Journal of Statistical Theory and Practice (2021) 15:14
https://doi.org/10.1007/s42519-020-00149-w

ORIGINAL ARTICLE

Deep Learning Approach to the Detection of Scattering Delay in Radar Images

John Lagergren[1] · Kevin Flores[1] · Mikhail Gilman[1] · Semyon Tsynkov[1]

Accepted: 12 November 2020

Abstract
In radar imaging, stochastic target models are routinely used to describe distributed scatterers. In such models, the reflectivity of a target or clutter is a realization of a stochastic process with certain autocorrelation properties. While most targets reflect the impinging electromagnetic radiation instantaneously, some targets with complicated geometry and/or material composition may exhibit delayed scattering. Detecting such delays will provide valuable data for target identification. However, the scattering delay can be confused with the signal propagation delay, and this difference is sometimes rather subtle. Due to the stochastic nature of the radar data, the classification errors are inevitable. The misclassification rate depends on the parameters characterizing the radar system, imaging scene, and observation settings. A convolutional neural network is applied to the problem of discrimination between the instantaneous and delayed targets in synthetic aperture radar images. A trained neural network demonstrates the discrimination quality commensurate with that of the benchmark maximum likelihood-based classifier.

Keywords Radar imaging · Dispersive targets · Classification · Convolutional neural network

1 Introduction

A radar interrogates its target by having a series of electromagnetic pulses bounce off it and then processing the scattered responses. Various radars perform various tasks, such as detection, tracking, classification, and imaging. Imaging radars, in particular, generate maps of the target reflectivity. They do so by measuring the

This article is part of the topical collection "Advances in Deep Learning" guest edited by David Banks, Ernest Fokoué, and Hailin Sang.

✉ Mikhail Gilman
mgilman@ncsu.edu

[1] Department of Mathematics, North Carolina State University, Raleigh, NC 27659, USA

Published online: 30 November 2020

travel delay, which is the signal round-trip time between the radar antenna and the target. The delay times are subsequently converted into distances in image coordinates using the signal propagation speed (i.e., the speed of light). Two-dimensional images are obtained by synthetic aperture radars (SAR), where the antenna is mounted on a moving platform (aircraft or satellite). A SAR sensor interrogates the target from a number of consecutive locations along the flight path and applies a coherent signal processing algorithm to the resulting set of reflected signals [5, 6, 10, 16].

Most targets reflect the impinging electromagnetic radiation instantaneously. Other targets exhibit scattering delay which may be due to their material composition or shape (e.g., presence of cavities) [1, 3, 4, 7]. The ability to detect the scattering delay is valuable because it may provide important information for target classification and identification. However, the increase in the signal travel time can often be incorrectly attributed to a longer travel distance. When this happens, the image acquires artifacts: certain details in the image do not correspond to any physical object and rather appear as implications of a mismatch between the signal processing assumptions and the actual physics of signal propagation and scattering.

A signal processing technique for the detection of a delayed target return has been proposed in [7] and further developed in [11]. It builds a coordinate-delay SAR (cdSAR) image, which is an approximate reconstruction of the scene reflectivity as a function of the spatial coordinates and scattering delay. The discrimination between the propagation and scattering delay is based on the assumption that whereas for a given target, the propagation delay is a function of the antenna position, the scattering delay is not because it is defined by the internal properties of the target, as illustrated in Fig. 1. Using the same coherent signal processing procedure as applied to the set of reflected signals in SAR, one can separate the

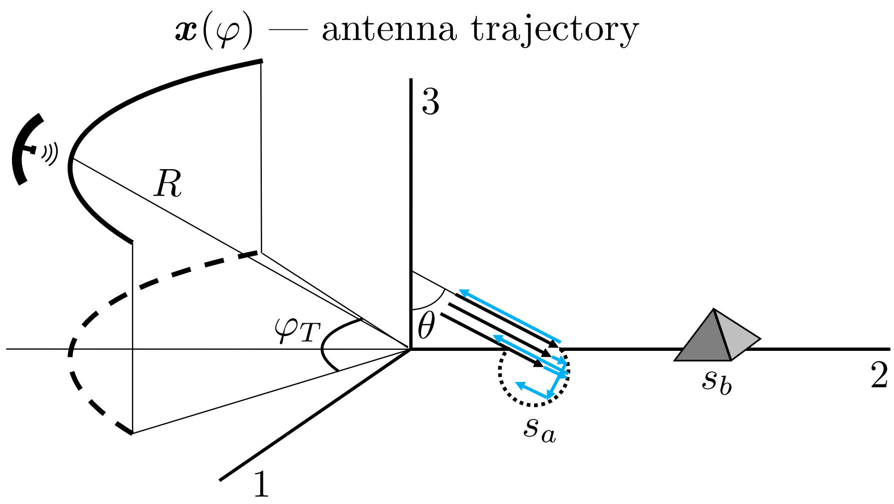

Fig. 1 Three-dimensional geometry of SAR imaging. A delayed scatterer is schematically shown as having an internal cavity where the multi-path reflection (blue arrows) of the incoming signal (black arrows) leads to scattering delays. Linear dimensions and angles are not to scale

constant (scattering) and variable (propagation) parts of the total delay, although this separation is not as good as resolution of the standard SAR images [11]. To discriminate between the instantaneous and delayed targets in cdSAR images, a maximum likelihood (ML)-based classifier has been developed, and its performance evaluated in numerical experiments [11, 12].

In the meantime, convolutional neural networks (CNNs) have emerged as a powerful tool for the fast and accurate classification of image data. Typically, CNNs are used in a supervised learning paradigm that requires a large set of input–output pairs, where the input is the image and the output is its true class [14]. While training the network typically requires a lot of computational resources, application of a trained network to a single input is very inexpensive, contrary to solving the optimization problem for a given input in the ML-based classification. Besides, the statistical model of cdSAR images and targets used in [11, 12] can easily produce large sets of simulated images of different types. Those sets can be used for training the network, and hence, it makes sense to try and employ CNNs as an alternative to the ML-based classification for fast discrimination between the instantaneous and delayed targets.

The goal of this work is to compare the capacity of the two classification methods (ML-based and CNN-based) to discriminate between the coordinate-delay radar images of instantaneous and delayed targets for different values of the system and target parameters. Additionally, we will explore the ability of these methods to generalize to the parameter values beyond those that they have been configured (or trained) for.

Note that conventional SAR provides image resolution in the directions called range and cross-range [6, 16]. For a linear antenna trajectory, these are the distance to the trajectory and the coordinate along the trajectory, respectively, while the localization in the third coordinate involves some external considerations, such as that all scatterers are located on a certain plane [10, Chapter 7], e.g., the Earth's surface. A standard two-dimensional SAR image usually corresponds to a small patch of this plane and is expressed in terms of the range and cross-range coordinates. As an extension of the standard SAR, cdSAR images are functions of the two spatial coordinates and, additionally, the delay time (i.e., three dimensions in total). In this work, we consider imaging restricted to a plane normal to the antenna trajectory. This leaves just one spatial coordinate, i.e., range, and the delay time, so the images are two-dimensional. This simplification still preserves the main challenge of the discrimination, as will be seen in Sect. 2. At the same time, antenna locations outside the aforementioned plane cannot be ignored because this is what enables the discrimination between the scattering and propagation delay. In what follows, we will use the term "range-delay image" for the latter type of two-dimensional radar images.

