. Compared to the original 2D classifier, this extension relaxes the requirement of knowing the cross-range position of the scatterer accurately. However, the improved robustness is not accompanied by a better classification performance, even though the measurements at the additional cross-range locations provide more information about the target. We conduct a series of numerical experiments to analyze several possible explanations of this effect.

2020 *Mathematics Subject Classification.* 35R30, 45Q05, 78A46, 86A22.

Key words and phrases. Delayed scattering of electromagnetic waves, coordinate-delay SAR images (cdSAR), range-delay ambiguity, target classification, convolutional neural network (CNN).

*Corresponding author: Semyon Tsynkov.

1. Introduction.

1.1. Delayed scattering and radar imaging.

Delayed scattering of electromagnetic waves can be caused by complex geometrical features of the scatterer, such as cavities or regular structures and/or special material properties [3, 4, 9]. Typically, we associate such features with man-made objects, so detection of scattering delay by means of remote sensing, e.g., radar imaging, is of interest for various applications.

The main obstacle to the direct detection of scattering delay is the ambiguity between the delay and increase of the signal travel time if the scattering object is placed at a larger distance from the radar antenna. This effect is called the range-delay ambiguity. It has been studied in [9, 12–14]. Under certain conditions, the range-delay ambiguity can be resolved by illuminating the object from various directions, similarly to how the along-track resolution is achieved in synthetic aperture radar (SAR) imaging [6, 11].

This work extends the results of [21] where we used deep learning to discriminate between the simulated coordinate-delay SAR (cdSAR) images of instantaneous and delayed scatterers. Whereas only one spatial coordinate, the range, has been considered in [21], in the current work we analyze the full-fledged 3D cdSAR images (two spatial coordinates and delay). Our goal is to identify and assess the potential benefits of using multiple cross-range coordinates for the robustness and quality of image classification.

As before, we use the simulated radar targets. A scenario was presented in [30] where the training on such targets did not generalize well and the corresponding algorithms demonstrated poor performance when applied to real-life data. Yet the advantage of simulated targets is that, unlike in the real world,[1] large datasets of targets and images with prescribed statistical characteristics can be generated, and the performance of the classifiers can be evaluated accurately. We consider two configurations, with and without the scattering delay, and with a tunable parameter that controls our ability to discriminate between the two. Building such configurations in real-life settings presents an extremely difficult task. Hence, our setup can be thought of as a testbed for studying the effect of different parameters, such as the grid (step) size and domain size, on the discrimination. Accordingly, we focus not on the specific percentages of successful or unsuccessful classifications, but on how those percentages vary as the corresponding simulation parameters change.

1.2. Motivation, research questions, and the roadmap.

Our main goal is to increase the robustness of the classifiers. In this work, robustness is interpreted as the maximum offset in the cross-range coordinate of the scatterer that does not significantly affect the system performance. Robustness with respect to the scatterer coordinates is important because in remote sensing, these coordinates are unknown, and the accuracy of their determination from the image is limited by the resolution, contrast, properties of the background, and other factors. If the cross-range coordinate of a 2D slice of the cdSAR image used for discrimination differs from the actual coordinate of the scatterer, then the classification performance decreases compared to that reported in [21]. To alleviate this deficiency, we propose a new 3D classifier that takes into account an interval of cross-range coordinates

[1] At the moment, the availability of real SAR images for algorithm training is indeed limited compared to conventional optical images (photographs). The situation may be improving though as additional companies produce large collections of SAR images, see, e.g., [31].

rather than a single value as in the 2D case. From the standpoint of convenience, the larger this interval the better. However, if the image coordinates are too far from the scatterer, the influence of the latter on the image is weak, and an increase in the amount of such inessential data may be detrimental. In our study we explore how large the cross-range interval can be so that the discrimination performance is not affected negatively.

Our second goal is to analyze the effect of including the additional cross-range coordinates on classification performance. For a given range-delay plane that contains the scatterer, the image at the neighboring cross-range locations is affected by the scatterer and therefore carries some information about the scatterer. This information can be used for classification. Yet any variation in the dimension of the input causes a change of architecture of the corresponding convolutional neural network (CNN). The difference in architecture can affect the performance of the CNN-based classifiers, so it is important to verify that different architectures produce similar results for comparable inputs.

SAR imaging is an established technology. An engineering perspective of the discipline can be found, e.g., in [8, 10, 19, 22, 23], while a mathematical approach to SAR imaging is presented in [5, 6, 11]. Dispersive targets have been of interest to SAR community for several decades, see [1, 3, 4, 9, 12–14, 20, 21, 24, 25]. Machine learning, on the other hand, is currently going through a phase of explosive growth. Therefore, we do not attempt to present a bibliography review and only mention two references that can provide some fundamental knowledge on the subject, see [15, 18]. The initial focus of deep learning was on optical images. Currently, there are more and more publications where the machine learning approach is applied to SAR, see, e.g., [33] and [32].

In what follows, Section 2 presents the mathematical fundamentals of the coordinate-delay synthetic aperture radar imaging, Section 3 describes the numerical experiments, and Section 4 discusses the results and identifies possible extensions of the current work.

2. Mathematical formulation of the problem.

2.1. Radar imaging operator and imaging kernel.

SAR images are obtained by illuminating the target from multiple antenna positions and processing the scattered signals received by the same antenna, see Figure 1. In this study, the target is characterized by the time-dependent target reflectivity function $\nu(t_{\boldsymbol{z}}, \boldsymbol{z})$, where $\boldsymbol{z} = (z_1, z_2, 0)$ is the target coordinate and $t_{\boldsymbol{z}} > 0$ is the scattering delay. The image is supposed to (approximately) reconstruct ν. We do not consider polarization of radar signals, i.e., adopt a scalar formulation. The propagation between the antenna and the target is assumed unobstructed while the scattering at the target is assumed weak (linearized). Then, the scattered signal $u^{(\mathrm{s})}(t, \boldsymbol{x})$ due to the emitted waveform $P(t)$ at the antenna position $\boldsymbol{x}$ is given by the following convolution-type expression:

$$u^{(\mathrm{s})}(t, \boldsymbol{x}) = \int_0^\infty \left(\int \nu(t_{\boldsymbol{z}}, \boldsymbol{z}) P\Big(t - \frac{2|\boldsymbol{x} - \boldsymbol{z}|}{c} - t_{\boldsymbol{z}}\Big)\, d\boldsymbol{z} \right) dt_{\boldsymbol{z}}, \tag{1}$$

where c is the speed of light. In formula (1), the antenna radiation pattern and geometrical attenuation are included into ν, while the integral without limits implies integration over the entire space. The retarded contribution to the argument of P is the sum of the round-trip travel time $2|\boldsymbol{x} - \boldsymbol{z}|/c$ and scattering delay $t_{\boldsymbol{z}}$. The

separation between these two terms in the sum is the main challenge for the remote sensing of dispersive targets. For a single antenna position, such separation is impossible [12]. This effect is called the range-delay ambiguity.

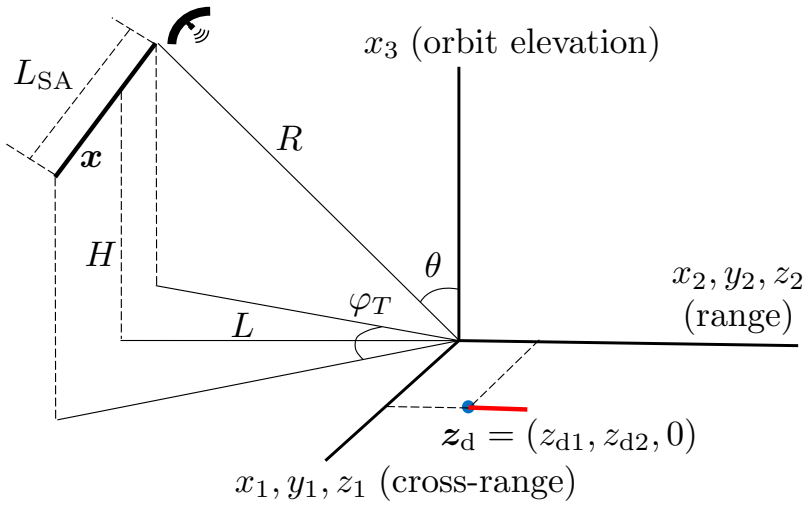

FIGURE 1. SAR imaging geometry. The targets (scatterers) are located on the plane $z_3 = 0$. The blue dot and red, straight segment correspond to the delayed and instantaneous scatterers, respectively. These two scatterers have a common reference point $\boldsymbol{z}_{\mathrm{d}}$.

The coordinate-delay SAR image $I(t_{\boldsymbol{y}}, \boldsymbol{y})$ is obtained by applying a matched filter to the received signal at each point of the data collection domain $\{\boldsymbol{x}\}$ called the synthetic aperture:

$$I(t_{\boldsymbol{y}}, \boldsymbol{y}) = \frac{1}{\mathcal{K}} \iint \underbrace{\overline{P\Big(t - \frac{2|\boldsymbol{x}-\boldsymbol{y}|}{c} - t_{\boldsymbol{y}}\Big)}}_{\text{matched filter}} u^{(\mathrm{s})}(t, \boldsymbol{x})\, dt\, d\boldsymbol{x}, \tag{2}$$

where the overbar denotes complex conjugate. Hereafter, we will use $\boldsymbol{y}$ to denote the image coordinates (while $\boldsymbol{z}$ will still be used for target coordinates). In equation (2), $\mathcal{K}$ is a normalization factor and the integration limits are determined by $\boldsymbol{y}$, pulse duration, and length of the synthetic aperture.

A typical SAR interrogating signal is a linear frequency-modulated pulse (narrowband):

$$P(t) = \exp(-i\omega_0 t) \exp\big(-iBt^2/(2\tau)\big) \mathbf{1}_{[-\tau/2,\, \tau/2]}, \quad \text{where} \quad 1 \ll B\tau \ll \omega_0 \tau. \tag{3}$$

In (3) and further on, $\mathbf{1}$ denotes the characteristic function (indicator) of an interval, ω_0 is the (circular) carrier frequency, and $B/(2\pi)$ and τ are the pulse bandwidth and duration, respectively. Another common assumption in SAR is that of a narrow total aperture angle:

$$\varphi_T \equiv \frac{L_{\mathrm{SA}}}{R} = \frac{L_{\mathrm{SA}}}{\sqrt{L^2 + H^2}} \ll 1, \tag{4}$$

where $\boldsymbol{x} = (x_1, -L, H)$ and $x_1 \in \mathbf{1}_{[-L_{\mathrm{SA}}/2,\, L_{\mathrm{SA}}/2]}$. In (4), L_{SA} is the length of the synthetic aperture, L is the horizontal distance to the target, and H is the orbit elevation, see Figure 1.

The imaging operator that maps the target reflectivity $\nu(t_{\boldsymbol{z}}, \boldsymbol{z})$ onto the cdSAR image $I(t_{\boldsymbol{y}}, \boldsymbol{y})$ has the following form:

$$I(t_{\boldsymbol{y}}, \boldsymbol{y}) = \iint W(t_{\boldsymbol{y}}, \boldsymbol{y}, t_{\boldsymbol{z}}, \boldsymbol{z}) \nu(t_{\boldsymbol{z}}, \boldsymbol{z})\, dt_{\boldsymbol{z}}\, d\boldsymbol{z}, \tag{5}$$

where the expression for the kernel W is obtained by substituting (1), (3), and (4) into (2) (see also [12]) and choosing the normalization factor as $\mathcal{K} = L_{\mathrm{SA}}\tau$:

$$W(t_{\boldsymbol{y}}, \boldsymbol{y}, t_{\boldsymbol{z}}, \boldsymbol{z}) \equiv W(\eta_{\boldsymbol{yz}}, \zeta_{\boldsymbol{yz}}, \psi_{\boldsymbol{yz}}) = \exp(-2i\beta\zeta_{\boldsymbol{yz}}) \cdot \Phi\big(\eta_{\boldsymbol{yz}}, \kappa(\zeta_{\boldsymbol{yz}} + \psi_{\boldsymbol{yz}})/2\big) \cdot \operatorname{sinc} \zeta_{\boldsymbol{yz}}. \tag{6}$$

In formula (6), $\beta = \omega_0/B \gg 1$, $\operatorname{sinc}\zeta = \sin(\zeta)/\zeta$, and

$$\kappa = \varphi_T^2 \omega_0 / B. \tag{7}$$

The quadratic dependence of κ on φ_T is due to the third term retained in the Taylor's expansion of the signal travel distances with respect to the aperture angle, see Figure 3 and [12, Section 2.2] for detail. In Section 2.2, we will see that it is the parameter κ of (7) that controls the system's ability to resolve the range-delay ambiguity.

The function $\Phi(v_1, v_2)$ in (6) is defined as

$$\Phi(v_1, v_2) \stackrel{\text{def}}{=} \int_{-1/2}^{1/2} \exp(2iv_1 s) \exp(iv_2 s^2)\, ds. \tag{8}$$

The arguments of the imaging kernel W of (6) are given by

$$\eta_{\boldsymbol{yz}} = \eta_{\boldsymbol{y}} - \eta_{\boldsymbol{z}}, \quad \zeta_{\boldsymbol{yz}} = \zeta_{\boldsymbol{y}} - \zeta_{\boldsymbol{z}}, \quad \psi_{\boldsymbol{yz}} = \psi_{\boldsymbol{y}} - \psi_{\boldsymbol{z}}, \tag{9}$$

where the dimensionless coordinates $\eta_{\boldsymbol{y}}$, $\eta_{\boldsymbol{z}}$, $\zeta_{\boldsymbol{y}}$, $\zeta_{\boldsymbol{z}}$, $\psi_{\boldsymbol{y}}$, and $\psi_{\boldsymbol{z}}$ are defined using the incidence angle θ, see Figure 1, and horizontal wavenumber $k_{\parallel} = (\omega_0/c)\sin\theta$:

$$\eta_{\boldsymbol{y}} = k_{\parallel}\varphi_T y_1, \qquad \eta_{\boldsymbol{z}} = k_{\parallel}\varphi_T z_1, \tag{10a}$$

$$\zeta_{\boldsymbol{y}} = \frac{B}{2}\Big(\frac{2y_2 \sin\theta}{c} + t_{\boldsymbol{y}}\Big), \qquad \zeta_{\boldsymbol{z}} = \frac{B}{2}\Big(\frac{2z_2 \sin\theta}{c} + t_{\boldsymbol{z}}\Big), \tag{10b}$$

$$\psi_{\boldsymbol{y}} = \frac{B}{2}\Big(\frac{2y_2 \sin\theta}{c} - t_{\boldsymbol{y}}\Big), \qquad \psi_{\boldsymbol{z}} = \frac{B}{2}\Big(\frac{2z_2 \sin\theta}{c} - t_{\boldsymbol{z}}\Big). \tag{10c}$$

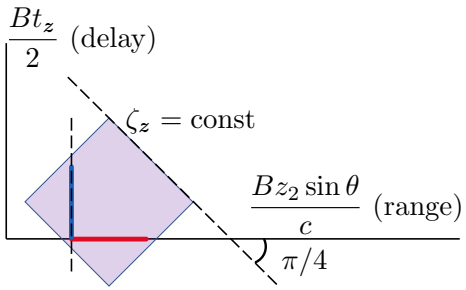

FIGURE 2. The range-delay plane that contains an instantaneous scatterer (red, straight segment) and a delayed scatterer (blue, straight segment) with the same dimensionless reference range $Bz_{\mathrm{d2}} \sin\theta/c$ (vertical dashed line). The purple square is the region where the data are sourced from for building the 2D classifiers in [21].

To reduce the cdSAR formulation to traditional SAR, one should replace $\Phi(v_1, v_2)$ with $\Phi(v_1, 0) \equiv \operatorname{sinc} v_1$ in (6) and eliminate the arguments $t_{\boldsymbol{y}}$ and $t_{\boldsymbol{z}}$ everywhere. For example, a regular (i.e., instantaneous) scatterer with reflectivity $\nu_{\text{inst}}(\boldsymbol{z})$ can be interpreted as $\nu(t_{\boldsymbol{z}}, \boldsymbol{z}) = \nu_{\text{inst}}(\boldsymbol{z})\delta(t_{\boldsymbol{z}})$. Formula (5) then reduces to

$$I_{\text{inst}}(\boldsymbol{y}) = \int W_{\text{inst}}(\boldsymbol{y}, \boldsymbol{z})\nu_{\text{inst}}(\boldsymbol{z})\, d\boldsymbol{z}. \tag{11}$$

The instantaneous imaging kernel W_{inst} of (11) is characterized by the range and cross-range resolution given by (see, e.g., [11, Chapter 2])

$$\Delta_{\text{R}} = \frac{\pi c}{B \sin\theta} \equiv \pi \frac{\omega_0}{B} \cdot \frac{1}{k_{\|}} \quad \text{and} \quad \Delta_{\text{A}} = \frac{\pi R c}{\omega_0 L_{\text{SA}}} \equiv \pi \frac{1}{k_{\|}\varphi_T}. \tag{12}$$

As there is no delay argument in either the reflectivity or the image, the range-delay ambiguity is not present in the traditional SAR formulation (11).

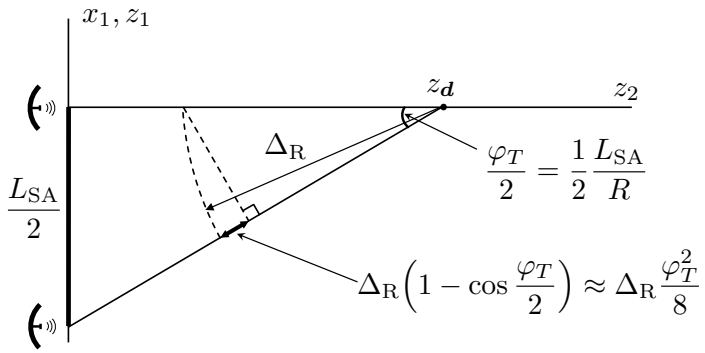

FIGURE 3. Quadratic dependence of the path length difference on the aperture angle.

2.2. **Range-delay ambiguity.** A SAR image should render an approximate reconstruction of the target reflectivity. This applies to both traditional SAR and cdSAR. In traditional SAR imaging (11), the best kernel would be $W_{\text{inst}} \sim \delta(\boldsymbol{y} - \boldsymbol{z})$, where δ is the Dirac delta function. Indeed, in this case we would have $I_{\text{inst}}(\boldsymbol{y}) \propto \nu_{\text{inst}}(\boldsymbol{z})$. However, the limitations due to the finite bandwidth and aperture size of a SAR sensor result in a finite width of the peak of the function W_{inst} around $\boldsymbol{y} = \boldsymbol{z}$. This finite width gives rise to a finite resolution (12).

Likewise, the best cdSAR imaging kernel in (5) would be $W \sim \delta(t_{\boldsymbol{y}} - t_{\boldsymbol{z}})\delta(\boldsymbol{y} - \boldsymbol{z})$. Yet the actual function W has a finite-width peak around $(t_{\boldsymbol{y}} = t_{\boldsymbol{z}}, \boldsymbol{y} = \boldsymbol{z})$. The cdSAR kernel W of (6) is different from W_{inst} of (11) in that $|W|$ decreases slowly along certain directions in the space of its arguments. Specifically, whereas $|\operatorname{sinc}\zeta| \le 1/|\zeta|$ as $|\zeta| \to \infty$, we have $|\Phi(v_1, v_2)| \sim 1/\sqrt{|v_2|}$ for $|v_2| \to \infty$ inside the sector $|v_1| \le |v_2|/2$. We can say that the range-delay ambiguity manifests itself as invariance of the imaging kernel w.r.t. range and delay in the limit of $\kappa \to 0$ in every plane given by equation[2]

$$\zeta_{\boldsymbol{z}} = \text{const}. \tag{13a}$$

[2]Note that the two terms in the definition of $\zeta_{\boldsymbol{z}}$ in (10b) can be traced back to the second and third terms in the argument of $P(\ldots)$ in (1).

Accordingly, in the three-dimensional space of coordinates $(z_1, z_2, t_{\boldsymbol{z}})$, we will call the planes (13a) ambiguity planes. An equivalent formulation in terms of the cdSAR image coordinates $(y_1, y_2, t_{\boldsymbol{y}})$ is

$$\zeta_{\boldsymbol{y}} = \text{const}, \tag{13b}$$

see formulae (9). It will be useful in the subsequent analysis. In Figure 2, the intersection of an ambiguity plane with a range-delay plane is shown by the sloped dashed line.

For finite κ, the function $\Phi(v_1, v_2)$ given by (8) has a central peak with respect to its second argument at $v_2 = 0$. The width of this peak corresponds to the value of φ_T where for a pair points separated in range by the resolution Δ_{R} of (12), the path difference $\Delta_{\mathrm{R}}\big(1 - \cos(\varphi_T/2)\big) \approx \Delta_{\mathrm{R}}\varphi_T^2/8$ is comparable to the carrier wavelength $\lambda = 2\pi c/\omega_0$, see Figure 3. This peak is what allows us to resolve the range-delay ambiguity, i.e., discriminate between the instantaneous and delayed scattering. Indeed, we can rewrite the second argument of Φ in (6) using (9) and (10) as follows:

$$\kappa(\zeta_{\boldsymbol{yz}} + \psi_{\boldsymbol{yz}})/2 = \kappa\frac{B\sin\theta}{c}(y_2 - z_2). \tag{14}$$

As the right-hand side of (14) depends only the spatial coordinates, it is not subject to the range-delay ambiguity, unlike other terms in (6). Moreover, the resulting range resolution appears inversely proportional to κ. However, the aforementioned slow decay of $|\Phi(v_1, v_2)|$ for large $|v_2|$ implies that the sidelobes will still be prominent, see [12] for detail. Hence, resolving the range-delay ambiguity may not be an easy task even when the target is illuminated from a series of locations $\{\boldsymbol{x}\}$.

2.3. **Statistical models of scatterers.** Unlike in the conventional photography where illumination is incoherent, radar imaging uses coherent signals [e.g., as expressed by (3)] to probe the targets. For this reason, radar images of extended objects always have grainy appearance due to the phenomenon called speckle [16, 17, 27]. The interpretation of instantaneous speckle presented in [27, Chapter 4] uses stochastic models of scatterer reflectivity. In the current work, we extend the approach used in [12, 13, 21] by adding the cross-range dimension. In particular, we define one-dimensional inhomogeneous targets using an inhomogeneous complex Gaussian white noise process $\mu(\zeta)$ with the following autocorrelation function:

$$\langle\overline{\mu(\zeta')}\mu(\zeta)\rangle = \delta(\zeta' - \zeta)\mathbf{1}_{0\le\zeta\le\zeta_{\max}}, \tag{15}$$

where $\zeta_{\max}$ is a parameter. The support of the autocorrelation in (15), i.e., $\mathbf{1}_{0\le\zeta\le\zeta_{\max}}$, is an L_1-integrable function of ζ and hence satisfies a sufficient condition for μ to exist, see [2].