Section 2 presents the key details about the physical imaging model that is used to generate the data, while a more thorough description can be found in [11, 12]. A brief exposition of the ML-based classification method is given in Sect. 3. Section 4 describes the CNN-based classification approach, including the network architecture and the training methodology. Section 5 presents the results of the CNN-based classification and their comparison with the ML-based method. Finally, Sect. 6 outlines the conclusions.

2 Essentials of the Range-Delay Radar Imaging

2.1 Scatterer Models and Image Statistics

For the purposes of discrimination between the propagation and scattering delay, we use the models of delayed and instantaneous targets, background, and noise developed and used in [11, 12]. To make this text self-sufficient, we present below the mathematical relations used to build the data for this publication with a small amount of comments.

The geometry of the imaging assumes the antenna trajectory in the form of an arc of a circle rather than a straight line, so the range direction is given by the radius drawn toward the middle point of this arc, see Fig. 1. The antenna coordinates are given by $\boldsymbol{x} = (x_1, x_2, x_3) = (-R\sin\theta\cos\varphi, -R\sin\theta\sin\varphi, R\cos\theta)$, where φ defines a point on an arc of a circle with the angular extent φ_T: $|\varphi| \le \varphi_T/2 \ll 1$, and θ is the incidence angle. The delayed and instantaneous targets are located at $z_a = (0, s_a, 0)$ and $z_b = (0, s_b, 0)$, respectively, such that for $s_b > s_a$, the delayed return from z_a can interfere from the instantaneous return from z_b. For $R \gg \max(|s_a|, |s_b|)$, the difference between the signal *travel* delays for these two targets is

$$\Delta t(\varphi) = \frac{2}{c}\big(|z_b - \boldsymbol{x}(\varphi)| - |z_a - \boldsymbol{x}(\varphi)|\big) \approx \frac{2}{c}\sin\theta(s_b - s_a)\cos\varphi, \tag{1}$$

see [11], i.e., depends on the antenna location. In (1), c is the speed of light, and the factor 2 is due to the two-way travel. However, this dependence is weak for small φ: $\cos\varphi \approx 1 - \varphi^2/2$, so if a single target at z_a exhibits a constant scattering delay equal to $\Delta t(0)$ in addition to the immediate return, the distinction from the configuration with two targets described by (1) may be not strong enough to be detectable.

The scatterer models employed in our study are based upon an extension of the stochastic speckle model [13, 18]. Each such model can be expressed using a universal delta-correlated circular Gaussian [2, 8, 9] coordinate-delay process $\rho(t_z, z)$ and a deterministic modulating function $f(t_z, z)$:

$$\nu(t_z, z) = f(t_z, z)\cdot\rho(t_z, z), \quad \text{where} \quad \langle\overline{\rho(t_z', z')}\rho(t_z, z)\rangle = \delta(t_z' - t_z)\delta(z' - z), \tag{2}$$

$\nu(t_z, z)$ is the reflectivity function, t_z is the scattering delay, and z is the scatterer coordinate with $z_3 = 0$, see Fig. 1. The target type is defined by the modulation function; in particular, we will be using the following three formulations:

1. $f_\mathrm{b}(t_z, z) = \sigma_\mathrm{b}\delta(t_z)$—instantaneous speckled background with the average intensity σ_b^2;
2. $f_t(t_z, z) = \sigma_t\delta(z)\cdot\mathbf{1}_{0\le t_z\le t_{\max}}$—delayed point scatterer at the origin of coordinates ($s_a = 0$, see Fig. 1) and maximum delay $t_{\max}$, where $\mathbf{1}_{\mathcal{D}}$ denotes the characteristic function of the domain $\mathcal{D}$;

3. $f_s(t_z, \boldsymbol{z}) = \sigma_s \delta(t_z)\delta(z_1) \cdot \mathbf{1}_{0 \le z_2 \le s_{\max}}$—instantaneous scatterer with the support on the z_2-axis between 0 and $s_{\max}$, see Fig. 1.

The total reflectivity is realized by one of the two scenarios:

$$
\begin{aligned}
\nu_{\text{s-model}}(t_z, \boldsymbol{z}) &= \nu_{\mathrm{b}}(t_z, \boldsymbol{z}) + \nu_s(t_z, \boldsymbol{z}) \\
&= f_{\mathrm{b}}(t_z, \boldsymbol{z})\rho_1(t_z, \boldsymbol{z}) + f_s(t_z, \boldsymbol{z})\rho_2(t_z, \boldsymbol{z}), \\
\nu_{\text{t-model}}(t_z, \boldsymbol{z}) &= \nu_{\mathrm{b}}(t_z, \boldsymbol{z}) + \nu_t(t_z, \boldsymbol{z}) \\
&= f_{\mathrm{b}}(t_z, \boldsymbol{z})\rho_1(t_z, \boldsymbol{z}) + f_t(t_z, \boldsymbol{z})\rho_2(t_z, \boldsymbol{z}),
\end{aligned}
\tag{3}
$$

where ρ_1 and ρ_2 are two different realizations of the stochastic process $\rho(t_z, \boldsymbol{z})$ defined in (2). The detection of the delayed target is understood as a discrimination between the two models of target reflectivity in (3).

A typical frequency-modulated radar pulse with the central frequency ω_0, duration τ, and bandwidth B can be formulated as

$$
P(t) = e^{-i\omega_0 t} e^{-iBt^2/(2\tau)} \cdot \mathbf{1}_{|t| \le \tau/2}.
$$

Radar imaging can be considered as a linear operator acting on the reflectivity function $\nu(t_z, \boldsymbol{z})$ and producing an image $I(t_y, \boldsymbol{y})$, which is an approximation to the reflectivity function derived from the scattered radar signals. As it was mentioned in the Introduction, the scatterers are assumed to be located at a certain horizontal plane that we define as $z_3 = 0$; accordingly, in what follows, we assume $y_3 = 0$ in the argument of I. The imaging operator relating $\nu(t_z, \boldsymbol{z})$ and $I(t_y, \boldsymbol{y})$ is of convolutional type [11]:

$$
I(t_y, \boldsymbol{y}) = \int_0^\infty \left(\iint \nu(t_z, \boldsymbol{z}) W(t_y - t_z, \boldsymbol{y} - \boldsymbol{z}) \, \mathrm{d}z_1 \, \mathrm{d}z_2 \right) \mathrm{d}t_z. \tag{4}
$$

The expression for the kernel in (4) is

$$
W(t_y, \boldsymbol{y}; t_z, \boldsymbol{z}) = e^{-2i\omega_0 T} \cdot \boldsymbol{\Phi}\big((k_\| \varphi_T (y_1 - z_1), k_\| \varphi_T^2 (y_2 - z_2)\big) \cdot \operatorname{sinc}(BT), \tag{5}
$$

where $k_\| = \omega_0 \sin\theta / c$, $\operatorname{sinc}\xi = \frac{\sin\xi}{\xi}$,

$$
T = \frac{y_2 - z_2}{c} \sin\theta + \frac{t_y - t_z}{2}, \tag{6}
$$

and

$$
\boldsymbol{\Phi}(v_1, v_2) = \int_{-1/2}^{1/2} e^{2iv_1 s} e^{iv_2 s^2} \, \mathrm{d}s. \tag{7}
$$

We are considering range-delay images, i.e., always take $y_1 = 0$ in the argument of $I(t_y, \boldsymbol{y})$ (note that the support of f_t and f_s does not extend beyond $z_1 = 0$). This leaves just two arguments of I: delay t_y and range y_2. It is convenient to use the following dimensionless coordinates:

$$
\begin{aligned}
\zeta &= \frac{B}{\omega_0} k_{\|} y_2 + B \frac{t_y}{2}, \\
\psi &= \frac{B}{\omega_0} k_{\|} y_2 - B \frac{t_y}{2},
\end{aligned} \tag{8}
$$

such that $I \equiv I(\zeta, \psi)$.