Using (15), we define the reflectivity functions for two types of inhomogeneous scatterers:

$$\nu_s(t_{\boldsymbol{z}}, \boldsymbol{z}) = \sigma_s\delta\big(k_\parallel\varphi_T(z_1 - z_{\mathrm{d}1})\big)\delta\Big(\frac{Bt_{\boldsymbol{z}}}{2}\Big)\mu\Big(\frac{B}{\omega_0}k_\parallel(z_2 - z_{\mathrm{d}2})\Big), \tag{16}$$

$$\nu_t(t_{\boldsymbol{z}}, \boldsymbol{z}) = \sigma_t\delta\big(k_\parallel\varphi_T(z_1 - z_{\mathrm{d}1})\big)\delta\Big(\frac{B}{\omega_0}k_\parallel(z_2 - z_{\mathrm{d}2})\Big)\mu\Big(\frac{Bt_{\boldsymbol{z}}}{2}\Big). \tag{17}$$

An instantaneous scatterer defined by formula (16) is localized in the cross-range coordinate, i.e., $z_1 = z_{\mathrm{d}1}$. Regarding the range coordinate, this scatterer is characterized by the reference location $\boldsymbol{z}_{\mathrm{d}}$ and length $s_{\max} = \omega_0\zeta_{\max}/(Bk_\parallel)$. In turn, formula (17) defines a delayed scatterer with the maximum delay $t_{\max} = 2\zeta_{\max}/B$

located at $\boldsymbol{z}_{\mathrm{d}}$, i.e., $z_1 = z_{\mathrm{d}1}$ and $z_2 = z_{\mathrm{d}2}$. The intensities of the scatterers are controlled by the constants $\sigma_s > 0$ and $\sigma_t > 0$. The supports of $\langle|\nu_s|^2\rangle$ and $\langle|\nu_t|^2\rangle$ are shown in Figures 1 and 2. Note that if the random process μ in the definitions (16) and (17) is the same, these supports intersect one and the same set of ambiguity planes (13a).

In addition to the inhomogeneous models (16) and (17), we will use an instantaneous speckle as a background (sometimes called clutter). The model of the speckle reflectivity function $\nu_{\mathrm{speckle}}(\boldsymbol{z})$ that is used in the conventional SAR imaging is a homogeneous two-dimensional white Gaussian noise [27]:

$$\big\langle\overline{\nu_{\mathrm{speckle}}(\boldsymbol{z}')}\nu_{\mathrm{speckle}}(\boldsymbol{z})\big\rangle = \sigma_{\mathrm{b}}^2\delta(\boldsymbol{z}'-\boldsymbol{z}), \tag{18}$$

where σ_{b}^2 is the average speckle intensity. We adapt this model for the cdSAR treatment by explicitly specifying the dependence on the delay time $t_{\boldsymbol{z}}$ as follows:

$$\nu_{\mathrm{b}}(t_{\boldsymbol{z}},\boldsymbol{z}) = \delta\Big(\frac{Bt_{\boldsymbol{z}}}{2}\Big)\nu_{\mathrm{speckle}}(\boldsymbol{z}). \tag{19}$$

Formulations (15)–(19) permit the following unifying expression for the autocorrelation of the scatterer reflectivity:

$$\big\langle\overline{\nu_\alpha(t'_{\boldsymbol{z}},\boldsymbol{z}')}\nu_\alpha(t_{\boldsymbol{z}},\boldsymbol{z})\big\rangle = \sigma_\alpha^2\mathcal{R}_\alpha(\tilde{t}_{\boldsymbol{z}},\tilde{z}_1,\tilde{z}_2;t_{\boldsymbol{z}},z_1,z_2), \tag{20}$$

where α is a scatterer type: $\alpha\in\{\mathrm{b},s,t\}$, $\tilde{z}_1 = z'_1 - z_1$, $\tilde{z}_2 = z'_2 - z_2$, $\tilde{t}_{\boldsymbol{z}} = t'_{\boldsymbol{z}} - t_{\boldsymbol{z}}$, and

$$\mathcal{R}_\alpha(\tilde{t}_{\boldsymbol{z}},\tilde{z}_1,\tilde{z}_2;t_{\boldsymbol{z}},z_1,z_2) = \delta\big(k_\parallel\varphi_T\tilde{z}_1\big)\delta\Big(\frac{B}{\omega_0}k_\parallel\tilde{z}_2\Big)\delta\Big(\frac{B\tilde{t}_{\boldsymbol{z}}}{2}\Big)a_\alpha(t_{\boldsymbol{z}},z_1,z_2). \tag{21}$$

In turn, the factors $a_\alpha(t_{\boldsymbol{z}},z_1,z_2)$ define the spatio-temporal localization of the scatterers:

$$a_{\mathrm{b}}(t_{\boldsymbol{z}},z_1,z_2) = \delta\Big(\frac{Bt_{\boldsymbol{z}}}{2}\Big), \tag{22a}$$

$$a_s(t_{\boldsymbol{z}},z_1,z_2) = \delta\big(k_\parallel\varphi_T(z_1-z_{\mathrm{d}1})\big)\delta\Big(\frac{Bt_{\boldsymbol{z}}}{2}\Big)\mathbf{1}_{z_{\mathrm{d}2}\le z_2\le z_{\mathrm{d}2}+s_{\max}}, \tag{22b}$$

$$a_t(t_{\boldsymbol{z}},z_1,z_2) = \delta\big(k_\parallel\varphi_T(z_1-z_{\mathrm{d}1})\big)\delta\Big(\frac{B}{\omega_0}k_\parallel(z_2-z_{\mathrm{d}2})\Big)\mathbf{1}_{0\le t_{\boldsymbol{z}}\le t_{\max}}, \tag{22c}$$

see (16), (17), and (19). Note that the multiplication of the delta-functions in (21)–(22) is legitimate because each coordinate or delay argument on the left-hand sides appear as an argument of only one delta-function on the corresponding right-hand side.

2.4. **Statistics of cdSAR images.** In order to realize the scenarios with and without scattering delays, we combine the statistical models from Section 2.3 in two different ways:

$$\nu_{\mathrm{inst}}(t_{\boldsymbol{z}},\boldsymbol{z}) = \nu_{\mathrm{b}}(t_{\boldsymbol{z}},\boldsymbol{z}) + \nu_s(t_{\boldsymbol{z}},\boldsymbol{z}), \tag{23a}$$

$$\nu_{\mathrm{del}}(t_{\boldsymbol{z}},\boldsymbol{z}) = \nu_{\mathrm{b}}(t_{\boldsymbol{z}},\boldsymbol{z}) + \nu_s(t_{\boldsymbol{z}},\boldsymbol{z}). \tag{23b}$$

Using the expressions for the scatterer reflectivities in Section 2.3, we can calculate the covariance matrices of the resulting cdSAR images. Due to the linearity of scattering in (1), we have

$$I(t_{\boldsymbol{y}},\boldsymbol{y}) \equiv I(\eta_{\boldsymbol{y}},\zeta_{\boldsymbol{y}},\psi_{\boldsymbol{y}}) = \sum_\alpha I_\alpha(\eta_{\boldsymbol{y}},\zeta_{\boldsymbol{y}},\psi_{\boldsymbol{y}}), \tag{24}$$

where $\alpha \in \{\mathrm{b}, s, t\}$ is the scatterer type, and

$$I_\alpha(\eta_{\boldsymbol{y}}, \zeta_{\boldsymbol{y}}, \psi_{\boldsymbol{y}}) = \iiint W(\eta_{\boldsymbol{yz}}, \zeta_{\boldsymbol{yz}}, \psi_{\boldsymbol{yz}}) \nu_\alpha(\eta_{\boldsymbol{z}}, \zeta_{\boldsymbol{z}}, \psi_{\boldsymbol{z}})\, d\eta_{\boldsymbol{z}}\, d\zeta_{\boldsymbol{z}}\, d\psi_{\boldsymbol{z}}, \tag{25}$$

see (5), where the coordinate transform $(t_{\boldsymbol{z}}, \boldsymbol{z}) \to (\eta_{\boldsymbol{z}}, \zeta_{\boldsymbol{z}}, \psi_{\boldsymbol{z}})$ is implied. Assuming that the two terms on the right-hand sides of both (23a) and (23b) are statistically independent and using (η, ζ, ψ) instead of $(\eta_{\boldsymbol{y}}, \zeta_{\boldsymbol{y}}, \psi_{\boldsymbol{y}})$ for the sake of brevity, we obtain

$$\big\langle \overline{I(\eta', \zeta', \psi')} I(\eta, \zeta, \psi) \big\rangle = \sum_\alpha \big\langle \overline{I_\alpha(\eta', \zeta', \psi')} I_\alpha(\eta, \zeta, \psi) \big\rangle. \tag{26}$$

Using (6), (20), (21), and (25), we express the expectations on the right-hand side of (26) as follows:

$$\begin{aligned} \big\langle \overline{I_\alpha(\eta', \zeta', \psi')} I_\alpha(\eta, \zeta, \psi) \big\rangle \sim \sigma_\alpha^2 \iiint & \overline{\Phi\big(\eta' - \eta_{\boldsymbol{z}}, \kappa(\zeta' - \psi' - \zeta_{\boldsymbol{z}} + \psi_{\boldsymbol{z}})\big)}\, \mathrm{sinc}(\zeta' - \zeta_{\boldsymbol{z}}) \\ & \cdot \Phi\big(\eta - \eta_{\boldsymbol{z}}, \kappa(\zeta - \psi - \zeta_{\boldsymbol{z}} + \psi_{\boldsymbol{z}})\big)\, \mathrm{sinc}(\zeta - \zeta_{\boldsymbol{z}}) \\ & \cdot \exp\big(-2i\beta(\zeta - \zeta')\big) a_\alpha(\eta_{\boldsymbol{z}}, \zeta_{\boldsymbol{z}}, \psi_{\boldsymbol{z}})\, d\eta_{\boldsymbol{z}}\, d\zeta_{\boldsymbol{z}}\, d\psi_{\boldsymbol{z}}. \end{aligned} \tag{27}$$

Notice that the absolute value of the covariance given by (27) quickly decreases as $|\zeta - \zeta'|$ increases due to the sinc factors in the integrand. This allows us to simplify calculations of cdSAR image covariance by restricting formula (27) to a single ambiguity plane, i.e., considering only $\zeta' = \zeta$. For the same reason, we take $\zeta_{\boldsymbol{z}} = \zeta$ and drop integration over $\zeta_{\boldsymbol{z}}$, arriving at

$$\begin{aligned} \big\langle \overline{I_\alpha(\eta', \zeta, \psi')} I_\alpha(\eta, \zeta, \psi) \big\rangle \sim \sigma_\alpha^2 \iint & \overline{\Phi\big(\eta' - \eta_{\boldsymbol{z}}, \kappa(-\psi' + \psi_{\boldsymbol{z}})\big)} \Phi\big(\eta - \eta_{\boldsymbol{z}}, \kappa(-\psi + \psi_{\boldsymbol{z}})\big) \\ & \cdot a_\alpha(\eta_{\boldsymbol{z}}, \zeta, \psi_{\boldsymbol{z}})\, d\eta_{\boldsymbol{z}}\, d\psi_{\boldsymbol{z}}. \end{aligned} \tag{28}$$

At the same time, due to the same factor $\mathrm{sinc}\,\zeta$ in the kernel (6), we assume that the image values corresponding to different values of ζ are nearly independent provided that the grid size in the coordinate ζ is at least π. Computation of the integral in (28) for each $\alpha \in \{\mathrm{b}, s, t\}$ is simplified further by the delta functions in (22).

2.5. **The discrimination problem.** Expression (28) defines the covariance for the circular Gaussian variables I_α. When $\kappa > 0$, the covariance functions for I_s and I_t given by (28) are different, as illustrated in Figure 4. This difference is controlled by two parameters, κ and $\zeta_{\max}$, defined in (7) and (15), respectively. In particular, according to (6) and (8), an increase of κ reduces the widths of the peaks of the blue and red curves in Figure 4 and thus improves separation between them. In turn, the value of $\zeta_{\max}$ represents the distance between the upper ends of the red and blue "sticks" in this figure, and hence, an increase of $\zeta_{\max}$ also increases the separation between the two covariances. Since the component I_{b} in the images resulting from the two scenarios in (23) is the same, the discrimination between these scenarios should improve as the product $\kappa\zeta_{\max}$ increases. This has been confirmed in [12] where a discrimination procedure was built based on the maximum likelihood principle. In the current work, we associate this discrimination with detection of the scattering delay with the help of radar imaging (see also our earlier publications [12–14, 21]).