We assume that the intervals of delays and coordinates in the expressions for f_t and f_s, respectively, are related by $ct_{\max}/2 = s_{\max} \sin\theta$ (cf. (1) and (6)). Introduce

$$
\zeta_{\max} = \frac{Bt_{\max}}{2} = \frac{B}{\omega_0} k_{\|} s_{\max}. \tag{9}
$$

Using (2)–(9), it is possible to obtain the second-order statistics of partial images, i.e., the images due to the reflectivity functions ν_{b}, ν_s, and ν_t in (3). We present these expressions in the following concise form:

$$
\left\langle I_\alpha(\zeta, \psi) \overline{I_\alpha(\zeta, \psi')} \right\rangle = \sigma_\alpha^2 K_\alpha H_\alpha(\zeta, \psi, \psi'). \tag{10}
$$

where the index $\alpha \in \{\mathrm{b}, s, t\}$ denotes the one of the three scatterer types used in (3). The dimensionless functions H_α normalized such that $\max |H_\alpha| \approx 1$ and the normalization coefficients K_α are expressed as follows [12]:

$$
\begin{aligned}
H_{\mathrm{b}}(\zeta, \psi, \psi') &= \Phi\left(0, \kappa \frac{\psi - \psi'}{2}\right), \\
H_t(\zeta, \psi, \psi') &= \Phi\left(0, \kappa \frac{\zeta + \psi}{2}\right) \overline{\Phi\left(0, \kappa \frac{\zeta + \psi'}{2}\right)} \frac{1}{\pi} \int_0^{\zeta_{\max}} \operatorname{sinc}^2(\zeta - \zeta_z)\, d\zeta_z, \\
H_s(\zeta, \psi, \psi') &= \frac{1}{\pi} \int_0^{\zeta_{\max}} \operatorname{sinc}^2(\zeta - \zeta_z) \Phi\left(0, \kappa \frac{\zeta + \psi}{2} - \kappa \zeta_z\right) \overline{\Phi\left(0, \kappa \frac{\zeta + \psi'}{2} - \kappa \zeta_z\right)}\, d\zeta_z,
\end{aligned} \tag{11}
$$

and

$$
K_{\mathrm{b}} = \frac{\omega_0}{Bk_{\|}} \frac{1}{k_{\|} \varphi_T} \cdot \pi^2, \quad K_t = \frac{2}{B} \pi, \quad K_s = \frac{\omega_0}{Bk_{\|}} \pi,
$$

where

$$
\kappa = \varphi_T^2 \frac{\omega_0}{B}. \tag{12}
$$

Note that on the left hand side of (10), we have taken one and the same value of ζ in the arguments of I. The reason is the rapid decay of $\operatorname{sinc}(BT)$ in (5) on its argument that allows us to consider $I(\zeta, \psi)$ and $I(\zeta', \psi')$ uncorrelated given $|\zeta - \zeta'| \geq \pi$, see [11, 12] for more details. Revisiting formula (1) and the subsequent discussion, we note that the instantaneous target at z_b and a "ghost" of the target at z_a due to the scattering delay equal to $\Delta t(0)$, see Fig. 1, have the same value of ζ; this relation is called the range-delay ambiguity, and the condition $\zeta = \text{const}$ defines what we will call an *ambiguity line* in the range-delay plane.

In addition to (10)–(11), we introduce I_n as yet another type of partial images to represent the noise term. This is done by allowing $\alpha = \mathrm{n}$ in (10) and formally setting

$$H_\mathrm{n}(\zeta,\psi,\psi') = 0 \quad \text{if}\, \psi \neq \psi', \quad \text{and} \quad H_\mathrm{n}(\zeta,\psi,\psi') = 1 \quad \text{if}\, \psi = \psi', \tag{13}$$

leaving the value of $\sigma_\mathrm{n}^2 K_\mathrm{n}$ to be specified in (14) below.

The relative scatterer intensities, or contrasts, are defined as follows:

$$p_\mathrm{n} = \frac{\sigma_\mathrm{n}^2 K_\mathrm{n}}{\sigma_\mathrm{b}^2 K_\mathrm{b}}, \quad q_s = \frac{\sigma_s^2 K_s}{\sigma_s^2 K_s + \sigma_\mathrm{b}^2 K_\mathrm{b} + \sigma_\mathrm{n}^2 K_\mathrm{n}}, \quad q_t = \frac{\sigma_t^2 K_t}{\sigma_s^2 K_t + \sigma_\mathrm{b}^2 K_\mathrm{b} + \sigma_\mathrm{n}^2 K_\mathrm{n}}. \tag{14}$$

In this work, we always take $q_s = q_t$ and $p_\mathrm{n} = 0.5$.

2.2 Sampled Range-Delay Images and the Metadata

The discrimination between target types will be performed using sampled images. This means defining a set of arguments for the image, i.e., a set of pairs $\{(\zeta_j, \psi_j)\}$ describing the image sampling points.

Besides the sampling points, a complete description of the data includes the values of parameters ζ_max, κ, p_n, and q. Using these values, we can calculate three covariance matrices for the random vectors $I_\mathrm{b}(\zeta_j, \psi_j)$, $I_s(\zeta_j, \psi_j)$, and $I_t(\zeta_j, \psi_j)$, respectively, using (10)–(14), whereas the covariance matrix for $I_\mathrm{n}(\zeta_j, \psi_j)$ is a diagonal matrix scaled with $\sigma_\mathrm{n}^2 K_\mathrm{n}$. In particular, if among the sampling points we have M *distinct* values of ζ, i.e., $\{\zeta_m,\ 1 \le m \le M\}$, then formula (10) yields the covariance matrices $\mathbf{M}_{m\alpha}$, $\alpha \in \{\mathrm{b}, s, t\}$, for all sampling points on the ambiguity line defined by $\zeta_m, 1 \le m \le M$ [12].

The sampling pattern used in this work is a 32×32 rectangular grid in (ζ, ψ) -coordinates, with $\zeta = (-4 \dots 27) \cdot \pi$ and $\psi = (-15 \dots 16) \cdot \pi$. In the range-delay plane, this pattern is shown in Fig. 2 together with the supports of f_t and f_s for the specified value of ζ_max. This pattern provides the data in the form suitable for both discrimination methods considered in this work, see Sects. 3 and 4.