Another factor that affects the discrimination between ν_{inst} and ν_{del} in (23) is the contrast. The numerical value of the contrast characterizes the share of intensity of the inhomogeneous scatterer, $\langle |I_s|^2 \rangle$ or $\langle |I_t|^2 \rangle$, in the total image intensity $\langle |I|^2 \rangle$.

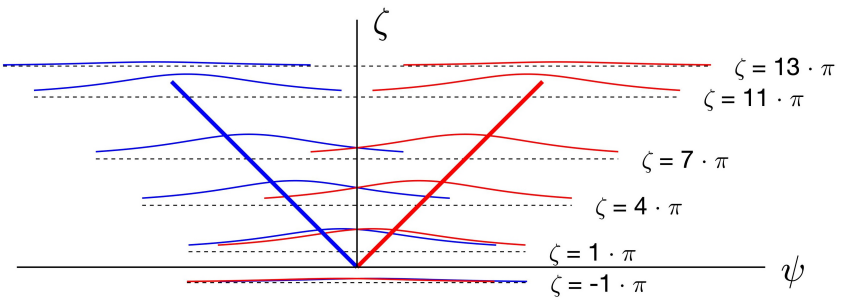

FIGURE 4. Schematic: separation of statistical characteristics of the instantaneous (the red stick) and delayed (the blue stick) scatterers in the range-delay plane containing the scatterers, see also Figure 2. The red and blue curves represent $\langle |I_s|^2(0,\zeta,\psi)\rangle$ and $\langle |I_t|^2(0,\zeta,\psi)\rangle$ given by (25), (28) as functions of ψ for a fixed ζ. For the case of $z_{\mathrm{d1}} = z_{\mathrm{d2}} = 0$, the maxima of these curves are located near $\psi = \zeta$ and $\psi = -\zeta$, respectively. However, this behavior is barely noticeable beyond $0 \le \zeta \le \zeta_{\max}$, see (15); in this illustration, $\zeta_{\max} = 12\pi$.

This share is defined by the relations between σ_s^2, σ_t^2, and σ_{b}^2 in (16), (17), and (19), respectively, and also the level of signal noise, see Appendix A for more details. In particular, the case of zero contrast means that the scatterer reflectivity is zero, hence the images due to ν_{inst} and ν_{del} are identical, whereas the value of contrast equal to one means that the entire signal is due to the inhomogeneous scatterer. Obviously, as the value of the contrast decreases, the maxima of the red and blue curves in Figure 4 become less prominent, thus making the discrimination between the two scattering scenarios harder.

As another illustration, Figures 5 and 6 display the average (top rows) and individual amplitude (bottom rows) of the sampled cdSAR images, with the discretization grid in the range-delay plane corresponding to the purple square in Figure 2, see also [21, Figures 2 and 3]. Figures 5 and 6 correspond to two different values of the cross-range coordinate: $\eta = k_\parallel \varphi_T z_{\mathrm{d1}}$ and $\eta = k_\parallel \varphi_T z_{\mathrm{d1}} + \pi$, respectively (according to (12), the cross-range distance between these two locations is Δ_{A}). We can see that the top plots in Figure 5 are mostly in agreement with the schematic in Figure 4. At the same time, due to the stochasticity, the bottom plots bear little or no resemblance to the corresponding top plots in the same figure (neither to Figure 4). The patterns in the top plots in Figure 6 are due to the cross-range sidelobes of the imaging kernel (6). As indicated by the color scales, these patterns have much smaller variations of intensity as compared to those in Figure 5; still, there is a clear difference between the two types of targets. At the same time, it appears very hard to notice the corresponding patterns in the bottom plots of the same figure.

From the pattern recognition perspective, discriminating between the averages (the images in the top rows in Figures 5 and 6) should not be too hard. Unfortunately, only the individual images (represented in the bottom rows of these figures) are available in the typical real world situations. It is then not much of a surprise

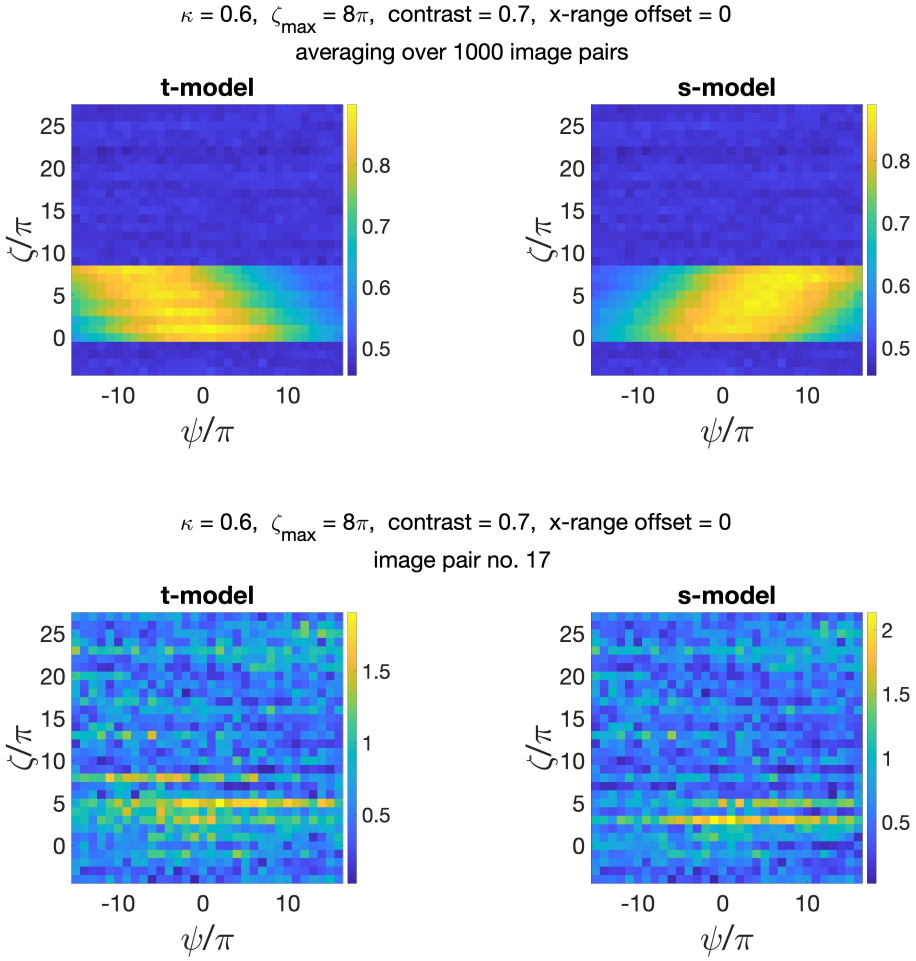

FIGURE 5. Magnitude of cdSAR images in the range-delay plane containing the inhomogeneous scatterers. Top plots: average over the dataset; bottom plots: individual images.

that the classification based on the individual images similar to those presented in these figures is subject to errors.

2.6. **Simulation of image ensembles.** The cdSAR images given by (24) can be described as multivariate zero-mean Gaussian variables. Expressions (26) with (28) provide the covariances, and hence, completely define the probability distribution functions (PDF) for these variables. Remember that when considered a function of the image values and values of model parameters, such PDF is called likelihood.

Technically speaking, knowledge of the PDF opens the way to calculating all metrics of the corresponding random variables. This includes the misclassification rate of the maximum likelihood based classifiers employed in [12–14]: its expectation could be calculated using the following procedure:

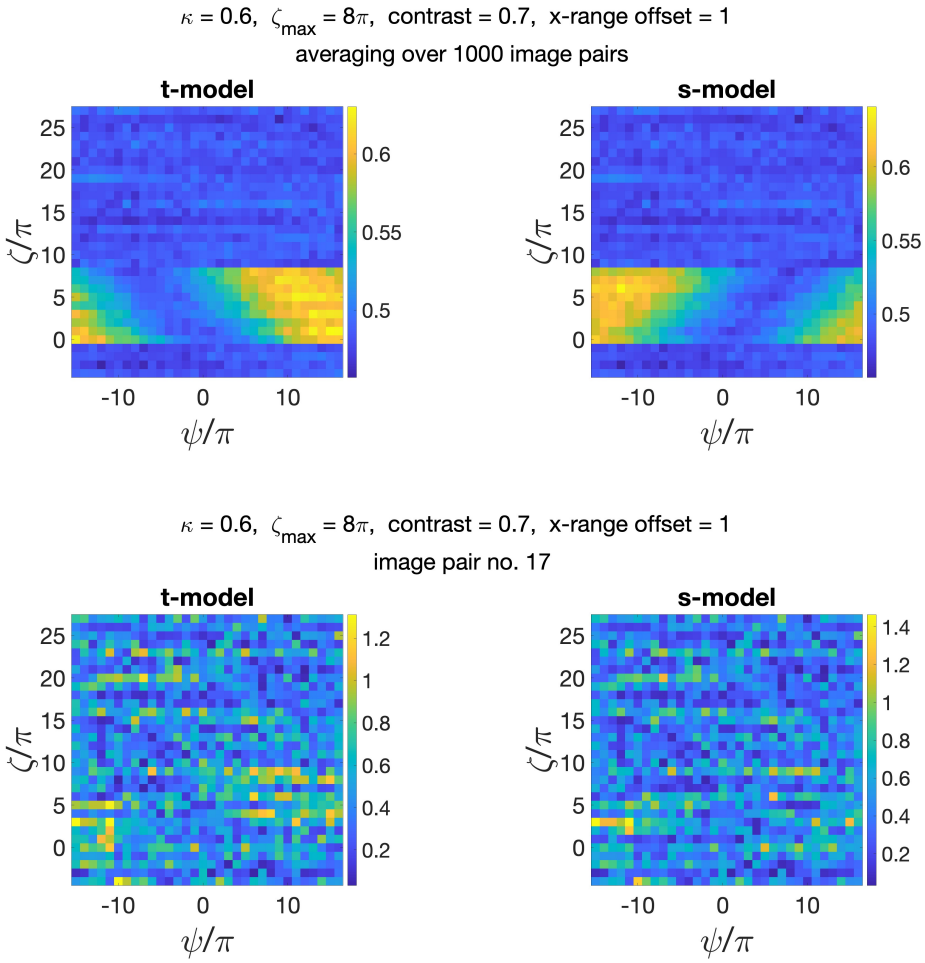

FIGURE 6. Same as in Figure 5, but in the range-delay plane shifted by π in the cross-range direction.

1. Choose one of the two scenarios in (23), specify the values of the relevant parameters, such as κ, $\zeta_{\max}$, and the value of the contrast q, and choose d discretization points.
2. Consider a discretized realization (i.e., a $2d$-dimensional vector) $I = I_r + iI_i$ given by (24), where $I_r \in \mathbb{R}^d$, $I_i \in \mathbb{R}^d$. This image may be due to either scenario (23a) with $q_s = q$ or scenario (23b) with $q_t = q$, where q_s and q_t are the contrasts for the corresponding scenarios, see Appendix A. Note that these two scenarios also give rise to two likelihood functions (in fact, two PDFs): $p_s(I, q_s)$ and $p_t(I, q_t)$, regardless of what model was used to generate I.
3. For this realization I, calculate the maxima

$$p_{s,\max}(I) = \max_{q_s'} p_s(I, q_s'), \quad p_{t,\max}(I) = \max_{q_t'} p_t(I, q_t') \tag{29}$$

by varying q_s' and q_t', respectively.

4. Depending on the scenario in item 1, we define the classification outcome $C_s(I)$ or $C_t(I)$ as follows: if the larger of the resulting maxima in (29) comes from the same scenario, then the outcome is assigned the value $C_s(I) = 1$ or $C_t(I) = 1$ (i.e., the classification is correct), otherwise $C_s(I) = 0$ or $C_t(I) = 0$ (misclassification).
5. The expected value of the classification outcome in item 4 is calculated using the PDFs p_s and p_t specified in item 3:

$$\langle C_s(q)\rangle = \int C_s(I)p_s(I,q)\,dI_r\,dI_i, \quad \langle C_t(q)\rangle = \int C_t(I)p_t(I,q)\,dI_r\,dI_i, \tag{30}$$

where both dI_r and dI_i mean integration over $\mathbb{R}^d$, and $q_s = q_t = q$.
6. The values obtained in item 5 are averaged over the two scenarios in item 1 to yield the average misclassification rate $M_{\text{avg}}(q)$:

$$M_{\text{avg}}(q) = 1 - C_{\text{avg}}(q) = 1 - \frac{\langle C_s(q)\rangle + \langle C_t(q)\rangle}{2}. \tag{31}$$

Because of the stochastic nature of SAR imaging, see Section 2.4, and the difficulties associated with resolving the range-delay ambiguity presented in Section 2.2, the number of discretization nodes d has to be large. Hence, the calculation of expectation in item 5 requires integration of rather nontrivial and discontinuous functions $C_s(I)$ and $C_t(I)$ over a multidimensional domain, see (30). This makes the previous method prohibitively expensive, even in the 2D case. An alternative approach we used in [12–14] is to calculate the experimental misclassification rate over representative ensembles of cdSAR images using the same maximum likelihood principle as that specified in items 3 and 4. These ensembles are created by sampling the aforementioned multivariate Gaussian PDFs. Essentially, this is a Monte-Carlo procedure applied to the calculation of the expectations in item 5, and its implementation requires the generation of image ensembles according to a certain multivariate Gaussian statistics. In addition, these ensembles can provide training sets for the classifiers based on neural networks. This approach was introduced in [21] and continued in the current work. The modifications required for a transition from 2D to 3D cdSAR images are described in Appendix B.

3. Numerical experiments.

3.1. General principles.

We discretize the cdSAR images using a regularly spaced 3D grid with step sizes π and $\pi/2$ in the non-dimensional range $By_2 \sin\theta/c$ and delay $Bt_{\boldsymbol{y}}/2$, respectively, see Figure 7. The discretization allows us to extract image subsets similar to those used in [21], see Figure 5.[3] For the non-dimensional cross-range coordinate $\eta = k_{\|}\varphi_T y_1$, we use step size π in most cases, whereas smaller sizes (such as $\pi/2$, $\pi/4$, etc.) have also been used, e.g., in Sections 3.3 and 3.4. The envelope of a typical discretized domain is shown in Figure 7.

In both 2D and 3D cases, the classifiers are realized as feed-forward convolutional neural networks (CNNs). The training procedure in each case is the supervised learning with cross-entropy as the loss function [18]. The implementations employs the PyTorch software package [28].

The classifiers are trained on a dataset containing 10 values of the target contrast in the interval between 0.0 and 0.9. The dataset is split into three parts used for training, validation, and testing (or evaluation), and containing 70%, 15%, and

[3] This way, we can perform the contrast scaling procedure as described in Appendix A.

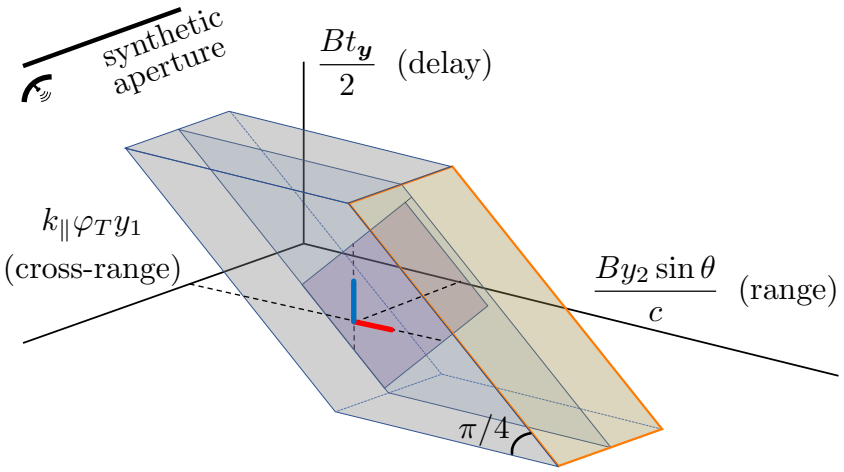

FIGURE 7. Coordinate-delay SAR imaging, cf. Figure 1. The envelopes of the 2D and 3D discretized domains are the purple square (also shown in Figure 2) and gray parallelepiped, respectively. The instantaneous scatterers and the purple square are in the plane of symmetry of the parallelepiped. The side of the parallelepiped that is aligned with the ambiguity planes $\zeta_{\boldsymbol{y}} = \text{const}$ is highlighted yellow.

15% of the images for each contrast, respectively. In training and validation, all target contrasts are combined, whereas in testing, we calculate the percentage of classification errors for each contrast separately. The average misclassification rate $M_{\text{avg}}(q)$, see (31), is plotted as a function of the contrast; such a graph is sometimes called the misclassification curve, see, e.g., Figure 8.

Note that in the process of training a deep learning classifier, its parameters (called weights and biases) are initialized randomly. Different initializations can lead to different final values of these parameters in the trained CNN, so the performance of the resulting classifiers may differ. In order to assess and monitor this effect, we routinely created ensembles of classifiers using the stochasticity of initializations and calculated the spread of the resulting misclassification rates. This spread, in terms of the standard deviation, is shown in some of the plots, see, e.g., Figures 8 and 11. When analyzing these sets of classifier instances, we have discovered that some ensembles of classifiers have a few outliers with much lower performance than the rest.[4] Such outliers are ignored when the mean and standard deviation of the misclassification rate are calculated, and we display the number of such outliers for each case in the plot legend. This procedure is legitimate because in our experiments, the test and validation performances are found to be very close, so the identification of outliers can be performed at the validation step.

[4] The outliers have been detected using the Matlab® `rmoutliers` routine using the default "median" method. The classifier is designated as an outlier if its performance is an outlier for at least two of the total 10 values of contrast.

3.2. **Performance of classifiers when the scatterer location is known.**

3.2.1. *2D classifiers with a cross-range offset.* In a set of experiments illustrated in Figure 8, we used 2D classifiers where the 2D slices of cdSAR images with a fixed offset in the cross-range coordinate have been taken as input, both for training/validation and testing. With the grid size in the non-dimensional coordinate $\eta \equiv \eta_{\boldsymbol{y}}$ [see (10)] equal to π, an index offset of n that is shown in the legend for each curve corresponds to a physical offset of $n\Delta_{\mathrm{A}}$, see (12).

We observe that the offset slices are not as good for the discrimination purposes as the one containing the scatterer. Figure 9 illustrates a similar experiment with a reduced grid size in the cross-range dimension; we see that the performance drops significantly as the absolute offset exceeds $\Delta_{\mathrm{A}}/4$, and as it reaches $\sim 1.5\Delta_{\mathrm{A}}$ (corresponding to $n = 6$), the 2D classifier becomes worthless. This is not very surprising because the azimuthal sidelobes of the imaging kernel are characterized by much lower contrast as compared to the central peak, see Figures 5 and 6, as well as [12, Figure 2].

On one hand, the offset value of $\Delta_{\mathrm{A}}/4$ mentioned above can serve as a measure of robustness of 2D classifiers. On the other hand, we can argue that a significant amount of "information" contained in the slices with the offset exceeding $\Delta_{\mathrm{A}}/4$ is due to the stochasticity of clutter and noise and, hence, irrelevant to the scatterer and its type, so learning it in other scenarios will not necessarily improve the discrimination quality.

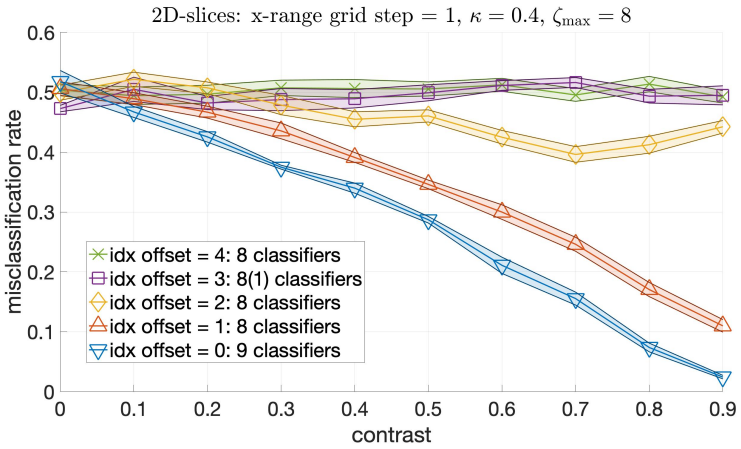

FIGURE 8. Misclassification curves (i.e., the dependence of the averaged misclassification rate on the target contrast) due to 2D classifiers with different cross-range offsets. Solid curves represent the average over the ensemble of classifiers, whereas the shaded regions around them illustrate the standard deviation. For each curve, the total number of classifiers, followed by the number of outliers (see Section 3.1), is listed. The grid step in the cross-range direction is Δ_{A}.

3.2.2. *3D classifiers with a cross-range offset.* In the following experiment, we explore whether the performance of the 3D classifiers is as sensitive to the shift of the

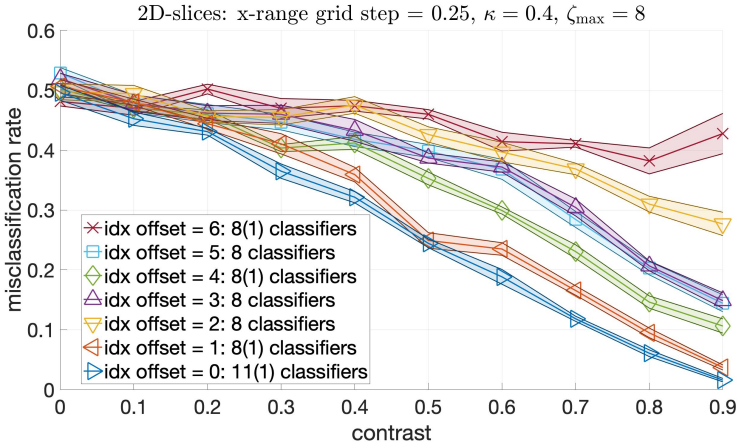

FIGURE 9. Same as in Figure 8, but with 4 times smaller grid step in the cross-range direction.

inhomogeneous scatterers in the cross-range direction as that of the 2D classifiers, provided that the magnitude of the shift is known.

With the projection of the discretization grid on the plane $\eta = 0$ fixed, we vary the number of discretization points in the cross-range direction. Each point corresponds to a 2D slice of the discretization grid. We call a collection of several adjacent slices of this kind a slab. Accordingly, a discretization grid with only one value of the cross-range coordinate (i.e., a 2D structure) is called a 1-slab. In this section, we also use 3-slabs and 7-slabs. Moreover, for each of the classifiers considered, we use two different grid sizes in the cross-range direction.

The resulting misclassification curves are presented in Figure 10. These plots show that as long as the scatterer is covered by the discretization grid, the performance of the 3D classifier with a moderate number of slabs is comparable to that of the centered 2D classifier. At the same time, once the scatterer is no longer within the discretization grid, we can observe that the classification quality quickly deteriorates, similarly to Figures 8 and 9.

We therefore conclude that the 3D classifiers offer an increased tolerance to inaccuracies in determining the scatterer position.

3.2.3. *Centered 3D classifiers.* A conjecture introduced in Section 3.2.1 about the fast decline of useful information as the cross-range offset increases pertains to 2D classifiers. Even if confirmed, it does not automatically apply to 3D classifiers that collect information from several cross-range offsets and may use the relations between them to infer additional information about the target. We address this topic in the next series of simulations where the cdSAR images have a total of 33 discretization points in the cross-range direction. We use 3-, 7-, and 11-slab classifiers with the cross-range coordinate of the inhomogeneous scatterer in the symmetry plane of the cdSAR discretization grid, see, e.g., Figures 10 and 11. In addition, we employ two 32-node discretization grids (and the associated classifiers) obtained by dropping one of the two outer cross-range coordinates: these cases are called "first-clipped" and "last-clipped," see Figure 11.