The covariance matrix $\mathbf{M}_\alpha$ is obtained by stacking all $\mathbf{M}_{m\alpha}$, $m = 1 \dots M$, diagonally such that, according to our previous discussion, the entries of $\mathbf{M}_\alpha$ corresponding to different ζ_m are zeros. Ensembles of multivariate zero-mean normal random vectors are then generated using these matrices: each such vector is a realization of a sampled partial image I_α. Depending on the scenario in (3), the partial images are combined as either

$$Q_j = I_\text{s-model}(\zeta_j, \psi_j) = I_\mathrm{b}(\zeta_j, \psi_j) + I_\mathrm{n}(\zeta_j, \psi_j) + I_s(\zeta_j, \psi_j). \tag{15a}$$

or

$$Q_j = I_\text{t-model}(\zeta_j, \psi_j) = I_\mathrm{b}(\zeta_j, \psi_j) + I_\mathrm{n}(\zeta_j, \psi_j) + I_t(\zeta_j, \psi_j). \tag{15b}$$

The vectors $\mathbf{Q}$ with entries of Q_j given by the left-hand sides of (15) represent sampled range-delay radar images. This way, the aforementioned four ensembles

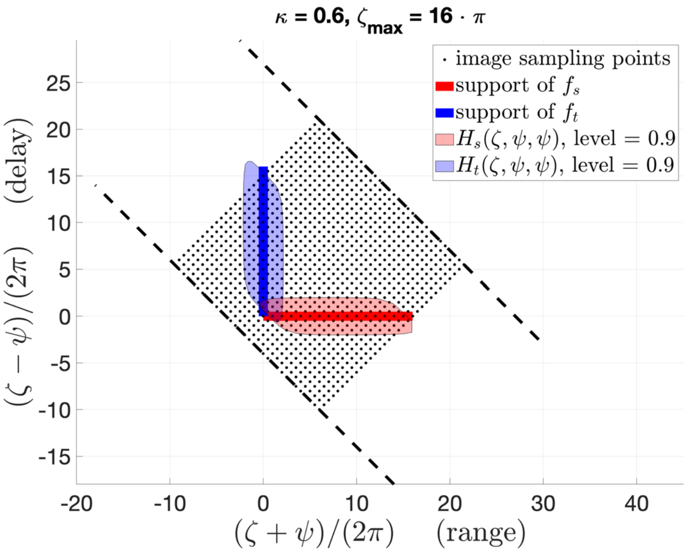

Fig. 2 The sampling pattern used for the discrimination between f_t and f_s in the range-delay radar images. The ambiguity lines (long dashes) correspond to the minimal and maximal values of ζ, see (8), in the sampling pattern, in this case $\zeta = -4\pi$ and $\zeta = 27\pi$ (not to be confused with the parameter $\zeta_{\max} = 16\pi$ that characterizes the scatterers)

of generated Gaussian random vectors give rise to two sub-ensembles of sampled images, one for the s-model and one for the t-model, with the same value of contrast: $q_s = q_t$.

In the discrimination problem, a single vector $\mathbf{Q}$ resulting from either (15a) or (15b) serves as data, whereas $\{(\zeta_j, \psi_j)\}$, κ, and $\zeta_{\max}$ are the known metadata. The goal is to determine which of the two scenarios in (15) has produced this data vector. The shaded areas in Fig. 2 display one level of the normalized expectations for $|I_s|^2$ and $|I_t|^2$, see (10). In general, an increase in κ squeezes these shaded areas toward the respective coordinate axes, thus improving the separation between the statistics of I_s and I_t [11, 12]. Similarly, an increase in $\zeta_{\max}$ also makes the two statistics more distinct. However, as the data are multivariate normal, any vector $\mathbf{Q}$ can result from either of the two models with a nonzero probability density. Hence, misclassifications will always take place, and we will evaluate the performance of a classifier by the average misclassification rate demonstrated on the two aforementioned sub-ensembles. Note that although the patterns due to f_s and f_t, see (3), are easily distinguishable in Fig. 2, the discrimination in realistic situations involving a *single* realization of the random processes described by (10) and (15), including the homogeneous background I_{b} and noise I_{n}, may become complicated, as illustrated in Fig. 3.

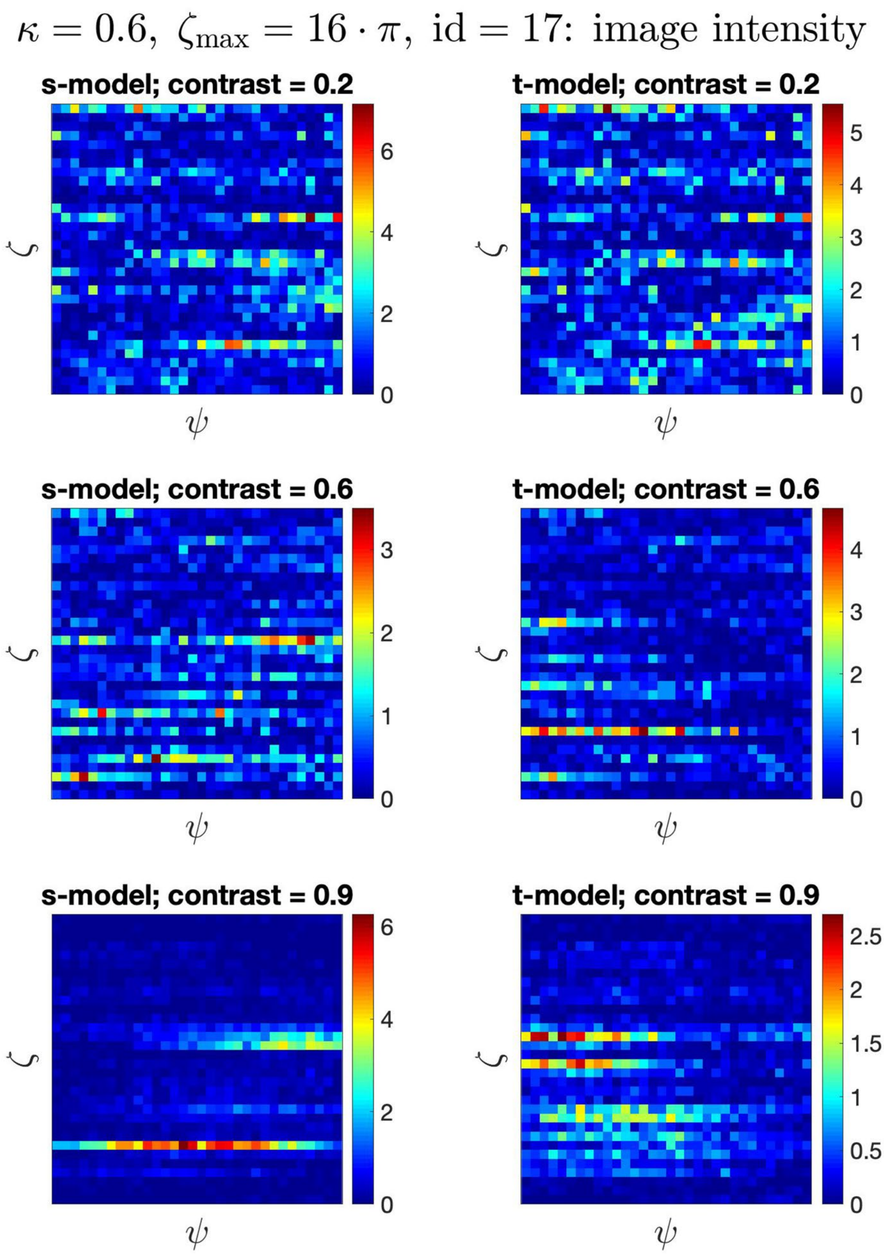

Fig. 3 Intensities for range-delay images generated according to (15). The left column of plots corresponds to the instantaneous target, whereas the right column corresponds to the delayed target. The rugged appearance of the images is due to the speckle. In the absence of the speckle, the level curves of image intensities would mostly have the slope of 1 and −1 in the left and right column, respectively, see also the plots of H_s and H_t in Fig. 2 (note that according to (8), the coordinate axes in these plots are rotated by 45° with respect to Fig. 2)