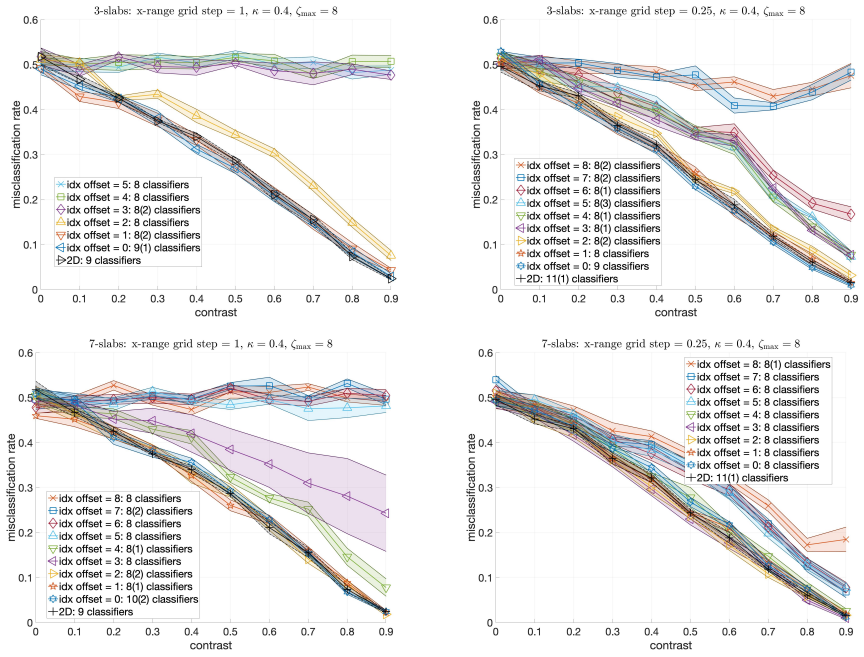

FIGURE 10. Misclassification curves for 3-slab (top row) and 7-slab (bottom row) classifiers applied to cdSAR images with cross-range grid sizes of $1\Delta_{\mathrm{A}}$ (left column) and $0.25\Delta_{\mathrm{A}}$ (right column). For both grid sizes, the scatterer is within the discretization grid as long as the index offset for the 3-slab and 7-slab does not exceed 1 and 3, respectively. For some reason that is still not fully understood, a larger than usual number of classifiers misbehave for offset=3 for the left bottom plot, so the procedure of outlier removal described in Footnote 4 on page 14 was not effective.

As the number of slabs increases, our simulations show no improvement in classification performance. In fact, the discrimination quality even slightly decreases, as can be seen in Figure 11. At the same time, the misclassification curves produced by the 2D and 3D classifiers appear qualitatively very similar (this also includes the cases in Figure 10 where the scatterer is covered by the discretization grid) despite the substantial differences in CNN architecture and format of the input data. This indicates that apparently, the 2D mechanism still plays a dominant role in classification with a 3D architecture.

There are several possible explanations to the foregoing observation. The first hypothesis is that the effect of "noise" information in the slabs far away from the scatterer, see Section 3.2.1, outweighs that of the additional "good" information contained in the same slabs. The second hypothesis is that the 3D classifier architecture is not as efficient as that in 2D. Yet the third possibility is that in the 2D case, extracting the correct slab (i.e., with the cross-range coordinate equal to that of the scatterer) is equivalent to providing a significant amount of information about the scatterer, while for the 3D classifier this is not so important as long as the scatterer is within the discretization grid. We further elaborate on these hypotheses

in Appendix C, but still do not have an unambiguous explanation of the observed effect. One obstacle is the lack of formal definition of information applicable to the phenomena we analyze.

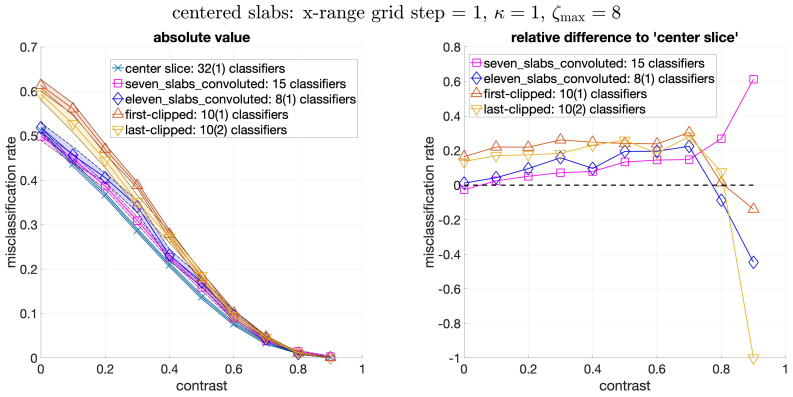

FIGURE 11. Left panel: Misclassification curves (see the caption to Figure 8) for different number of slices in the 3D discretization grids. Right panel: the relative difference between the 3D and 2D misclassification rates (the 2D classifier is labeled as "center slice" on the left panel). For this plot, the values for the two largest contrasts (i.e., 0.8 and 0.9) are not representative because the number of misclassifications is very low.

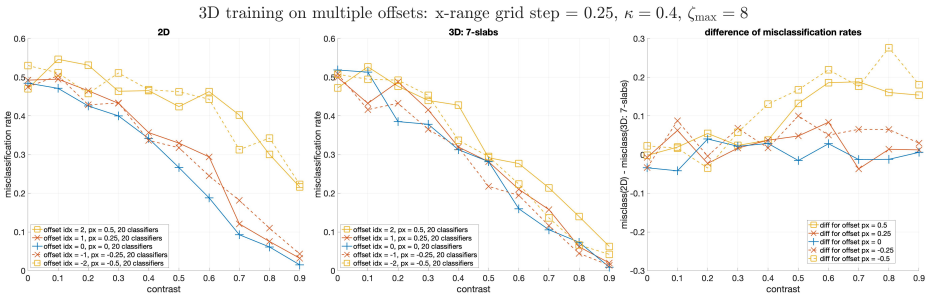

FIGURE 12. Averaged performance of classifiers trained on multiple cross-range offsets, see Section 3.3. The training offsets range from $-0.5\Delta_{\mathrm{A}}$ to $0.5\Delta_{\mathrm{A}}$ with a step of $0.25\Delta_{\mathrm{A}}$. Each curve corresponds to a test dataset having a single value of the cross-range offset. Left: 2D classifiers. Center: 7-slab classifiers. Right: difference in performance for the same offset values.

3.3. **Robustness of the 3D classifier.** In real-world remote sensing scenarios, the true location of the scatterers is unknown. Given the finite SAR resolution, see (12), and the presence of noise, the coordinates of individual scatterers can be

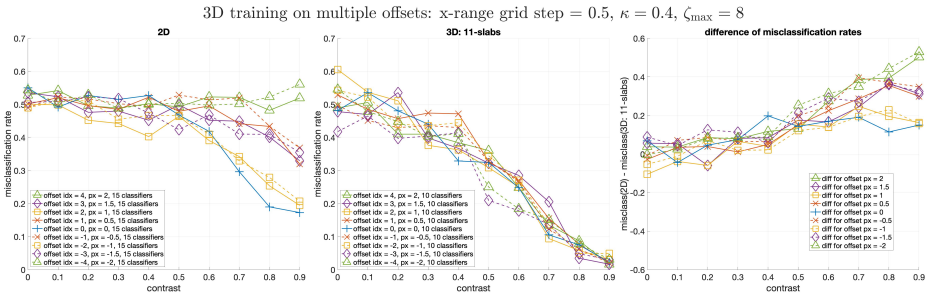

FIGURE 13. Same as in Figure 12 but for the offsets from $-2\Delta_{\mathrm{A}}$ to $2\Delta_{\mathrm{A}}$ with a step of $0.5\Delta_{\mathrm{A}}$. The 3D classifiers are 11-slabs.

determined only with some finite accuracy. In particular, it means that the actual value of the cross-range offset is unknown. In the context of the current work, it has two implications. First, a cross-range offset can lead to a significant increase of the misclassification rate, especially for 2D classifiers, see Sections 3.2.1 and 3.2.2. Second, having a classifier trained and tested on the same offset can be considered as an indirect way of providing the offset information to the classifier. This is true for all the results presented in Section 3.2.

In order to simulate the lack of information about the actual cross-range coordinate of the scatterers, we perform the training on datasets that combine several values of the cross-range offset. The resulting classifiers can be tested on each offset separately, the same way as the target contrasts are currently treated. This will make the misclassification rate a function of two variables: contrast and offset. Unlike in Section 3.2.2, we choose to restrict the offset value such that the location of the inhomogeneous scatterers corresponds to one of the inner slices. In particular, for 7- and 11-slab classifiers used in this experiment, the dataset includes inhomogeneous scatterers on slices 2 through 6 and 2 through 10, respectively. The resulting misclassification curves are displayed in Figures 12 and 13. We see that the performance is close to that reported in Figure 10.

Two conclusions can be drawn from the previous result. First, we observe that the offset information does not play a significant role in the training. Second and most important, we clearly see an advantage of using multi-slab classifiers in terms of robustness, i.e., tolerance to inaccuracies in predicting the cross-range location of the scatterer. As a measure of robustness, we can use the magnitude of the cross-range offset that does not result in a substantial degradation of the classifier performance. For the cases shown in Figures 12 and 13, the robustness increases by a factor of about 2 and 8, respectively, as compared to that for 2D classifiers, see Section 3.2.1.

3.4. **Discretization in the cross-range direction.** Natural guidelines for choosing the grid sizes are the spatial and temporal scales of intensity of the imaging kernel $|W|$, see (6) and Figures 5 and 6. In particular, for the range and cross-range directions, these will be the resolution sizes Δ_{R} and Δ_{A}, respectively, see (12). Discretization with a smaller step will result in the correlated image values. Nevertheless, it is a legitimate question whether a smaller grid size may improve the

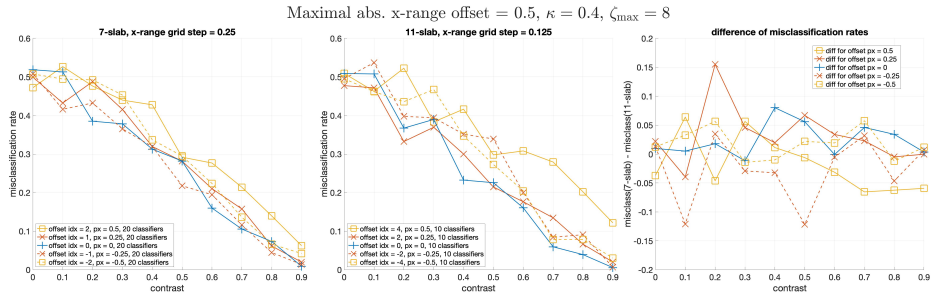

FIGURE 14. Performance of multi-slabs with different cross-range grid steps, see Section 3.4. In the left and center panels, the curves with the same line style and color correspond to different offset indices but same offset distance. Similarly to Figure 12, the right panel plots the difference between the data displayed in the center and left panels.

classification performance. In order to address this issue, we create pairs of multi-slab classifiers with different cross-range grid sizes and compare their performance.