3 Maximum Likelihood Based Classifier

The true values of the intensities σ_{b}^2, σ_s^2, σ_t^2, and σ_{n}^2 (see the definitions of f_{b}, f_s, and f_t on page 3), as well as the choice of model in (15) are unknown to the classifier. Accordingly, we consider the matrices

$$\mathbf{M}_{\text{s-model}} = \mathbf{M}_{\mathrm{b}} + \mathbf{M}_{\mathrm{n}} + \mathbf{M}_s \quad \text{and} \quad \mathbf{M}_{\text{t-model}} = \mathbf{M}_{\mathrm{b}} + \mathbf{M}_{\mathrm{n}} + \mathbf{M}_t$$

as functions of these unknown intensities, i.e.,

$$\mathbf{M}_{\text{s-model}} = \mathbf{M}_{\text{s-model}}(\sigma_{\mathrm{b}}^2, \sigma_{\mathrm{n}}^2, \sigma_s^2) \quad \text{and} \quad \mathbf{M}_{\text{t-model}} = \mathbf{M}_{\text{t-model}}(\sigma_{\mathrm{b}}^2, \sigma_{\mathrm{n}}^2, \sigma_t^2),$$

see also (10). For each data vector $\mathbf{Q}$, we define the likelihood functions as the Gaussian probability densities:

$$\begin{aligned} p_{\text{s-model}}(\mathbf{Q};\sigma_{\mathrm{b}}^2, \sigma_{\mathrm{n}}^2, \sigma_s^2) &= \frac{1}{\sqrt{\det(2\pi\mathbf{M}_{\text{s-model}})}} \exp\left(-\frac{1}{2}\mathbf{Q}^{\mathbf{T}}(\mathbf{M}_{\text{s-model}})^{-1}\mathbf{Q}\right), \\ p_{\text{t-model}}(\mathbf{Q};\sigma_{\mathrm{b}}^2, \sigma_{\mathrm{n}}^2, \sigma_t^2) &= \frac{1}{\sqrt{\det(2\pi\mathbf{M}_{\text{t-model}})}} \exp\left(-\frac{1}{2}\mathbf{Q}^{\mathbf{T}}(\mathbf{M}_{\text{t-model}})^{-1}\mathbf{Q}\right), \end{aligned} \tag{16}$$

The classification based on the maximum likelihood principle is realized via the following comparison:

$$\max_{\sigma_{\mathrm{b}}^2,\sigma_{\mathrm{n}}^2,\sigma_s^2} p_{\text{s-model}}(\mathbf{Q};\sigma_{\mathrm{b}}^2, \sigma_{\mathrm{n}}^2, \sigma_s^2) \lessgtr \max_{\sigma_{\mathrm{b}}^2,\sigma_{\mathrm{n}}^2,\sigma_t^2} p_{\text{t-model}}(\mathbf{Q};\sigma_{\mathrm{b}}^2, \sigma_{\mathrm{n}}^2, \sigma_t^2). \tag{17}$$

The model yielding the larger of the two maxima in (17) assigns its type to the vector $\mathbf{Q}$. If this turns out to be the t-model, we declare that the underlying target contains a delayed component (see (3)).

4 CNN Method

The convolutional neural network (CNN) takes as input a $2 \times 32 \times 32$ real-valued array representing a 32×32 complex-valued image, see (15) and Figs. 2 and 3. The output is a probability score between 0 and 1 for an image to be due to the s-model. If the computed score is less than 0.5, then the underlying scatterer is classified as a t-model, otherwise as an s-model.

The network architecture is shown in Fig. 4. The convolutional layers employ the standard 3×3 kernels. Each convolutional layer is followed by batch normalization [15], a "Leaky ReLU" nonlinear activation function, and dropout [19] with rate 0.5. Batch normalization is used for improving the speed, performance, and stability of the CNN, "Leaky ReLU" introduces nonlinearity, and dropout is an easy method for preventing overfitting. The two-dimensional output of the first fully connected layer was used to monitor the learned decision boundary. The final fully connected layer is followed by a "sigmoid" activation function

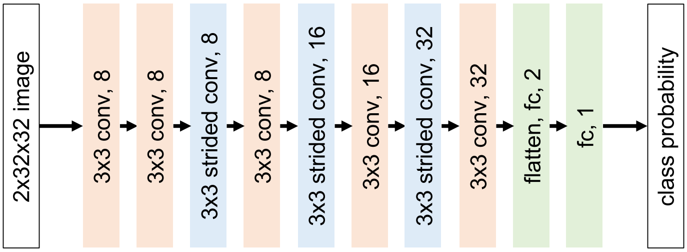

Fig. 4 Convolutional neural network architecture for classification of coordinate-delay radar images. White blocks indicate inputs and outputs, orange blocks indicate convolutional layers with associated size and number of filters, blue blocks indicate strided convolutional layers that downsample by a factor of 2, and green blocks indicate fully connected layers with associated number of neurons

to constrain the CNN outputs to between 0 and 1. For discrimination, the CNN outputs are rounded (i.e., to 0 for outputs < 0.5 and 1 for outputs ≥ 0.5). Binary cross-entropy is used as an objective function to train the CNN model.

A separate CNN is trained for each combination of values of κ and $\zeta_{\max}$, see Sect. 2.2. In order to facilitate the comparison with the ML-classifier of Sect. 3, each training set contains images corresponding to different values of contrast q, see (14). Unless specifically stated otherwise, we use 10 values of q uniformly spread between 0.0 and 0.9. For each contrast, the training set contains 2000 images generated using each of the two models in (15). The resulting set of 40,000 images is randomly partitioned into 80% training and 20% validation sets. The CNN is trained for a total of 100 epochs with a batch size of 64 using the Adam [17] optimizer with default values. The model parameters that result in the best validation accuracy are saved. Note that no augmentations (e.g., flips, rotations, additional noise) are applied to the data during training.

5 Results and Discussion

Our goal for this study is to compare the performance of CNN-based classification against the ML-based approach that we consider a benchmark. To evaluate performance of the two methods, 400 t-model and 400 s-model images are generated for each value of contrast from 0.0 to 0.9, totaling 8000 test images per each combination of κ and $\zeta_{\max}$. The ML-based and CNN-based classifiers are applied to each image, and the quality of discrimination is evaluated for each method and each contrast by averaging the misclassification rates shown on s-model and t-model subensembles, see Sect. 2.2.