Similarly to Section 3.3, the numerical experiment involves training on multiple cross-range offsets and testing on individual offsets. Figure 14 presents the results for the 7-slab and 11-slab classifiers with the cross-range grid sizes of $0.25\Delta_{\mathrm{A}}$ and $0.125\Delta_{\mathrm{A}}$, respectively. The envelopes of the training grids for both classifiers correspond to one and the same shape in the coordinate-delay space (illustrated by the parallelepiped in Figure 7), with the cross-range offsets not exceeding $0.5\Delta_{\mathrm{A}}$ by absolute value. At the same time, the training dataset for the 7-slab classifier includes 5 different offset values, whereas that for the 11-slab has 9 offset values. Despite the differences in discretization and training, the test results presented in Figure 14 don't show a significant difference in performance between the two classifiers. We see here that halving the grid size did not lead to an increase in the classification performance.

4. **Discussion.** We studied the classification performance for the problem of detecting the scattering delay in SAR images. Our approach extends that of [21] where two-dimensional classifiers were built based on convolutional neural networks. In the current work, we considered three-dimensional classifiers and compared their performance to that of the 2D ones.

The main advantage of the 3D classifiers is their improved robustness. Compared to 2D, they impose less stringent requirements on how accurately one should know the position of the scatterers. On the other hand, while the 3D classifiers use more input data, it does not translate into a better classification performance. Altogether, we consider the transition from 2D to 3D beneficial for classification problems although some questions related to classification performance still need to be addressed (see Appendix C).

We also note that, while the problem of identifying the dispersive targets in SAR images is important for applications, it has received little attention in the literature. Work [9] is based on a coordinate-delay imaging functional, similar to what we use in this paper. Work [1] employs sub-banding, while work [3] relies on a special

parameterization of a delayed target response. A direct comparison of classification performance between our current methodology and any of these earlier approaches would be difficult to make because of considerable dissimilarities in the setup and assumptions. Therefore, in this work we chose to compare the results against those obtained using our own previously developed techniques.

The targets in this study were chosen so as to enable a convenient formulation of a "benchmark" classification case. Yet the reflectivity models we employ are not very realistic. The development and adaptation of a realistic dispersive target model is an important future goal. Other extensions may include the following topics.

- Special study of those rare cases that result in a misclassification for high target contrasts. Do these cases have something in common? Can the performance for high contrasts be improved even further?
- Combination of classification with the determination of the target position. Our preliminary experiments demonstrated that the YOLO technology [29] can be successfully utilized in such a scenario.
- Incorporation of scatterers with the off-grid coordinates. This addresses the issue of the so-called "inverse crime", see [7, Section 5.3].
- Defining and quantifying information for the purpose of classification, see Appendix C, and revisiting the results presented so far with this concept as a new tool.
- Analysis of imagery where multiple targets with different properties may be present.
- Analysis of targets that are both extended in space and delayed in time.
- Going beyond the plain delay detection and obtaining a resolution with respect to the delay variable. More sophisticated models of target dispersion will be needed here.

Appendix A. **Contrast scaling procedure and its validation.** In the 2D settings of [12–14, 21], the contrast has been defined as follows:

$$q^{(\mathrm{2D})} = \frac{K_s^{(\mathrm{2D})}\sigma_s^2}{K_s^{(\mathrm{2D})}\sigma_s^2 + K_{\mathrm{b}}^{(\mathrm{2D})}\sigma_{\mathrm{b}}^2 + K_{\mathrm{n}}^{(\mathrm{2D})}\sigma_{\mathrm{n}}^2} = \frac{K_t^{(\mathrm{2D})}\sigma_t^2}{K_t^{(\mathrm{2D})}\sigma_t^2 + K_{\mathrm{b}}^{(\mathrm{2D})}\sigma_{\mathrm{b}}^2 + K_{\mathrm{n}}^{(\mathrm{2D})}\sigma_{\mathrm{n}}^2}, \quad (32)$$

see, in particular, [12, equation (88) and Appendix]. The parameters $K_\alpha^{(\mathrm{2D})}$ are expressed via the radar parameters such as the bandwidth, incident angle, and the aperture angle, and their values represent the scaling constants that appeared in the calculation of $\langle |I_\alpha|^2 \rangle$ for different α, as in Section 2.3. Note that the noise term $K_{\mathrm{n}}^{(\mathrm{2D})}\sigma_{\mathrm{n}}^2$ in (32) characterizes the intensity of the additive noise component I_{n} that is realized as an uncorrelated circular Gaussian process.

To validate our three-dimensional setup, we have extracted the 2D sub-images with the same geometry and discretization grid as in our previous 2D studies, see Figures 2, 5, and 7, and with the cross-range coordinate equal to that of the scatterer, i.e., $\eta = k_\parallel \varphi_T z_{\mathrm{d1}}$. We found that the statistics of image brightness and performance of the existing 2D classifiers for the sub-images did not match our previous results if formula (32) were used to define the contrast. The mismatch may be due to the cross-range sidelobes of the imaging kernel (6).

To make the contrast values in this work comparable to our previous results, we employed a definition similar to (32):

$$q = \frac{K_s\sigma_s^2}{K_s\sigma_s^2 + K_{\rm b}\sigma_{\rm b}^2 + K_{\rm n}\sigma_{\rm n}^2} = \frac{K_t\sigma_t^2}{K_t\sigma_t^2 + K_{\rm b}\sigma_{\rm b}^2 + K_{\rm n}\sigma_{\rm n}^2}, \tag{33}$$

and scaled the constants K_α in (33) as follows. We generated a significant amount of sampled single-component images I_α using the approach described at the end of Section 2.6. Similarly to the 2D case, we added noise to the right hand side of (24), with the intensity 2 times smaller than that of the background, i.e., $\langle|I_{\rm n}|^2\rangle = 0.5\langle|I_{\rm b}|^2\rangle$ [21], resulting in

$$q_s = \frac{K_s\sigma_s^2}{K_s\sigma_s^2 + 1.5K_{\rm b}\sigma_{\rm b}^2} = q, \quad q_t = \frac{K_t\sigma_t^2}{K_t\sigma_t^2 + 1.5K_{\rm b}\sigma_{\rm b}^2} = q, \tag{34}$$

depending on which of the two scenarios in (23) is considered. The values of the scaling factors K_α were determined by comparing the 95% percentiles of the intensity (i.e., $|I_\alpha|^2$) for the corresponding scatterer types between the sub-images and their 2D counterparts.

To further validate the 3D setting, we generated representative datasets of 3D cdSAR images with the same target and radar parameters, including the values of the contrast, as those employed in [21]. From these images, similarly to the foregoing contrast scaling procedure, we have extracted the 2D sub-images by taking $\eta = k_\parallel\varphi_T z_{\rm d1}$. Then we have applied the existing 2D classifiers to the resulting 2D datasets and confirmed that the misclassification curves come out similar to those in [21].

In addition, we wanted to see whether the classification quality of 2D classifiers is sensitive to the discretization grid. In particular, we considered a 2D cross-section of the 3D cdSAR image aligned with the inhomogeneous scatterers in the cross-range coordinate. While the shape of this figure is a parallelogram (shown in light blue in Figure 7), the discretization grid used in [21] covers a smaller square (shown in purple) with one side parallel to the ambiguity direction. Since this square is sufficiently large to cover the supports of both scatterers used in the experiment (instantaneous and delayed), we expected that the additional pixels in the same plane will not significantly affect the classification. The numerical experiments confirmed our expectations.

Appendix B. **Generation of correlated Gaussian variables.** The generation of multivariate Gaussian variables requires construction of the covariance matrices. In general, large matrices slow down the computations. Note that one of the goals of the transition from formula (27) to (28) was to reduce the size of the covariance matrices. Still, we have found that even with this simplification, the resulting matrices for the 3D case are too large for the efficient generation of image ensembles.[5] Specifically, since I_α in (28) is complex-valued, the size of the covariance matrix for an ambiguity plane with d sampling locations (see item 2 on page 12) is $2d \times 2d$. In order to capture the differences between the types of individual images illustrated in Figures 5 and 6, the value of d should be large. In particular, for the 32×32 discretization grid used in [12–14], each of the three scatterer types on the right hand sides of (23) leads to 32 covariance matrices of the size 64×64. For the 3D geometry in the current study, the number of discretization points in each ambiguity

[5]Note that in the two-dimensional treatment of cdSAR imaging used in our previous studies, the ambiguity planes of (13) turn into ambiguity lines by dropping the cross-range coordinate.

plane will be significantly larger than 32, and accordingly, the size of the covariance matrix will be significantly larger than 64×64.

The main bottleneck in the process of generation of correlated Gaussian is the factorization of the covariance matrix. The need for the factorization can be explained as follows. We assume that an efficient procedure of generation of i.i.d. standard normal variables is available, e.g., in the form of the Matlab© `randn` function. This way we can generate the Gaussian vectors $\mathbf{u} \in \mathbb{R}^{2d}$ with the covariance $\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \rangle = \mathbf{I}$, where $\mathbf{I}$ is the identity matrix. For a covariance matrix $\mathbf{C} \in \mathbb{R}^{2d \times 2d}$, the correlated zero-mean Gaussian vectors $\mathbf{v}$ satisfying $\langle \mathbf{v}\mathbf{v}^{\mathrm{T}} \rangle = \mathbf{C}$ can be obtained using a linear transform $\mathbf{v} = \mathbf{L}\mathbf{u}$, where $\mathbf{L} \in \mathbb{R}^{2d \times 2d}$ satisfies $\mathbf{C} = \mathbf{L}\mathbf{L}^{\mathrm{T}}$, see, e.g., [26]. Indeed, it is easy to show that

$$\langle \mathbf{v}\mathbf{v}^{\mathrm{T}} \rangle = \mathbf{L}\langle \mathbf{u}\mathbf{u}^{\mathrm{T}} \rangle \mathbf{L}^{\mathrm{T}} = \mathbf{L}\mathbf{I}\mathbf{L}^{\mathrm{T}} = \mathbf{C}.$$

In turn, the calculation of the matrix $\mathbf{L}$ is realized in the Matlab© `mvnrnd` function via the Cholesky factorization. We have found that the Cholesky factorization step for covariance matrices of the required size becomes computationally expensive.

In order to have an efficient way of simulating large datasets needed for training the neural networks, we have developed an alternative approach to the simulation of discretized 3D cdSAR images. In particular, we populate the support of the target or clutter with a circular Gaussian noise scaled with the corresponding σ_α, see Section 2.3. The convolution product of the resulting "sources" with the three-dimensional imaging kernel (6) is evaluated numerically. It can be shown that the covariance of the image values obtained from this procedure is very close to the one resulting from the "continuous" formula (28). Additional validation of this method is described in Appendix A.

Appendix C. Useful and "noise" information in the offset slabs of cdSAR images. In Section 3.2.3, we reported an unexpected result that, the performance of a zero-offset 2D classifier could not be improved upon by employing 3D classifiers, including the cases where the 3D discretization grid contains the 2D one as a subset. In other words, while each additional value of the cross-range coordinate brings along the same number of new discretization points as the original 2D grid, the classification performance remains unaffected.

This result is somewhat counterintuitive as additional samples would normally be interpreted as additional information, i.e., something useful for discrimination. To gain further insight into this, we have conducted several numerical experiments that we describe below.

In the first experiment, to highlight the role of the offset slabs, we performed the train/test sequence on the 3D cdSAR images from which we removed the "center slice" (i.e., all discretization points with the cross-range coordinate equal to that of the inhomogeneous scatterers). In Figures 15 and 16, such classifiers are called "PS-clipped". Figure 15 demonstrates a decent discrimination performance of the classifier based on the remaining data. Consequently, some information about the scatterers can be derived from the offset slabs. Moreover, we can observe that the classifiers with and without the center slice have common features. In particular, an increase in the number of slabs (in Figure 15, from 11 to 32) may lead to a decrease of performance, cf. Figure 11.