Besides comparing the classification performance, we also explored the ability of the classification methods to adapt to missing training data or incorrect metadata. In particular, the value of the parameter $\zeta_{\max}$ characterizes the length of an instantaneous scatterer or the maximum delay of a dispersive scatterer, see (9). The formulation of maximum likelihood procedure assumes that this parameter is known, i.e., no optimization w.r.t. $\zeta_{\max}$ is made in (17). Similarly, a CNN is trained on a dataset generated with a fixed value of $\zeta_{\max}$, see Sect. 4. At the same time, in a remote sensing paradigm, it may be justified to consider all parameters of the target unknown, including $\zeta_{\max}$. One way of addressing this discrepancy is to change the formulations of the classification procedures, in particular, extend optimization in (17) to include the parameter $\zeta_{\max}$ for the ML-based method and include samples with different values of $\zeta_{\max}$ into the training dataset for the CNN-based method. We leave these options for future studies. Instead, we explore the ability of the two methods to generalize: for the ML-based classifier, this means that the values of $\zeta_{\max}$ (and, hence, functions H_α and matrices $\mathbf{M}_{\text{s/t-model}}$, see (10) and (16), respectively) at the classification stage, see (17), will be different from those used for dataset generation, see (15), whereas for the CNN-based classifier, the matrices $\mathbf{M}_{\text{s/t-model}}$ used to generate the training and validation datasets will be different from those used to generate the test dataset. This study is described in Sect. 5.1.

In terms of adaptation to different target contrasts, CNN-based method is different from the ML-based method in that in the latter, the target contrast is not part of the configuration of the classifier; rather, its value can be retrieved from the argument of the maximum in (17). Since no meaningful comparison between the two classification methods can be seen in this case, Sect. 5.2 presents the results on adaptation to missing/incorrect values of target contrast only for the CNN-based classifier.

5.1 Classification Performance with Matching and Mismatching Metadata

Figure 5 plots the average misclassification rates for the ML-based and CNN-based classifiers for $\kappa = 0.6$ and two different values of $\zeta_{\max}$: 8π and 16π. The upper left and bottom right plots illustrate the cases where the value of $\zeta_{\max}$ used for training/configuration of a classifier matches that of the test dataset, whereas the upper right and bottom left plots correspond to the cases where these values are different. The latter cases illustrate the ability of the classifier to generalize, i.e., perform with incorrect metadata.

Both methods demonstrate better performance with the larger value of $\zeta_{\max} = 16\pi$ as compared to $\zeta_{\max} = 8\pi$; this result is expected. We also observe that the discrimination quality shown by the two methods is very close in all four cases. This is a remarkable result given a significantly different nature of the CNN-based classifier as compared to the ML-based classifier. Both methods demonstrate a reasonable adaptation capability; however, the incorrect metadata reduces the classification quality (observe that the plots in each “off-diagonal” panel in Fig. 5 are higher than those in the “diagonal” panel in the same row).

In turn, Fig. 6 plots the average misclassification rates for the ML-based and CNN-based classifiers for $\zeta_{\max} = 8\pi$ and two different values of κ: 0.4 and 1. The arrangement of panels in Fig. 6 is similar to that of Fig. 5. For the upper left and bottom right plots, we see that the quality of classification is better for the larger value of κ, as expected. Similarly to Fig. 6, we observe that the performance of the two classifiers is very close. We also see that in one of the two adaptation cases, namely, the bottom left plot, the performance of the CNN-based classifier is significantly better than that of the ML-based classifier, whereas in the second case (the

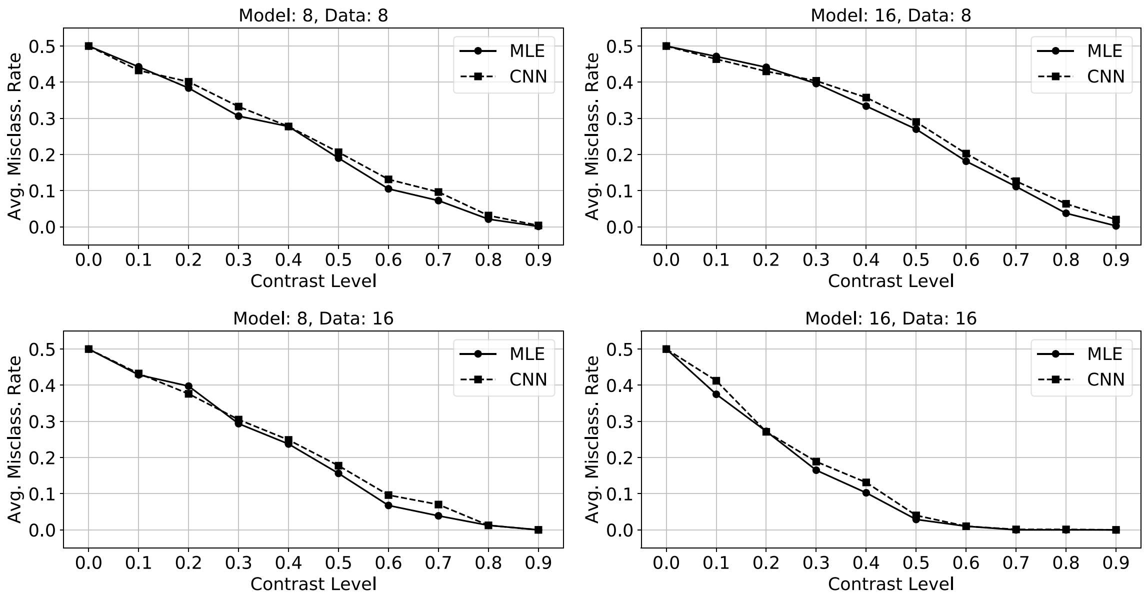

Fig. 5 Average misclassification rates for the ML-based and CNN-based classifiers for different values of $\zeta_{\max}$, see (9). The titles of each panel in the format "Model: XX, Data: YY" mean that for the ML-based method, the value of $\zeta_{\max} = \mathrm{XX} \cdot \pi$ is used at the classification step, see (17), whereas for the CNN-based method, $\zeta_{\max} = \mathrm{XX} \cdot \pi$ is used for generation of training/validation datasets, while $\zeta_{\max} = \mathrm{YY} \cdot \pi$ is used to generate the common test dataset for the two plots in this panel

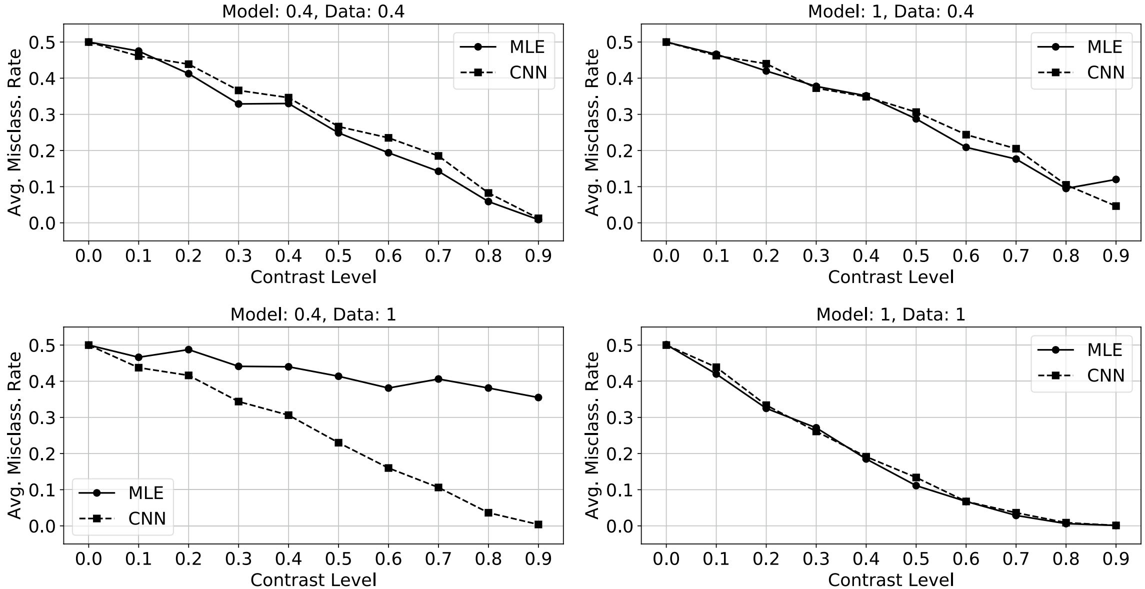

Fig. 6 Average misclassification rates for the ML-based and CNN-based classifiers for different values of κ, see (12). The titles of each panel in the format "Model: XX, Data: YY" mean that for the ML-based method, the value of κ = XX is used at the classification step, see (17), whereas for the CNN-based method, κ = XX is used for generation of training/validation datasets, while κ = YY is used to generate the common test dataset for the two plots in this panel

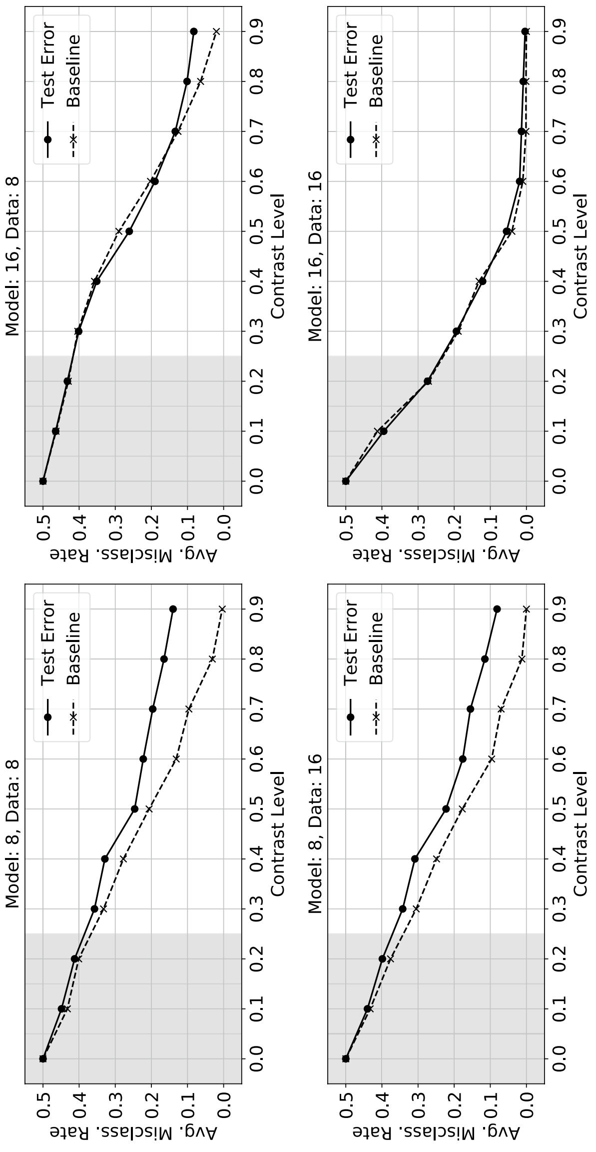

Fig. 7 Average misclassification rates for the CNN-based classifier trained on the contrasts 0.0, 0.1, and 0.2. (This range of contrasts is illustrated by shading.) On each plot, the "baseline" curve corresponds to training on the entire set of contrasts, see the "CNN" plots in Fig. 5. The titles of the panels are the same as in Fig. 5

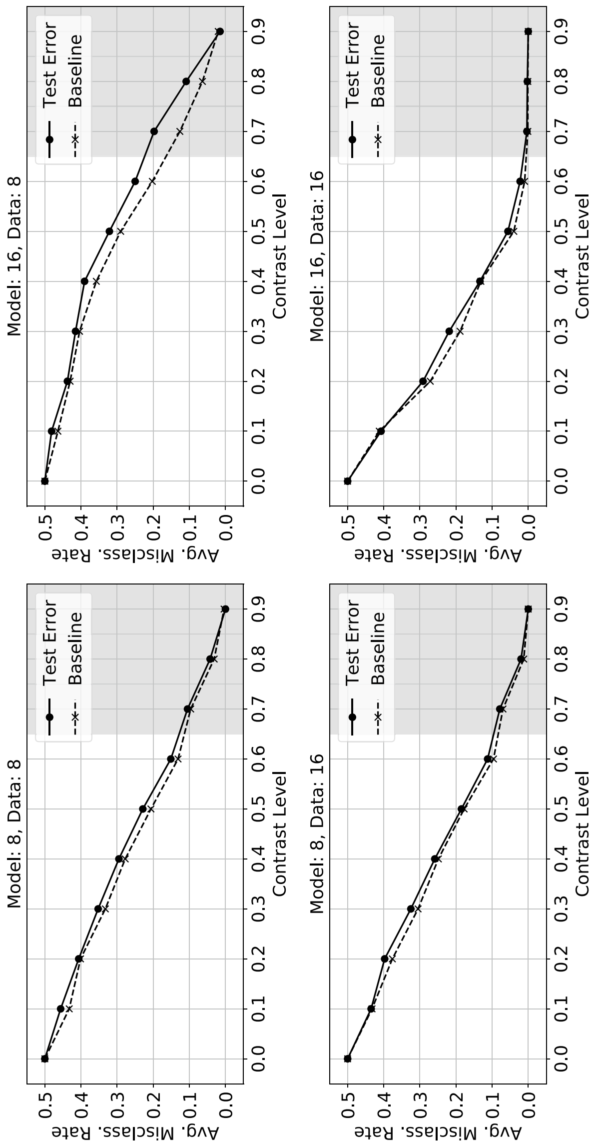

Fig. 8 Same as in Fig. 7, but for the training on the contrasts 0.7, 0.8, and 0.9