A notable observation made from similar plots in Figure 16 is that the "PS-clipped" 32-slab classifier outperforms the 2D classifiers with the offsets Δ_{A} and $2\Delta_{\mathrm{A}}$

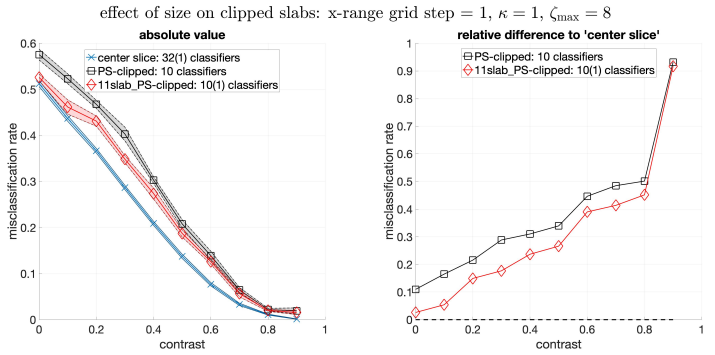

FIGURE 15. Similar to Figure 11 but for multi-slabs with the central slab removed.

for moderate and high contrasts. Therefore, in some situations the performance of a multi-slab is better than that of the best of its 2D components. In other words, here we have examples where a 3D classifier is able to "combine the information" from several 2D slices, so that adding more slices does not decrease the discrimination quality, as compared to Figures 11 and 15.

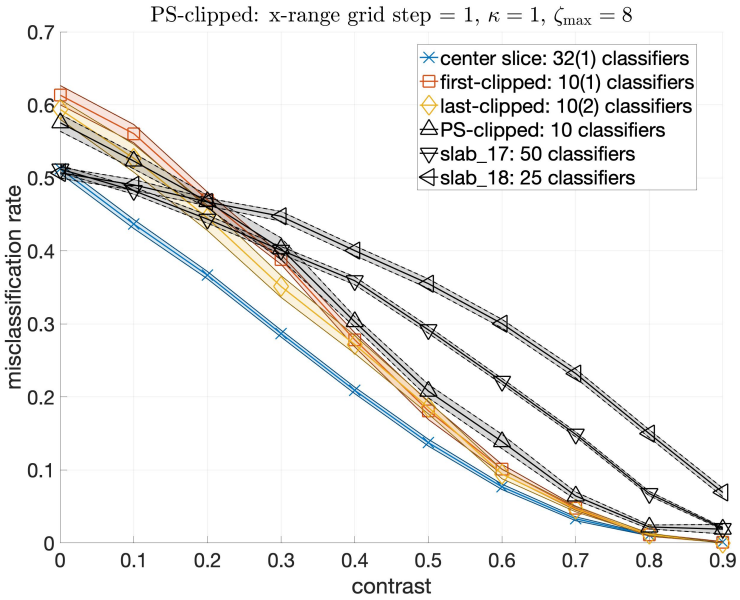

FIGURE 16. Multi-slabs with the slab aligned with the scatterer removed, as compared to the offset 2D slabs.

The next experiment attempts to reveal the relation between the three-dimensional structure in cdSAR images and the discrimination quality. We have compared performances of two types of classifiers for the 7-slab and 11-slab discretization grids: the first type contained three-dimensional convolutional kernels, whereas the classifiers of the second type had convolutions only in the range-delay plane, whereas

in the cross-range coordinate the kernels were fully connected. Note that the role of convolutions in CNN is to detect geometrical features, hence the fully connected option is equivalent to disregarding the geometrical structure in the cross-range direction and treating each slice as a separate feature map, see [15, Chapter 9]. Figure 17 shows no significant difference in performance for 7-slabs, whereas for 11-slabs, the architecture with convolutions in the cross-range dimension shows a superior performance as compared to that without the convolutions. This observation indirectly supports the hypothesis that capturing the structure of cdSAR images in the cross-range direction may increase the classification performance, although such a gain could not be seen in the experiments with the original images, as described in Section 3.2.3.

Altogether, the key benefit of adding the cross-range coordinate to cdSAR images is the increase of classification robustness (see Sections 3.2.2 and 3.3). In addition, the 3D discretization grids can collect more useful information about the scatterers compared to the 2D grids. Yet the question of whether or not the additional information will translate into a gain in classification performance requires further investigation.

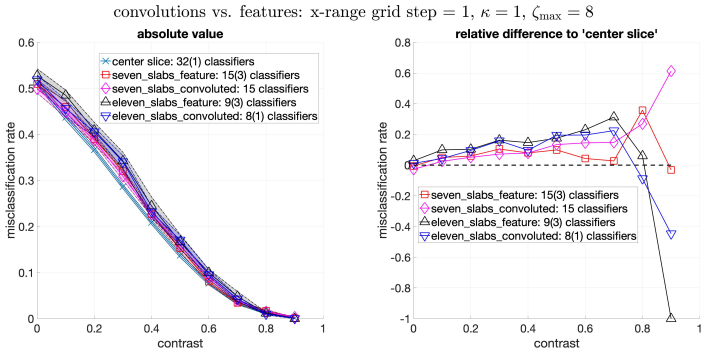

FIGURE 17. Multislabs with and without convolution in the cross-range dimension. The organization of the left and right panels is similar to that in Figure 11.

Acknowledgments. This work was supported by the US Air Force Office of Scientific Research (AFOSR) under awards number FA9550-21-1-0086, FA9550-23-1-0101, and FA9550-24-1-0172. The authors are thankful to Prof. Kevin Flores and Dr. Kyle Nguyen (NCSU) for their help with running the simulations.

REFERENCES

[1] R. A. Albanese and R. L. Medina, Materials identification synthetic aperture radar: progress toward a realized capability, *Inverse Problems*, **29** (2013), 054001 (11pp).

[2] J. C. Allen and S. L. Hobbs, Spectral estimation of non-stationary white noise, *J. Franklin Inst. B*, **334** (1997), 99-116.

[3] B. Borden, Dispersive scattering for radar-based target classification and duct-induced image artifact mitigation, in *NATO Symposium on Non-Cooperative Air Target Identfication Using Radar*, North Atlantic Treaty Organization, Research and Technology Organization, 1998, 14.1-14.7.

[4] V. C. Chen and H. Ling, *Time-frequency Transforms for Radar Imaging and Signal Analysis*, Artech House Radar Library, Artech House, Norwood, MA, 2002.

[5] M. Cheney, A mathematical tutorial on synthetic aperture radar, *SIAM Rev.*, **43** (2001), 301-312 (electronic).
[6] M. Cheney and B. Borden, *Fundamentals of Radar Imaging*, vol. 79 of CBMS-NSF Regional Conference Series in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2009.
[7] D. Colton and R. Kress, *Inverse Acoustic and Electromagnetic Scattering Theory*, vol. 93 of Applied Mathematical Sciences, 3rd edition, Springer, New York, 2013.
[8] I. G. Cumming and F. H. Wong, *Digital Processing of Synthetic Aperture Radar Data. Algorithms and Implementation*, Artech House, Boston, 2005.
[9] M. Ferrara, A. Homan and M. Cheney, Hyperspectral SAR, *IEEE Transactions on Geoscience and Remote Sensing*, **55** (2017), 1682-1695.
[10] G. Franceschetti and R. Lanari, *Synthetic Aperture Radar Processing*, Electronic Engineering Systems Series, CRC Press, Boca Raton, FL, 1999.
[11] M. Gilman, E. Smith and S. Tsynkov, *Transionospheric Synthetic Aperture Imaging*, Applied and Numerical Harmonic Analysis, Birkhäuser/Springer, Cham, Switzerland, 2017.
[12] M. Gilman and S. Tsynkov, Detection of delayed target response in SAR, *Inverse Problems*, **35** (2019), 085005 (38pp).
[13] M. Gilman and S. Tsynkov, Statistical characterization of scattering delay in synthetic aperture radar imaging, *Inverse Problems & Imaging*, **14** (2020), 511-533.
[14] M. Gilman and S. Tsynkov, Divergence measures and detection performance for dispersive targets in SAR, *Radio Science*, **56** (2021), e2019RS007011.
[15] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, Adaptive Computation and Machine Learning, MIT Press, Cambridge, MA, 2016, http://www.deeplearningbook.org.
[16] J. W. Goodman, Some fundamental properties of speckle, *J. Opt. Soc. Am.*, **66** (1976), 1145-1150.
[17] J. W. Goodman, Statistical properties of laser speckle patterns, in *Laser Speckle and Related Phenomena*, Springer, 1984, 9-75.
[18] T. Hastie, R. Tibshirani and J. Friedman, *The Elements of Statistical Learning*, 2nd edition, Springer Series in Statistics, Springer, New York, 2009, Data mining, inference, and prediction.
[19] C. V. Jakowatz Jr., D. E. Wahl, P. H. Eichel, D. C. Ghiglia and P. A. Thompson, *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*, Springer, 1996.
[20] A. D. Kim and C. Tsogka, Synthetic aperture imaging of dispersive targets, *IEEE Transactions on Computational Imaging*, **9** (2023), 954-962.
[21] J. Lagergren, K. Flores, M. Gilman and S. Tsynkov, Deep Learning Approach to the Detection of Scattering Delay in Radar Images, *J. Stat. Theory Pract.*, **15** (2021), Paper No. 14, 19pp.
[22] J.-S. Lee and E. Pottier, *Polarimetric Radar Imaging from Basics to Applications*, Optical Science and Engineering, CRC Press, Boca Raton, 2009.
[23] D. Massonnet and J.-C. Souyris, *Imaging with Synthetic Aperture Radar*, Engineering Sciences: Electrical Engineering, EFPL Press. Distributed by CRC Press, Lausanne, Switzerland, 2008.
[24] J. McCorkle and L. Nguyen, *Focusing of Dispersive Targets Using Synthetic Aperture Radar*, Technical report, ARMY RESEARCH LAB ADELPHI MD, 1994.
[25] R. Medina, J. Penn and R. Albanese, *Dielectric Response Data on Materials of Military Consequence*, Technical Report AFRL-HE-BR-TR-2002-0155, United States Air Force Research Laboratory, Human Effectiveness Directorate, Directed Energy Bioeffects Division, Biomechanics and Modeling Branch, Brooks AFB, San Antonio, TX, 2002.
[26] T. Muschinski, G. J. Mayr, T. Simon, N. Umlauf and A. Zeileis, Cholesky-based multivariate Gaussian regression, *Econometrics and Statistics*, **29** (2024), 261-281.
[27] C. Oliver and S. Quegan, *Understanding Synthetic Aperture Radar Images*, Artech House Remote Sensing Library, Artech House, Boston, 1998.
[28] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan, T. Killeen, Z. Lin, N. Gimelshein, L. Antiga, A. Desmaison, A. Kopf, E. Yang, Z. DeVito, M. Raison, A. Tejani, S. Chilamkurthy, B. Steiner, L. Fang, J. Bai and S. Chintala, Pytorch: An imperative style, high-performance deep learning library, in *Advances in Neural Information Processing Systems*, Curran Associates, Inc., **32** (2019), 8024-8035, URL http://papers.neurips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf.
[29] J. Redmon, S. Divvala, R. Girshick and A. Farhadi, You only look once: Unified, real-time object detection, in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, 779-788.

[30] T. Scarnati and B. Lewis, A deep learning approach to the synthetic and measured paired and labeled experiment (SAMPLE) challenge problem, in *Algorithms for Synthetic Aperture Radar Imagery XXVI*, International Society for Optics and Photonics, **10987** (2019), 109870G.

[31] Umbra News, Umbra releases over $4 million in free SAR data, https://umbra.space/blog/umbra-releases-over-4-million-in-free-sar-data/, 2024.

[32] T. Zhang, T. Zeng and X. Zhang , Synthetic aperture radar (SAR) meets deep learning, *Remote Sensing*, **15** (2023), 303. URL https://www.mdpi.com/journal/remotesensing/special_issues/synthetic_aperture_radar_meets_deep_learning.

[33] X. X. Zhu, S. Montazeri, M. Ali, Y. Hua, Y. Wang, L. Mou, Y. Shi, F. Xu and R. Bamler, Deep learning meets SAR: Concepts, models, pitfalls, and perspectives, *IEEE Geoscience and Remote Sensing Magazine*, **9** (2021), 143-172.

Received for publication June 2024; early access July 2024.