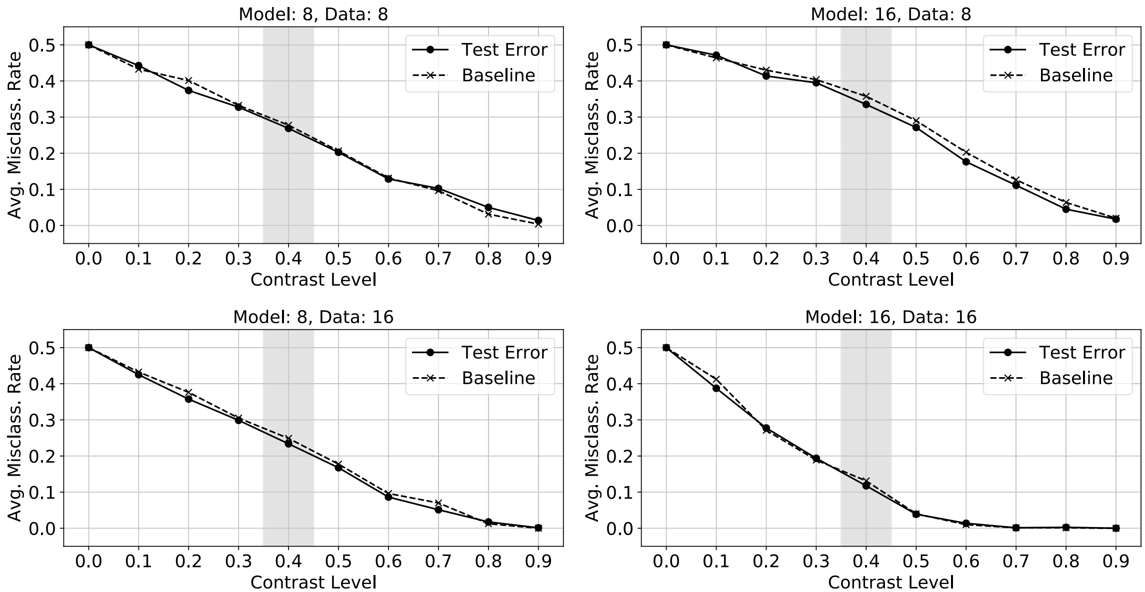

Fig. 9 Same as in Fig. 7, but for the training on a single contrast of 0.4

upper right plot), the two classifiers demonstrate similar adaptation performance. In all cases, the adaptation performance is inferior to that shown with the correct metadata; the same effect was also observed in Fig. 5. Note that the adaptation scenarios illustrated by Fig. 6 are not realistic because, unlike $\zeta_{\max}$ in Fig. 5, the parameter κ characterizes the radar system and should be considered known with a good accuracy, see (12).

5.2 Adaptation of CNN to Different Target Contrasts

Figure 7 demonstrates the performance for the CNN-based classifier trained on the contrasts 0.0, 0.1, and 0.2, other parameters being similar to Fig. 5. We can see that the discrimination quality has decreased with respect to the baseline case shown in Fig. 5, especially for higher contrasts. At the same time, training on the three highest contrasts, see Fig. 8, as well as a single contrast of 0.4, see Fig. 9, yields very little to no loss in the discrimination quality for all contrasts.

6 Conclusions

We have successfully applied the convolutional neural networks (CNNs) to the problem of discrimination between the instantaneous and delayed scatterers in radar images as presented in [11, 12]. The quality of the CNN-based classification is very close to that demonstrated by the benchmark maximum likelihood-based classifier. Both classifiers can generalize, i.e., operate with incorrect metadata or missing training data, although sometimes with a moderate loss of performance.

It should be emphasized that the results obtained in this work have been achieved using the standard CNN architecture, in particular, the most popular 3×3 convolutional kernels, see Fig. 4. At the same time, the images used in this study are highly anisotropic, in particular, the correlation between the neighboring pixels in the same row is much stronger than that across the rows, see Fig. 3. This characteristic feature of the range-delay radar images is due to the properties of the imaging kernel (5), see also the discussion following (11). In our future studies, we will attempt to increase the discrimination quality by adjusting the CNN architecture to the properties of the data. Another possible direction of future research is combining the two classification methods, e.g., by including the output of the ML-based classifier [such as the arguments of the maxima in (17)] into the data for the CNN-based classifier.

Acknowledgements This material is based upon work supported by the US Air Force Office of Scientific Research (AFOSR) under Award Number FA9550-17-1-0230 and by the National Science Foundation under Grant DMS-1638521 to the Statistical and Applied Mathematical Sciences Institute. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the funding agencies.

Funding US Air Force Office of Scientific Research (AFOSR), Award Number FA9550-17-1-0230; National Science Foundation, Grant DMS-1638521 to the Statistical and Applied Mathematical Sciences Institute.

Availability of data and materials (data transparency) Not applicable.

Code availability (software application or custom code) Not applicable.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Albanese RA, Medina RL (2013) Materials identification synthetic aperture radar: progress toward a realized capability. Inverse Problems 29:054001
2. Allen JC, Hobbs SL (1997) Spectral estimation of non-stationary white noise. J Frankl Inst B 334(1):99–116. https://doi.org/10.1016/S0016-0032(96)00060-9
3. Borden B (1998) Dispersive scattering for radar-based target classification and duct-induced image artifact mitigation. In: NATO symposium on non-cooperative air target identification using radar. North Atlantic Treaty Organization, Research and Technology Organization, pp 14.1–14.7
4. Chen VC, Ling H (2002) Time-frequency transforms for radar imaging and signal analysis. Artech House Radar Library. Artech House, Norwood
5. Cheney M, Borden B (2009) Fundamentals of radar imaging, CBMS-NSF regional conference series in applied mathematics, vol 79. SIAM, Philadelphia
6. Cumming IG, Wong FH (2005) Digital processing of synthetic aperture radar data. Algorithms and Implementation. Artech House, Boston
7. Ferrara M, Homan A, Cheney M (2017) Hyperspectral SAR. IEEE Trans Geosci Remote Sens 55(3):1–14
8. Gallager RG (2008) Circularly-symmetric Gaussian random vectors. http://www.rle.mit.edu/rgallager/documents/CircSymGauss.pdf
9. Gallager RG (2008) Principles of digital communication, vol 1. Cambridge University Press, Cambridge
10. Gilman M, Smith E, Tsynkov S (2017) Transionospheric synthetic aperture imaging. Applied and numerical harmonic analysis. Birkhäuser/Springer, Cham. https://doi.org/10.1007/978-3-319-52127-5
11. Gilman M, Tsynkov S (2019) Detection of delayed target response in SAR. Inverse Problems 35:085005. https://doi.org/10.1088/1361-6420/ab1c80
12. Gilman M, Tsynkov S (2020) Statistical characterization of scattering delay in synthetic aperture radar imaging. Inverse Problems Imaging 14(3):511–533. https://doi.org/10.3934/ipi.2020024
13. Goodman JW (1984) Statistical properties of laser speckle patterns. In: Dainty JC (ed) Laser speckle and related phenomena. Springer, Berlin, pp 9–75
14. Hastie T, Tibshirani R, Friedman J (2009) The elements of statistical learning. Springer series in statistics. Springer, New York. https://doi.org/10.1007/978-0-387-84858-7
15. Ioffe S, Szegedy C (2015) Batch normalization: accelerating deep network training by reducing internal covariate shift
16. Jakowatz CV Jr, Wahl DE, Eichel PH, Ghiglia DC, Thompson PA (1996) Spotlight-mode synthetic aperture radar: a signal processing approach. Springer, Berlin
17. Kingma DP, Ba J (2014) Adam: a method for stochastic optimization
18. Oliver C, Quegan S (1998) Understanding synthetic aperture radar images. Artech House Remote Sensing Library. Artech House, Boston
19. Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R (2014) Dropout: a simple way to prevent neural networks from overfitting. J Mach Learn Res 15(56):1929–1958

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.